Managers' Productivity and Recruitment in the Public Sector: The Case of School Principals*

Pablo Muñoz Mounu Prem

Abstract

We study whether differences in management can explain variation in productivity and how more effective managers can be recruited in absence of high-powered incentives. To investigate this, we first extend the canonical teacher value-added model to account for school principals, and we document substantial variation in their ability to improve students' learning. Teachers' survey responses and quasi-experimental designs based on changes in school management validate our measure of principal effectiveness. Then, we leverage the timing of adoption of a civil service reform and show that despite having relatively rigid wages, public schools were able to attract more effective managers after increasing the competitiveness and transparency of their personnel selection process.

Keywords: Public sector, Recruitment, School principals, Managers

---

*This version: April, 2021. 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 grateful to Ernesto Dal-Bó, Alonso Bucarey, Christopher Campos, Dante Contreras, Alessandra Fenizia, Deon Filmer, Patrick Kline, Mathieu Pedemonte, Miguel Purroy, Tatiana Reyes, Damian Vergara, and Christopher Walters for great feedback; and to the participants at the UC Berkeley Labor Economics and the Political Economy seminars, the World Bank Research Group, the CAF Development Bank of Latin America, and the 2nd Bank of Italy Human Capital workshop for their suggestions. We thank the Civil Service, 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 and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 for financial support. José Niño and Luis Serrano provided excellent research assistance. Pablo Muñoz, FGV EPGE Brazilian School of Economics and Finance. Email: pablo.munoz@fgv.br. Mounu Prem, Universidad del Rosario, Department of Economics. Email: francisco.munoz@urosario.edu.co

1 Introduction

Management is a key resource of both private and public enterprises (e.g., Bloom et al., 2013, 2015b), but identifying and recruiting effective managers remains challenging. This problem is especially ubiquitous in the public sector, where discretionary appointments and patronage can be pervasive (e.g., Xu, 2018; Colonnelli et al., 2020), and where incentive schemes are hard to define and mostly absent (Lazear and Shaw, 2007; Finan et al., 2017). However, research on the effectiveness and allocation of managerial talent in the public sector has significantly lagged behind that of politicians and managers in the private sector, despite the high practical relevance of managers for state capacity (Bertrand et al., 2020).

Empirical progress on this area has faced at least two important hurdles. First, the dearth of data makes it difficult to assess performance in the public sector. Second, it is hard to find sources of quasi-experimental variation in personnel selection policies. In this article, we overcome these limitations by leveraging the institutional setting and the rich administrative data of Chilean schools. By focusing on managers in the educational context, we can construct objective measures of their performance to then assess the impact of a civil service reform that made the selection process of public-school principals more competitive and transparent.

We develop a novel extension of the well-known teacher value-added model to disentangle the effect of the school principal from that of her teaching staff and other school-related factors. We find that a one standard deviation increase in principal effectiveness raises students' course grades by 0.29 standard deviations. Teachers' surveys and event studies around the timing of arrival/departure of principals validate our measure of their effectiveness. Leveraging our measure of principals' effectiveness, coupled with quasi-experimental evidence from the adoption of a civil service reform, we study the role of personnel selection policies in the public sector. We find that, despite having a relatively rigid wage scheme, public schools were able to attract more effective managers (a 0.06 SD increase) after increasing the competitiveness and transparency of their selection process.

The first part of our study measures principal effectiveness. We estimate a value-added model using more than 7.7 million student-subject-year observations. This model relates students' academic achievement to school characteristics, and to the fixed effects of 64,770 teachers and 8,061 principals. In Chile, students do not take nationwide standardized tests every year, which prevents us from using test scores to measure students' achievement. Instead, we use students' course grades in Mathematics and Spanish and account for potential biases following Peteck and Pope (2019). We argue that course grades are better suited to our purposes for several reasons. First, course grades—and not test scores—determine grade retention and high school graduation. Importantly, course grades also determine college admission and access to student financial aid.[1] Second, 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).[2] Third, as is common in value-added models, we examine performance longitudinally, which mitigates concerns about grade inflation. Finally, grades are better at capturing non-cognitive traits valued by the labor market (e.g., Bowles and Gintis, 1976; Bowles et al., 2001; Heckman et al., 2006).

To accommodate principal effects into the teacher value-added model is challenging. On the one hand, it is necessary to distinguish principals' effectiveness from teachers' effectiveness. On the other hand, it is necessary to disentangle the effect of the principal from other school-level factors. To account for the former challenge, we follow the seminal work by Abowd et al. (1999) and estimate a two-way fixed effects model that leverages personnel and student switches across principals (*within* the largest connected set) to separately identify their effectiveness and that of their teaching staff. To accomplish the latter, 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 it also overcomes problems of weak identification that could arise if we added school fixed effects in this setting (e.g., Jochmans and Weidner, 2019).

[1] High-school GPA, along with entrance exam scores, are the key components of the composite scores used for post-secondary admissions and financial aid eligibility (see Hastings et al., 2013). In 2017, course grades and contextual course grades had an average weighting of 40% in the college admission score.

[2] Figure A.1 in the appendix show a strong association between course grades and test scores. A one standard deviation increase in course grades is associated with a 0.6 standard deviation increase in contemporaneous test scores, and with a 0.8 standard deviation increase in college admission scores.

To validate our estimates of principal quality, we present quasi-experimental evidence from an analog to the ideal experiment of random assignment of principals to schools. In the spirit of Chetty et al. (2014) and Angrist et al. (2017), our design exploits the turnover of high and low value-added principals for identification. Using event studies of principals' arrivals and departures, we find that student achievement changes sharply after event time *as predicted* by our measure of principal effectiveness. We complement this evidence with a specification diagnostic in the spirit of Rothstein (2010). Focusing on a subset of students who switched between schools to transit from primary to secondary level, we show that the effectiveness of the principal in the school of destination does not affect their achievement in the school of origin. Finally, in the spirit of recent research looking at managerial practices within schools (e.g., Di Liberto et al., 2015; Bloom et al., 2015a), we study teachers' perceptions to validate our measure of principal effectiveness. We show that more effective principals are associated with a larger fraction of their teaching staff highly agreeing with positive statements about them.

Having established that school principals matter, we go on to ask whether a better selection process can improve the allocation of principal effectiveness. We begin by documenting that the compensation of most public-school principals is rigid and mainly based on statutory payments, and that the association between principals' effectiveness and wages is mostly driven by bonus payments at private schools. We interpret these findings through the lenses of a two-sided matching model (e.g., Abowd and Farber, 1982; Logan, 1996), which underscores the importance of personnel selection for the allocation of talent in absence of high-powered incentives. Then, by leveraging quasi-experimental variation from a civil service reform, we directly assess the effects that better personnel selection has on the allocation of principal effectiveness.

Since the eighties, the recruitment of public schools' principals in Chile has been the exclusive responsibility of the municipalities. This has given local politicians a significant degree of discretion over the appointments of school personnel, a feature that—in a similar context—has been associated with negative effects on student outcomes (e.g., Akhtari et al., 2018). To reduce politicians' discretion over the appointment of school principals, in 2011 Chile enacted a reform that modified the selection of school leaders. Under the new system of selection, local politicians still have a say in principals' election but only after a competitive and transparent competition has shortlisted a subset of candidates based on their merit and suitability. These competitions are publicly advertised and led by a third-party human resources agency and are overseen by the Civil

Service, the agency responsible for selecting the highest-level bureaucrats of the central government. To assess the effects of the new selection system on principal effectiveness, we estimate a difference-in-differences regression that compares the change in principal's quality after a principal is appointed under the new selection system to the change in principal's quality after principal turnover at private schools. The dynamic version of our difference-in-differences approach provides visual support to our identification strategy.

We find that merit-based selection increased principal effectiveness in public schools by 0.06 standard deviations; *ceteris paribus* this would be enough to close half of the gap in course grades between public and private schools after 5 years. We find similar results when we only keep public schools for estimation, and therefore we identify the effect of ADP selection from variation in the timing of adoption; we also show that this result is robust to recent developments in the literature (i.e., De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2020). Reassuringly, placebo exercises looking at principal turnover at private schools as well as pre-reform principal turnover at public schools fail to detect positive effects on principal quality. We conclude with suggestive evidence that merit-based selection has positive long-run effects as it increases the average composite score that determines post-secondary enrollment and eligibility for student financial aid.

Related literature

Our paper contributes to different branches of the economics literature. First, our work relates to a large body of research on the measurement of productivity—more specifically, to models of value-added (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). We add to this literature by extending the canonical model to accommodate managers' effects. In doing so, we also contribute new evidence to a growing literature 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 organizations (e.g., McCormack et al., 2014; Bloom et al., 2015b; Lavy and Boiko, 2017; Rasul and Rogger, 2018; Fenizia, 2019), as well as to an incipient literature on the effectiveness of school principals (e.g., Branch et al., 2012; Coelli and Green, 2012; Dhuey and Smith, 2014; Grissom et al., 2015).

Second, our work also speaks to research on personnel economics (see Lazear and Shaw, 2007; and Finan et al., 2015 for reviews). Recent literature has shown that patronage is a common feature in public sector appointments (Xu, 2018; Akhtari et al., 2018; Colonnelli et al., 2020; Voth and Xu, 2019), with mixed findings regarding its effects on state capacity.[3] More related to our work, Estrada (2019) studies the effect of decreasing the share of teachers hired under discretion and finds that it has a positive effect on school level outcomes. Likewise, Scot et al. (2020) show that civil service examinations can be an effective way to screen state judge candidates in Brazil, and Moreira and Pérez (2021) show that performance-based hiring for US bureaucrats led to less turnover and better quality workers (although it had no effect on performance). More broadly, our paper also relates to the literature showing that different selection policies and incentive methods can attract different types of public sector servants (e.g., Ferraz and Finan, 2011; Dal Bó et al., 2013; Finan et al., 2017; Deserranno, 2019; Deserranno et al., 2019; Ashraf et al., 2020). We contribute to this literature by showing that a more transparent selection system based on third-party screening led to the selection of more effective managers. We also document that newly appointed principals are slightly younger and more likely to have previously worked in administration positions in addition to the private schooling sector.

Finally, our paper also contributes to the educational 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) or on the measurement of principal value-added in isolation (e.g., Coelli and Green, 2012; Dhuey and Smith, 2014; Grissom et al., 2015). We add to this literature by providing an estimation framework that properly accounts for key actors in the education production function to disentangle principal effectiveness. Our study is also related to recent work on school personnel (e.g., Rothstein, 2015; Biasi, 2018; Loyalka et al., 2019) and the labor market of school principals (e.g., Cullen et al., 2016). In contrast to them, we focus on the demand side of the labor market and show that in absence of high-powered incentives, public schools can attract more effective managers by improving their selection process.

The rest of this paper is organized as follows. The next section discusses institutional

[3] For instance, Xu (2018) shows that patronage in the British Empire decreased revenue and investment, while Voth and Xu (2019) finds that patronage can lead to better selection, which in turn leads to better outcomes.

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 studies principals' labor market and the effects of the new selection system on the allocation of principal quality in public schools. It also offers a brief discussion about our findings. Finally, section 5 concludes.

2 Background 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.

Institutional setting

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.[4] 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 (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.[5] 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. As of 2018, enrollment at private, subsidized private, and public schools represented 7, 53, and 40 percent, respectively.

[4]A reform in 2008 established a new voucher targeted to low-income students. This represented the first major change to the voucher policy program. For an evaluation of this policy, see Neilson (2019).

[5]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.

Alongside the privatization of the educational system, the 1981 reform also decentralized it by transferring control of public schools from the central government to municipal authorities. Administrative departments of municipal education and municipal education corporations were created to administrate the public schools. Not surprisingly, the efficacy and probity of these departments/corporations were 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 municipalities or 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. During the period we study, public school teachers once again belong to a national teachers' union and their wages are determined by a rigid formula that is negotiated between their union and the government. Wages are subject to seniority increments and other adjustments such as allowances for leadership responsibilities, professional training, and for working in difficult conditions.[6]

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 increased the attributions of school principals and created a new system to appoint them. After the reform, all new principals at public schools i) must be elected in a public and competitive contest run by an external human resources firm and overseen 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 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 a 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. For more details about the reform, see "Ley 20.501 Calidad Y Equidad de la Educación".

Like many countries, Chile has a nationwide standardized curricula.[7] The Chilean cur-

[6]Teachers in private schools are also eligible for some of these allowances, but they are mainly subject to the Private Labor Code, implying that their wages are often individually negotiated with the schools.

[7]Countries that have a national curriculum include: France, Hungary, Ireland, Italy, Japan, Korea,

riculum 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 A.2 in the appendix, the government not only provides teachers with the curriculum guides and official textbooks, but also with lesson plans and exams. Curriculum guidelines establish minimum content goals and fundamental objectives for education, which ultimately determine course grades and grade retention. Students are evaluated continuously throughout the year, and, in general, each subject's annual grade is based on more than four evaluations. Teachers in a particular subject determine the course grade in that subject. Grades are awarded on a scale from 1 to 7 in intervals of 0.1, with a minimum passing grade of 4. The Ministry of Education also administers a national standardized test called SIMCE. This test is taken annually but only by students in the 4th, 8th, and 10th grades. Finally, to gain admission into higher education, most students take a standardized college entrance exam known as PSU. Students must complete exams in mathematics and language, and many students also take optional tests in other subjects. Entrance exam scores, along with high-school GPA, are the primary components of the composite scores used for post-secondary admissions, scholarships, and student loan eligibility (Hastings et al., 2013).[8]

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 panel spans the period from 2011 to 2016 and has information on the academic performance of all students, 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

the Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, and the United Kingdom. While most states in the U.S. follow common guidelines for a core curriculum, there is no national curriculum as such.

[8] Students who are not admitted through the centralized admission system enroll in non-selective technical or professional schools or in newer universities operating outside the centralized system.

and Spanish.[9] 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, hours of contract). For a subset of these teachers, we recover their perceptions about the school principal 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, fraction of disadvantaged students, fraction of parents with a college degree, parents' income level, the identity of the school principal, and whether she or he was elected through the new selection system. We complement this data with characteristics of the municipalities where the schools are located. We consider all schools to estimate principals' and teachers' effectiveness, except preschools, adults' schools, and special education 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 their expenditures. 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 compensation components regulated by law but unrelated to performance, such as payments for experience and for teacher certification; it also includes other payments 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 compensation components 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 subsidized private schools, and other discretionary payments and gratifications related to transportation, food, and holidays. On average, principals earn around 2,700 USD per month. According to the representative survey CASEN (in 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). The average monthly wage of 2,739 USD corresponds to roughly 11 times the

[9] 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.

legal minimum wage.

Table 1 presents detailed descriptive statistics of students, schools, and principals at public and private schools. Panel A, which focuses on students' course grades, test scores, and grade retention, shows that students attending private schools obtain higher course grades and test scores (0.25 and 0.5 standard deviations respectively) and are almost 4 percent points less likely to fail a grade. In terms of school characteristics, Panel B shows that private schools tend to serve students who obtain (on average) higher scores for college admission. They also serve more students, have larger classrooms, have fewer teachers per student, and have slightly better attendance. Some of these differences are likely related to the fact that only 21% of private schools are in rural areas versus 63% 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 32% in private schools. Finally, Panel C of Table 1 presents the descriptive statistics of principals' demographics and wages. Compared to private schools, public schools pay lower wages to the principals and their compensation relies more on statutory payments and less on bonuses. In public schools, 35% of the wage corresponds to the base, 56% to statutory payments, and only 11% to bonuses, while in private schools these figures correspond to 51, 26, and 24 percent. In terms of demographic characteristics, public school principals have more tenure and most of them are male; this is 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 principals' effectiveness. We consider a specification that relates academic achievement to student characteristics, school characteristics, and to the teachers and school principal, as follows:

$$\begin{aligned} Y_{it+1} = \quad & \gamma_t + \rho_{g(i,t)} + \beta_0 f(Y_{it-1}, \bar{Y}_{it-1}, \rho_{g(i,t)}) + \beta_1 X_{it} + \\ & \underbrace{\mu_{j(i,t)}}_{\text{teacher FE}} + \underbrace{\theta_{p(i,t)}}_{\text{principal FE}} + \underbrace{\phi_0 X_{s(i,t)t} + \phi_1 \bar{X}_{s(i,t)}}_{\psi_s:\quad \text{school CRE}} + e_{it+1}, \end{aligned} \tag{1}$$

where Y_{it+1} is the course grade obtained by student i in year $t+1$, and γ_t and ρ_g stand for year and grade fixed effects. To remove systematic bias from teachers evaluating their own students, we follow Peteck and Pope (2019) and focus on future, instead of contemporaneous, course grades while restricting the sample to students for whom the teacher, in a given subject, changed between t and $t+1$.[10] As is 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 for student's age A_{it} and principal tenure $A_{p(i,t)}$. Since we only observe a subset of principals switching between schools, we do not include school fixed effects. Instead, we use correlated random effects (e.g., Mundlak, 1978; Chamberlain, 1980) to account for school heterogeneity. Specifically, we include a combination of fixed and time-varying characteristics of each school. Time-variant school characteristics include total enrollment, the fraction of disadvantaged students, the share of low-income and high-income parents, and the share of parents with a college degree. Fixed school characteristics include the across-time average of the previous list of time-variant characteristics as well as indicators for whether the school is public, subsidized private, or private, and for whether it is located in a rural area.[11] Building on 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.

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. 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. We exclude preschools, adults' schools, and special education schools from our analysis; and we also exclude classes that had more than one teacher per year and eliminate the bottom and top one percent of classroom size outliers. By taking teacher effects into account directly, our empirical model disentangles the effectiveness of the school principal from the effectiveness of the teaching staff. We find 9,120,301 such students (70% of our original sample). Finally, since teacher and principal effects in (1) are identified by movers and can only be compared within connected sets (e.g., Abowd et al., 1999), we estimate our model within the largest connected set of teachers

[10] Intuitively, our specification gives 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. This avoids confounding easy-graders with high value-added teachers.

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

and principals. We end-up with an estimation sample of 7,735,683 student-subject-year observations (85% of the previously restricted dataset), including 1,977,231 students, 64,770 teachers, and 8,061 principals. Reassuringly, as shown by Table A.1 in the appendix, we do not find strong evidence of selective sample attrition in terms of grades, subject, attendance, student performance, or teachers' characteristics.

The standard deviation of our principal fixed effects is 0.39; in other words, one standard deviation in principals' effectiveness is associated with a 0.39 standard deviations increase in students' course grades. Although for each principal $p = (1, ..., P)$, her estimated effectiveness $\hat{\theta}_{p(i,t)}$ is an unbiased estimate of her true effect on students' achievement, the standard deviation of $\hat{\theta}_{p(i,t)}$—which summarizes the overall variability in principals' effectiveness—is an upwardly biased estimate of the standard deviation of $\theta_{p(i,t)}$. This occurs because $\hat{\theta}_{p(i,t)}$ equals $\theta_{p(i,t)} + \hat{\epsilon}_p$, where $\hat{\epsilon}_p$ is a least squares sampling error. This can be particularly serious when estimating teacher value-added because of the small populations of students used to identify their value-added, but it is less of a concern in our case as there are many more observations of students per principal than there are per teacher.

In our two-way fixed effects setting, quadratic forms of $\hat{\theta}_{p(i,t)}$ might also be biased due to "bottlenecks" in the connected set (e.g., Jochmans and Weidner, 2019; Kline et al., 2020). To account for this issue, we compute the adjusted standard deviation of $\hat{\theta}_{p(i,t)}$ using the formula proposed in Krueger and Summers (1988):

$$SD(\theta_p) \approx \sqrt{\text{var}(\hat{\theta}_p) - \sum_{p=1}^{P} \hat{\sigma}_p^2 / P},$$

where $\hat{\sigma}_p^2$ is the standard error of $\hat{\theta}_{p(i,t)}$. Consistent with the prevalence of bias in variance components, the adjusted standard deviation of principal effectiveness is 0.29, implying that the "true" variance of principal effectiveness is almost half of its unadjusted counterpart.[12] We leverage the previous estimate of the true variation in principal effects and the estimation error to rescale our measure of principal effectiveness by its "reliability". Following the teacher value-added literature (e.g., Aaronson et al., 2007), we shrink our

[12]It is worth noticing that this adjustment neglects the covariances among the ϵ_p, slightly underestimating the standard deviation of $\theta_{p(i,t)}$.

estimates of principal effectiveness to obtain an adjusted measure θ_p^*, as follows:

$$\theta_p^* = \hat{\theta}_p \times \frac{\sigma_p^2}{\sigma_p^2 + \hat{\sigma}_\epsilon^2}, \tag{2}$$

where $\hat{\theta}_p$ is our OLS estimate of the value added by principal p, σ_p^2 is our estimate of the "true" variation in principal effectiveness (calculated as described above), and $\hat{\sigma}_\epsilon^2$ is the noise associated with the estimate of principal p's effect, namely, the estimation error for $\hat{\theta}_p$. To alleviate concerns related to measurement error, we use this adjusted measure of principal effectiveness θ_p^* through the paper.

Finally, in appendix Table A.2, we document how estimated principal effectiveness correlates with observable characteristics. To do so, we regress the estimated principal fixed effects from (1) on age, age squared, gender, and indicators for holding a college degree and for their experience in previous "schooling type" of positions. Female principals appear to be on average more productive than their male counterparts. Principal effectiveness is also strongly correlated with experience and depicts a concave profile at public schools. While interesting, these correlations could be explained by differential selection patterns into schools and managerial career. Thus, they should not be interpreted causally.

On the validity of our estimates

Our 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 nor 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, even if we could compute principal and teacher fixed effects, those estimates would be noisier as the estimation sample would be significantly smaller. Furthermore, recent research by Cuesta et al. (2020) has shown that low-performing students are underrepresented in test days, generating distortions in school quality information.

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 principals' and teachers' effectiveness. This is because course grades—and not test scores—is the metric that determines high school graduation and advancement into higher level courses. Moreover, it also captures the non-cognitive dimensions of students' development such as perseverance, dependability, and consistency (e.g., Bowles and Gintis, 1976), and research has shown that these non-cognitive traits are strongly related to both school and labor market outcomes (e.g., Bowles et al., 2001; Heckman et al., 2006). Furthermore, regarding our empirical specification, we examine students' performance longitudinally, which should help to mitigate concerns about grade inflation. Finally, 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 A.1 in the appendix shows a strong relationship between test scores and course grades. Panel A considers a sample of students for whom we observe SIMCE test scores and course grades contemporaneously for Math and Spanish (between 2011 and 2016); and shows that a one standard deviation increase in course grades is associated with a 0.6 standard deviation increase in test scores. Likewise, panel B considers a sample of students for whom we can compute their college admission score (in the 2017 process); and shows that a one standard deviation increase in course grades is associated with a 0.8 standard deviation increase in test scores.

While 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' setting 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. 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 and teacher fixed effects, the correlation between the assignment of students to principals and other determinants of students achievement is innocuous. Although 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 some of the validation exercises proposed in this literature (e.g., Rothstein, 2010; Chetty et al., 2014)

We begin by presenting quasi-experimental evidence from an analog to the ideal experiment of random principal assignment to schools. In the spirit of Chetty et al. (2014) and the omnibus test in Angrist et al. (2017), our design exploits principal turnover for identification. To understand the design, suppose an effective principal moves from school A to another school B at time $t = 0$. Because of this change, students at school A will have a less effective principal on average than they had in the previous years (before $t = 0$). If our estimates of principal effectiveness have predictive content, we would expect grades' growth at school A to decrease after a high value-added principal exits. Using event studies of principals' arrivals and departures, we find that student achievement changes sharply across time *as predicted* by the change in principal effectiveness, when high or low value-added principals enter or exit a school.

This quasi-experimental design rests on the identification assumption that principal turnover within a school is uncorrelated with student and school characteristics. Although untestable, this assumption is plausible insofar as teachers and students are unlikely to immediately switch to a different school because the principal changed. We begin with event studies of course grades around the entry and exit of low and high value added principals (Figure 1). Let year 0 denote the school year that a principal enters or exits a school and define all other school years relative to that year (e.g., if the principal enters in 2013, year 2011 is -2 and year 2015 is +2). We define an entry event as the arrival of a principal whose effectiveness is either in the top or bottom 20% of the distribution of principal effectiveness, and we define exit events analogously.

Figure 1, panel A, plots the impact of the exit of a low value-added principal on mean course grades. The series plots school-year means of standardized course grades in the two years before and after a low value-added principal exits the school. We do not condition on any other covariates in this figure: each point simply shows average course grades for different years within a school. The change in mean course grade gains in the school in which the low value-added principal exits are 0.07 SD from year -1 to 1. The null hypothesis that this change is 0 is rejected with $p < 0.001$. More importantly, the magnitude of the increase in mean test score gains is very similar to the change in mean principal effectiveness, which is 0.09. The hypothesis that the observed impact on mean score gains equals the increase in mean value-added is not rejected ($p = 0.35$), consistent with the idea that our estimates of principal effectiveness are forecast unbiased.

The remaining panels of Figure 1 repeat the event study in Panel A for other types of arrivals and departures. Panel B examines course grades around the entry of a low value-added principal. Grades drop sharply relative to prior years when a low value-added principal takes leadership of a school, showing that the entry of the low value-added principals lowers the achievement of students. Panels C and D analyze the arrival and departure of high value-added principals. Course grades in the school lead by a new principal fall relative to prior grades when high value-added principals exit and rise when high value-added principals leave. In all but the last panel, the change in course grade gains is significantly different from 0 with $p < 0.05$ ($p = 0.6$ in the last panel) but is not significantly different from what one would forecast based on the change in mean principal effectiveness. Together, these event studies provide reassuring evidence that principals can have a significant impact on the academic achievement of students, independent of school characteristics.

Our data also allows us to implement a falsification test similar to that in Rothstein (2010). For this exercise, we focus on a subset of students who switched schools and who were consequently exposed to more than one principal. The intuition of the test is simple: if the effectiveness of the principal in the school of destination impacts students' learning in the school of origin, that would be evidence of model misspecification. To better approximate the ideal experiment that assigns students to different principals, we focus on students attending schools that do not offer secondary education at the time of their transition to high school. We find 11,542 such events. To perform this test, we compute "jackknife" estimates of principal effectiveness, i.e., we estimate principal fixed effects again but now excluding all observations of the students who switched schools and consider specification diagnostic for both a model of gains and a model with lagged dependent variable. Table 2 presents the results. Reassuringly, we do not find much evidence of positive correlation between the standardized course grade gains of the students (the pre-assignment variable) and the effectiveness of their future principal (the treatment variable).

Two-way fixed effects specifications are simple and tractable. Nevertheless, when used for estimating worker and firm fixed effects, these specifications are prone to be criticized (see Card et al., 2018 for a discussion). This is because OLS estimates of worker and firm effects will be biased unless worker mobility is uncorrelated with the time-varying residual components of wages, a strong assumption on workers' mobility if one considers some models of wage determination (e.g., Gibbons et al., 2005). Since our model also considers additive teacher and principal effects, one might be worried about the bias of measure of principal quality. We address this issue in the spirit of Card et al. (2013) and plot the mean course grades of the students taught by teacher j before and after the teacher started working under a new principal p. For this, we first residualize course grades using all controls in our main specification (including lagged course grades), but excluding teachers' and principals' fixed effects. Figure 2 presents these profiles.

We see that teachers who moved from working under a principal with students in the lowest (1st) quartile of course grades to working under a principal with students in the highest (4th) quartile experienced a large average gain in their students' course grade, while those who moved in the opposite direction experienced large loses. Moving within a quartile group, by comparison, is associated with relatively small changes in residualized course grades. Moreover, although we do not condition on holding teacher-principal relationships for at least 2 years, the trends prior and after moving are very similar across groups, and the mean change in course grades for teachers who move in opposite directions between quartile groups (e.g, from quartile 1 to quartile 2, versus from quartile 2 to quartile 1) are of similar magnitude and uniformly of opposite sign. While not perfect, this figure is consistent with the symmetry implications of the additive two-way fixed effects model with exogenous mobility.

Finally, we analyze teacher surveys to contrast our estimates with the perception of the school community. In the spirit of recent research looking at managerial practices within schools (e.g., Di Liberto et al., 2015; Bloom et al., 2015b), we explore how teachers perceive the practices of school principals according to their estimated effectiveness. We start by analyzing a set of surveys that 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—which are publicly available for the years 2010, 2011, 2014, and 2015—to create a dummy variable at the survey-respondent level that equals one if the teacher "highly agrees" with a given statement about the school 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 a simple regression of the fraction of teachers highly agreeing with a given statement about the school principal on our estimated measure of principal effectiveness. Figure 3 presents the effect size and confidence intervals based on bootstrapped standard errors.

We find that 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 engages teachers*, *principal knows teacher needs*, *principal engages parents*, and *principal knows students needs* by around 6%; *principal makes good decisions*, *principal includes teachers*, *principal is effective*, and *principal does a good job* by around 5%; and *principal promotes a good work climate*, *principal is good at communicating*, and *principal can be trusted* by around 3 to 4%. Table 3 in the appendix presents our point estimates and shows the robustness of these results to accounting for multiple hypothesis testing using the step-down procedure proposed in Romano and Wolf (2005),[13] and also to a permutation exercise where we randomly reshuffle the principal fixed effects 1,000 times and then 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 (to gauge how likely would it be to obtain our results just by chance). We consider the previous results as *prima facie* evidence that our estimates of principal effectiveness are sensible.

4 The Labor Market and Selection of School Principals

In this section, we study the labor market of school principals in Chile. We begin by documenting that the compensation of most public-school principals is rigid and mainly based on statutory payments. In this context, we study the extent to which improving personnel selection can help to bring more effective principals to public schools.

[13] A similar approach is taken by Bassi et al. (2020) in the context of heterogeneous treatment effects of a policy evaluation in Chile.

Descriptive analysis of wages at public and private schools

Public sector compensation usually does not include pay for performance (Finan et al., 2015), and although there is a good rationale for this,[14] 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 teachers' firing policies (e.g., Staiger and Rockoff, 2010; Boyd et al., 2011; Cowen and Winters, 2013), is also relevant to the Chilean case. To study this, we use administrative data on wages from public and subsidized private schools from 2015 to 2017. Figure A.3 in the appendix presents some features of our data. Perhaps not surprisingly, we find that hourly wages (residualized with respect to year and county fixed effects) at public schools are significantly less spread and 0.09 log points lower than those at the voucher-private schooling sector. Like in the US, wages in Chilean public schools also rely less on pay-for-performance. On average, the bonus component of wages represents 22% of the principal's salary in voucher-private schools but only 9% in public schools.

To study whether workers' characteristics command the same price in public and voucher schools, we estimate the following mincer type regression model:

$$\ln(\text{wage}_{pt}) = \alpha + \beta_0 \text{Voucher}_{pt} + \beta_1 [X_{pt} - \bar{X}] + \beta_2 \text{Voucher}_{pt} \times [X_{pt} - \bar{X}] + \rho_{m(p,t)} + \gamma_t + \epsilon_{pt}, \quad (3)$$

where $\ln(\text{wage}_{pt})$ represents the logarithm of the average hourly wage paid to principal p at time t, Voucher_{pt} is an indicator that equals one if the principal works at a voucher-private school (and zero otherwise), 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_2, and it represents the factor price differential between sectors. Importantly, the vector X_{pt} includes principal characteristics such as our measure of her effectiveness $\hat{\theta}_p$, tenure, tenure squared, an indicator for whether the principal is female, and for whether she has a permanent contract. This specification also allows us to study how the different components of wages relate to principal effectiveness. For this, we decompose the dependent variable $\ln(\text{wage}_{pt})$ into two components: $\ln(\text{base}_{pt})$ and $\ln(\text{wage}_{pt}/\text{base}_{pt})$, where "base" corresponds to the sum of the minimum legal wage and the statutory

[14]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).

payments described in section 2, and base_{pt} corresponds to the total wage minus the bonuses.

Table 4 presents the point estimates and bootstrap standard errors (100 replications) obtained from these regressions. Columns 1 and 2 show the association between the log wage of school principals and their characteristics, while columns 3 to 6, replicate this analysis but decompose log wages into its base and a bonus components. Our estimates reveal a sizable and statistically significant wage premium in voucher-private schools. On average, voucher schools pay 15% more than public schools, and most of this premium is driven by the bonus components of wages. Regarding the association between wages and principals' effectiveness, we fail to reject the null of no association between the variables in public schools; however we find a modest, although statistically significant, association at voucher-private schools where increasing principal effectiveness by one standard deviation is associated with a 2% increase in wages, a correlation that is also driven by the bonus components of wages. The results in this table also reveal interesting patterns. For instance, we find that the tenure profile is salient at public schools, but not at voucher-private schools, a result consistent with the prevalence of fixed wage schemes in the public sector. More interestingly, we find that the size of the gender wage gap is large—almost 11%—at voucher-private schools, but close to zero at public schools, a finding in line with recent evidence by Biasi and Sarsons (2020) showing that flexible pay reforms can increase the gender wage gap.

The relationship between wages and self-selection is a core topic in labor economics. Indeed, the seminal observation by Roy (1951) that insofar as higher quality workers demand higher compensation, employers paying higher wages can attract those workers has become pervasive in the economics literature. However, this view underestimates the role of labor demand. Higher wages might not suffice nor be the only relevant variable because workers' matching in the labor market also depends on: i) their idiosyncratic taste, i.e., workers might have specific preferences for the public or private sector (e.g., Deserranno, 2019; Ashraf et al., 2020), and ii) the labor demand that they face (i.e., the personnel selection process of the employers constraints workers' choice *de facto*). Indeed, the intuition derived from models with two-sided selection (e.g., Abowd and Farber, 1982; Logan, 1996) is that schools could offset the "labor supply effect" by making informed choices; in other words, selection can accentuate or counteract the self-sorting of workers à la Roy. For the interested reader, in Appendix B, we present a thorough exposition of a two-sided matching model for the labor market. We build on Logan

(1996)'s model, which is itself a variant of the deterministic two-sided matching models studied in game theory, and simulate the allocation of talent under different selection schemes.

In the following section, we assess the extent of this dimension of the labor market by leveraging quasi-experimental variation in the introduction of a merit-based personnel selection policy in public schools.

Selection and recruitment of public school principals

With the purpose of strengthening the managerial performance and administration of public education, in 2011 Chile defined a new procedure to select school principals at public schools. 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. 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".[15]

We exploit the non-eligibility of private schools and the timing of adoption of this within public schools to study the impact of the new selection policy on the allocation of principal effectiveness. We use digitized data from all the competitions for the position of school principal between 2012 and 2016 to identify the time when a new principal was appointed under the new selection process. As shown by Figure A.4 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 appointing a principal. As expected, the number of principals elected under the new regime increased over time,

[15] 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 (composed by a Deputy Director, an Inspector General and a Chief Technician) without having to call a contest for those positions and they can fire up to 5% of teachers with a bad or regular teacher evaluation.

with around 370 new principals elected every year since 2012. Table A.3 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 public 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, less rural, and with a larger share of disadvantaged students, although these differences are smaller and less precise. Importantly, when comparing late to early adopters, we don't find statistically significant differences in terms of student achievement. Finally, column 7 presents summary statistics for all private schools. Private schools are similar to the Ever ADP public schools in terms of size and rural status, but they serve children from better socioeconomic backgrounds and have higher test scores.

Regarding principals' characteristics, in Table 5 we present the differences between public school principals who were selected by the new system and those who were not. For comparison purposes, we add the characteristics of principals at private schools. Compared to other public school principals, 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.); they are also more likely to have worked in the private schooling sector in the past. Moreover, principals elected under the new system are slightly younger and more likely to have a college degree. In the same vein of the previous descriptive statistics, Table A.4 in the appendix 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 the data suggests that these principals are leaving schools with more students, located 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 preference for amenities hypothesis, these principals are arriving at municipalities that have higher income and more years of schooling.

To assess the effects of this new selection system, we compare the change in principal's effectiveness triggered by a principal's turnover under the new selection system to the change in principal's effectiveness triggered by a principal's turnover at private schools. More specifically, we estimate the following *difference-in-differences* regression:

$$y_{st} = \alpha_s + \alpha_t + \beta_1 \times ADP_{st} \times \text{ Principal Turnover}_{st} + \beta_2 \times \text{Principal Turnover}_{st} + \sum_t \Phi_t' X_s I[year = t] + \epsilon_{st}, \quad (4)$$

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. Principal Turnover$_{st}$ is a dummy variable that equals one from the first year (after 2012) when the school selected a new principal, and ADP_{st} is a dummy variable that takes the value one from the first year a public school appointed a principal using the new selection system.[16] X_s is a vector of predetermined (as of 2010) school characteristics including income per student, share of disadvantaged students, total enrollment, test scores, as well as municipality level controls including poverty rate, average household income, unemployment rate, average years of education, and literacy rate. We interact this set of controls with year fixed effects, thus adding flexible time trends parametrized by these school and municipality characteristics. Finally, α_s and α_t are school and year fixed effects, and ϵ_{st} is an error term robust to heteroscedasticity and clustered at the school level.

Our parameter of interest is β_1, and it captures the difference in the change of principal effectiveness after a public school appoints a principal using the new system and the change in principal effectiveness after a private school appoints a new principal. For those schools that had a principal turnover, we include a window of four years around the adoption to facilitate the study of the timing of the effect (results are robust to not imposing this restriction). Results are shown in Table 6. Estimates from column 1 suggest that—relative to the effect of principal turnover on principal effectiveness at private schools—the turnover at public schools due to the appointment of a new principal elected under the ADP system increases principal effectiveness by 0.06 standard deviations. Reassuringly, we find that there is a non-significant negative change in effectiveness after a principal turnover at private schools. 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.

[16] In Chile, this new selection system is known by the acronym of *Alta Dirección Pública* (ADP).

(2009) and truncate our analysis sample based on a propensity score that estimates the probability that a school appoints a principal under the new system. We also estimate an effect of 0.06 standard deviations using this truncated sample.

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 principals's effectiveness. To partially address this concern, we estimate a variation of model (4) with a dynamic treatment. Specifically, we estimate the following regression:

$$\begin{aligned} y_{st} \; &= \alpha_s + \alpha_t + \textstyle\sum_{j=-4}^{-2} \beta_j \times ADP_s \times I[k=j] + \sum_{j=0}^{4} \beta_j \times ADP_s \times I[k=j] \\ &+ \textstyle\sum_{j=-4}^{-2} \delta_j \times I[k=j] + \sum_{j=0}^{4} \delta_j \times I[k=j] + \sum_t \gamma_t' X_s I[year=t] + \epsilon_{st}, \end{aligned} \quad (5)$$

where k corresponds to the year relative to the first time a school appointed a principal using the new selection system if the school is public or the year relative to the first time a private school experienced principal turnover. The estimation sample includes all types of school independent of whether they elected a principal via the new selection process or not. Figure 4 presents this dynamic version of the *difference-in-differences* approach, where we plot the parameters of interest β_js, which capture the difference in principal effectiveness in period j relative to the omitted period (-1) for schools that had a principal turnover via the ADP system relative to the difference in principal effectiveness in period j relative to the omitted period (-1) for private schools that experience principal turnover. This figure provides visual support for our identification strategy, as the point estimates are around zero and not significant in the pre-period. A joint test for the coefficients being all equal to zero in the pre-period cannot be rejected at conventional levels. Moreover, 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 and municipality characteristics.

We now present the results when we only keep public schools for estimation, and therefore we identify the effect of ADP selection from variation in the timing of adoption. We find similar results: schools that selected a principal using the ADP system experienced a statistically significant increase in principal effectiveness of 0.045 standard deviations when compared to *never* and to *late* adopters (see column 6 of Table 6). Recent literature on this type of two-way fixed estimation have shown that estimates from this model can substantially differ from the group's ATE in the presence treatment heterogeneity (e.g., Borusyak and Jaravel, 2017; De Chaisemartin and d'Haultfoeuille, 2020). We assess the relevance of this concern by following De Chaisemartin and d'Haultfoeuille (2020) and computing the number of estimates with a negative weight. We find that only 7% of our estimates have a negative weight (the sum of the weights is −0.019). We also compute the decomposition of the two-way fixed effect estimate following Goodman-Bacon (2018). We find that more than 68% of our estimate is computed from differences between *treated* and *never treated* and only 11% comes from the comparison between "late" and "early" treated (see Figure A.6). Together, these results suggest that the concerns regarding this staggered difference-in-differences estimation should be minor. However, as a robustness check, in columns 8 and 9 of Table 6 we present the estimation using the models suggested by De Chaisemartin and d'Haultfoeuille (2020) and Callaway and Sant'Anna (2020). In both cases we find a positive and significant effect that ranges between 0.035 and 0.061.[17]

Finally, as additional robustness checks, we perform two placebo exercises. First, we consider any principal turnover that happened before the reform (2008-2011) in public schools as a treatment (column 9). Second, we consider any principal turnover that happened after the reform (post 2011) in private schools as a treatment (column 10). 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 increase principal effectiveness in these placebo exercises. If anything, our results suggest that turnover is associated with a decrease in principal effectiveness, although these results are not significant at standard levels. As shown by Table A.5, this finding remains unchanged if we considered the models proposed by De Chaisemartin and d'Haultfoeuille (2020) and Callaway and Sant'Anna (2020).

[17] In Figure A.5, we present the dynamic versions of these two-way fixed effect models, that leverage variation within public schools.

The impact of the ADP reform on equity and long-term outcomes

Before concluding, we briefly discuss the plausible impacts of this reform on educational equity and longer-term student outcomes.

The civil service reform studied here aimed to improve equity by boosting achievement at public school, thus a natural benchmark to assess its effectiveness is the public-private gap in terms of course grades, which equals 0.17 standard deviations within our time span (see Table 1). We have shown that a one standard deviation increase in principal effectiveness raises course grades by 0.29 standard deviations, and that the impact of ADP adoption is a 0.06 standard deviations increase in principals' effectiveness. This implies that—*ceteris paribus*—the reform created course grade gains of 0.017 standard deviations per year; enough to reduce the public and private course grade gap by half in 5 years. Insofar as improving gains in course grades leads to better long-run outcomes, we would expect the positive effects on educational equity to persist in students' adult life.

In Chile, course grades—along with college entrance exams—are a key component of the composite scores used to determine scholarship and student loan eligibility as well as for post-secondary admissions. Among the students accepted into college in 2017, the correlation between their standardized course grades and college admission score was 0.83 (see figure A.1 panel B). This strong correlation is due to two facts. First, course grades and contextual course grades have an average weighting of 40 percent into the admission score. Second, since the entrance exams in Chile are oriented to measure how much of the school curriculum has been learned,[18] course grades are also correlated with the students' performance in the entrance exams. In 2017, a one standard deviation increase in students' course grades was associated with 0.47 and 0.38 standard deviation higher scores in Math and Spanish, respectively.

Leveraging data from the centralized college admissions system (between 2010 and 2017) and our preferred specifications (4) and (5) with school and municipality controls, we can study the impact of "ADP" appointments on college entrance exams and application scores. Naturally, for this analysis we need to restrict our estimating sample to high schools whose students apply to higher education via the centralized college admissions

[18]The Chilean entrance exams are more like the American College Testing than the SAT as the former is oriented to measure how much of the school curriculum has been learned while the latter attempts to measure cognitive aptitudes (González Adonis et al., 2017).

system. We present our findings in Table 7 and Figure 5. Results suggest that the appointment of a principal elected under the new selection system increases the average score (between Math and Spanish) by 0.08 standard deviations. Importantly, the final application score that determines admission at a given institution-major pair (i.e, a degree) increases by 0.13 standard deviation after the appointment of a new ADP principal.[19] The latter results, coupled with i) the positive sorting induced by admission scores into enrollment at more selective institutions (e.g., Rodríguez et al., 2016), and ii) the large positive returns associated to more selective institution-mayors (e.g., Hastings et al., 2013; Zimmerman, 2019), leads us to expect positive long-run impacts of this policy on income and other non-pecuniary outcomes (e.g., Oreopoulos and Salvanes, 2011; Bautista et al., 2020).

Although we cannot say for sure what the welfare impact of this policy was, a number of factors suggest positive and potentially large effects on educational quality and equity. For one thing, school principals are a salient component of the educational production function and have an impact on *all* the students attending their schools, thus policies oriented to recruit better principals might be an effective way to boost school quality at a relatively low cost. In addition, reforms like the one studied here can be an alternative to achieve accountability and flexibility in public education (Abdulkadiroğlu et al., 2011) in settings where politicians or unions have discretion over the appointment of public school personnel. Policies such as providing management training for principals in public schools (Fryer et al., 2017) or endowing public schools with greater autonomy (Clark, 2009) might enhance the positive effects of a more competitive and transparent recruitment.

5 Conclusion

Education has become an essential public service and a landmark of state capacity as it is an important determinant of individual earnings, macroeconomic growth, and equity (Barro, 1991; Card, 2001; Chetty et al., 2017). In Chile, government expenditure in education represents more than one fifth of the budget, and the efficiency and equity of

[19]The final application score is a weighted average of a student's score in the entrance exams and her course grades, with weights defined by each institution-major. In our analysis, we consider the weights of the institution-major that is most preferred by a student as revealed by her preferences in the college application process.

educational policy is often at the center of the political debate. In this paper, we have assessed the extent to which personnel selection can help to improve public schools' quality. To do so, we extended the the canonical teacher value-added model to measure principals' effectiveness while accounting for teachers' and schools' effects. We document substantial variation in the ability that school principals have to improve students' achievement. Quasi-experimental evidence looking at principals' arrivals and departures coupled with teacher surveys and a falsification test provide *prima facie* evidence that our measure of principal effectiveness is sensible. Then, leveraging our estimates of principal quality, we have studied the labor market of school principals. We began by documenting that, relative to their private counterparts, public schools reward tenure more and do not penalize women in terms of wages. However, principal effectiveness is not compensated in public schools, a finding that underscores the role that personnel selection plays to accentuate or counteract the allocation of talent in the public sector.

Using quasi-experimental variation coming from a civil service reform, we assess the extent to which improving personnel selection can be an effective way to enhance the recruitment of candidates and ultimately the allocation of talent in public schools. Our results show that improving the competitiveness and transparency of the recruitment process can increase the average effectiveness of public schools' leaders by 0.06 standard deviations, enough to close half of the gap in course grades between public and private schools after 5 years. We also present suggestive evidence of positive effects of this policy on long-run outcomes, as it improved the scores used for college admission and student financial aid. Since school leadership affects all students within a school, we believe that personnel policies intended to improve the quality of school principals might be a cost-effective way to increase state capacity.

References

Aaronson, D., L. Barrow, and W. Sander (2007): "Teachers and student achievement in the Chicago public high schools," *Journal of Labor Economics*, 25, 95–135.

Abdulkadiroğlu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane, and P. A. Pathak (2011): "Accountability and flexibility in public schools: Evidence from Boston's charters and pilots," *The Quarterly Journal of Economics*, 126, 699–748.

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

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67, 251–333.

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, E. Davenport, and S. S. Lee (2020): "Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services," *American Economic Review*, 110, 1355–94.

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.

Barro, R. J. (1991): "Economic growth in a cross section of countries," *The Quarterly Journal of Economics*, 106, 407–443.

Bassi, M., C. Meghir, and A. Reynoso (2020): "Education quality and teaching practices," *The Economic Journal*, 130, 1937–1965.

Bautista, M. A., F. González, L. R. Martinez, P. Munoz, and M. Prem (2020): "Does Higher Education Reduce Mortality? Evidence from a Natural Experiment in Chile," *Evidence from a Natural Experiment in Chile (October 23, 2020)*.

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.

Bertrand, M., R. Burgess, A. Chawla, and G. Xu (2020): "The glittering prizes: Career incentives and bureaucrat performance," *The Review of Economic Studies*, 87, 626–655.

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.

Biasi, B. and H. Sarsons (2020): "Flexible Wages, Bargaining, and the Gender Gap," Tech. rep., National Bureau of Economic Research.

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., R. Lemos, R. Sadun, and J. Van Reenen (2015a): "Does management matter in schools?" *The Economic Journal*, 125, 647–674.

Bloom, N., C. Propper, S. Seiler, and J. Van Reenen (2015b): "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.

Borusyak, K. and X. Jaravel (2017): "Revisiting event study designs," *Available at SSRN 2826228*.

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.

Callaway, B. and P. H. Sant'Anna (2020): "Difference-in-differences with multiple time periods," *Journal of Econometrics*.

Card, D. (2001): "Estimating the return to schooling: Progress on some persistent econometric problems," *Econometrica*, 69, 1127–1160.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 36, S13–S70.

Card, D., J. Heining, and P. Kline (2013): "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly Journal of Economics*, 128, 967–1015.

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.

Chetty, R., J. N. Friedman, E. Saez, N. Turner, and D. Yagan (2017): "Mobility report cards: The role of colleges in intergenerational mobility," Tech. rep., national bureau of economic research.

Clark, D. (2009): "The performance and competitive effects of school autonomy," *Journal of political Economy*, 117, 745–783.

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 (2020): "Patronage and selection in public sector organizations," *American Economic Review*, 110, 3071–99.

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.
Cuesta, J. I., F. González, and C. L. Philippi (2020): "Distorted quality signals in school markets," *Journal of Development Economics*, 147, 102532.
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.
De Chaisemartin, C. and X. d'Haultfoeuille (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 110, 2964–96.
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.
Deserranno, E., M. Stryjan, and M. Sulaiman (2019): "Leader selection and service delivery in community groups: Experimental evidence from uganda," *American Economic Journal: Applied Economics*, 11, 240–67.
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.
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.
Estrada, R. (2019): "Rules versus discretion in public service: Teacher hiring in Mexico," *Journal of Labor Economics*, 37, 545–579.
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.
——— (2017): "The personnel economics of the developing state," *Handbook of economic field experiments*, 2, 467–514.
Fryer, R. G. et al. (2017): "Management and student achievement: Evidence from a randomized field experiment," Tech. rep., National Bureau of Economic Research.
Gibbons, R., L. F. Katz, T. Lemieux, and D. Parent (2005): "Comparative advantage, learning, and sectoral wage determination," *Journal of Labor Economics*,

23, 681–724.

González Adonis, P., V. Arancibia Clavel, and D. Boyanova (2017): "Talento académico, vulnerabilidad escolar y resultados en la prueba de selección universitaria," *Estudios pedagógicos (Valdivia)*, 43, 171–191.

Goodman-Bacon, A. (2018): "Difference-in-differences with variation in treatment timing," 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*.

Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2013): "Are some degrees worth more than others? Evidence from college admission cutoffs in Chile," Tech. rep., National Bureau of Economic Research.

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.

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*, 303–328.

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.

Jochmans, K. and M. Weidner (2019): "Fixed-effect regressions on network data," *Econometrica*, 87, 1543–1560.

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.*

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

Kline, P., R. Saggio, and M. Sølvsten (2020): "Leave-out estimation of variance components," *Econometrica*, 88, 1859–1898.

Krueger, A. B. and L. H. Summers (1988): "Efficiency wages and the inter-industry wage structure," *Econometrica: Journal of the Econometric Society*, 259–293.

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 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.

MCCORMACK, J., C. PROPPER, AND S. SMITH (2014): "Herding cats? Management and university performance," *The Economic Journal*, 124, F534–F564.

MOREIRA, D. AND S. PÉREZ (2021): "Civil Service Reform and Organizational Practices: Evidence from the 1883 Pendleton Act," .

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

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

OREOPOULOS, P. AND K. G. SALVANES (2011): "Priceless: The nonpecuniary benefits of schooling," *Journal of Economic perspectives*, 25, 159–84.

PETECK, N. AND N. POPE (2019): "The Multidimensional Impact of Teachers on Students," *Working Paper*.

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.

RODRÍGUEZ, J., S. URZÚA, AND L. REYES (2016): "Heterogeneous economic returns to post-secondary degrees: Evidence from chile," *Journal of Human Resources*, 51, 416–460.

ROMANO, J. P. AND M. WOLF (2005): "Stepwise multiple testing as formalized data snooping," *Econometrica*, 73, 1237–1282.

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.

SCOT, T., R. DAHIS, AND L. SCHIAVON (2020): "Selecting Top Bureaucrats: Admission Exams and Performance in Brazil," *Available at SSRN*.

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.

VOTH, J. AND G. XU (2019): "Discretion and Destruction: Promotions, Performance, and Patronage in the Royal Navy," *Working Paper*.

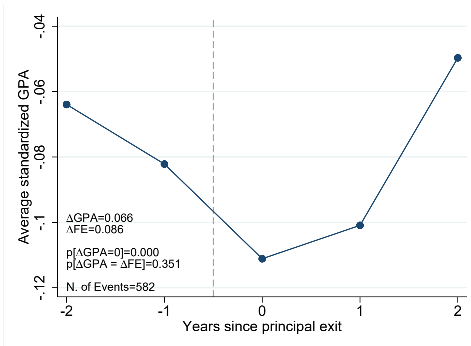
WOOLDRIDGE, J. M. (2010): "Correlated random effects models with unbalanced panels," *Michigan State University, Department of Economics*.

XU, G. (2018): "The costs of patronage: Evidence from the british empire," *American Economic Review*, 108, 3170–98.

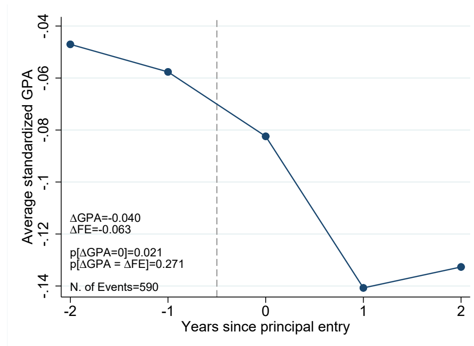
ZIMMERMAN, S. D. (2019): "Elite colleges and upward mobility to top jobs and top incomes," *American Economic Review*, 109, 1–47.

Figures and Tables

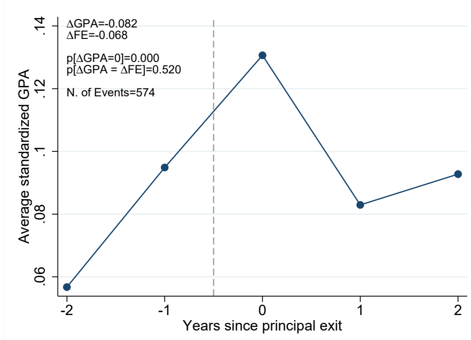
Figure 1: Impacts of Principal Entry and Exit on Student's Performance



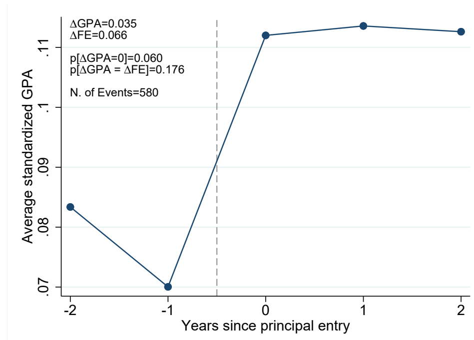
A. Low Value-Added Principal Exit



B. Low Value-Added Principal Entry



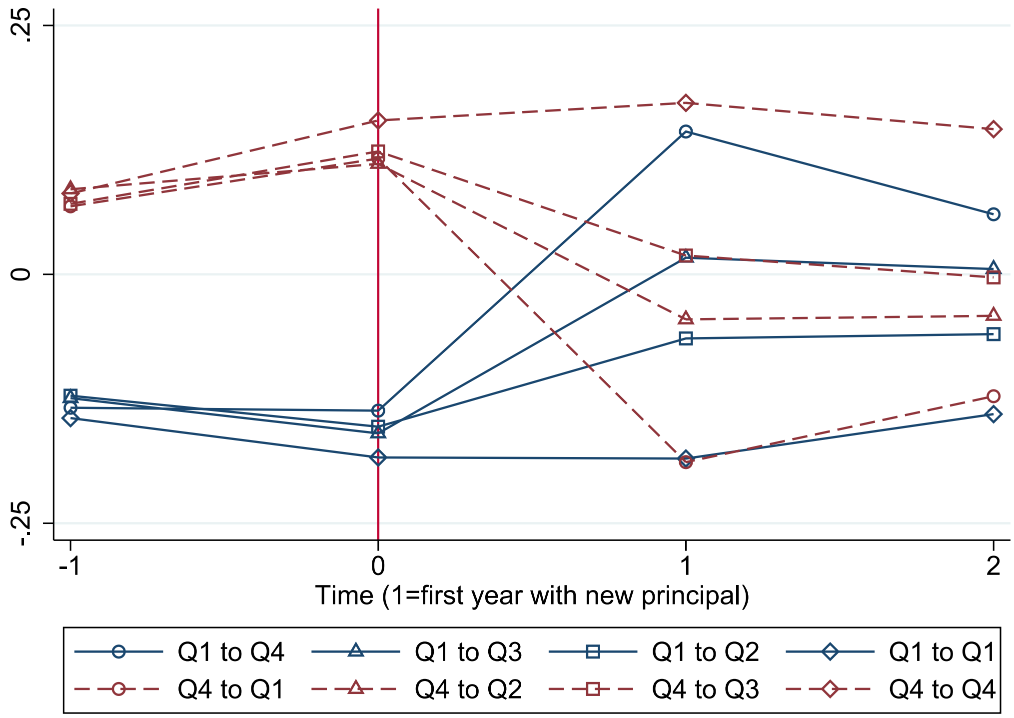
C. High Value-Added Principal Exit



D. High Value-Added Principal Entry

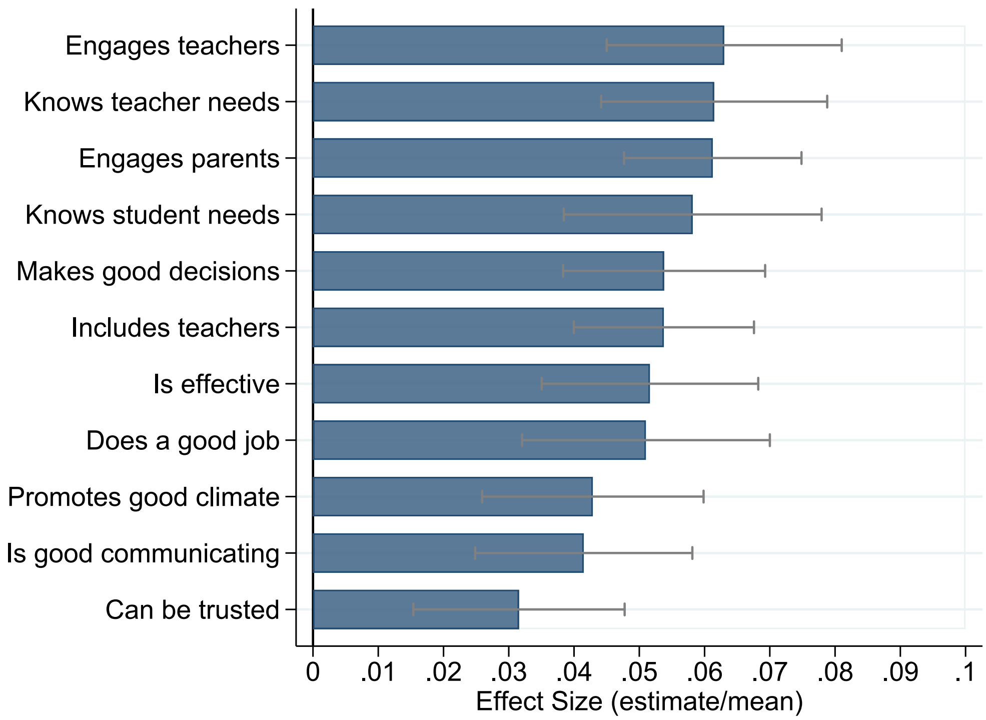
Notes: These figures plot event studies of of standardized course grades as principals arrive at or leave a school at year t=0. Panels A and B plot the impact of the exit and entry of a low value-added principal (principals with VA in the bottom 20% of the distribution) on mean course grades. Likewise, Panels C and D plot the impact of the exit and entry of a high value-added principal (principals with VA in the top 20% of the distribution) on mean course grades. To construct each panel, we first identify the set of principals who entered or exited a school between 2012 and 2015 and define event time as the school year relative to the year of entry or exit. Each panel reports the change in mean grades' gains (current minus lag grades) from t=-1 to t=1 and the change in mean estimated VA. We report p-values from a tests of the hypotheses that the change in achievement gains from t=-1 to t=1 equals the change in VA and that the change in achievement gains equals 0.

Figure 2: Mean Residualized GPA of Teachers who change Principal, classified by Quartile of Principals' Mean Residualized GPA at Origin and Destination



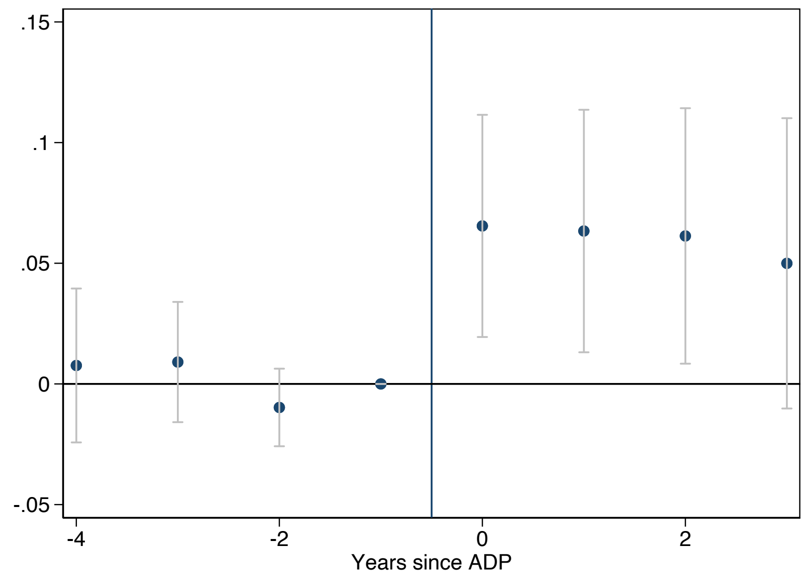
Notes: This figure plots the mean residualized course grades of teachers who changed principal in 2011-2016. We consider the first time a teacher switches to work under a new principal but we do not condition on holding the old or new job relationship for a minimum number of years. Each principal is classified into quartiles based on mean residualized course grades of the students at her school. Course grades are residualized with respect to the same set of controls considered in our main specification (1), except teacher and principal fixed effects.

Figure 3: Teachers' Survey Responses

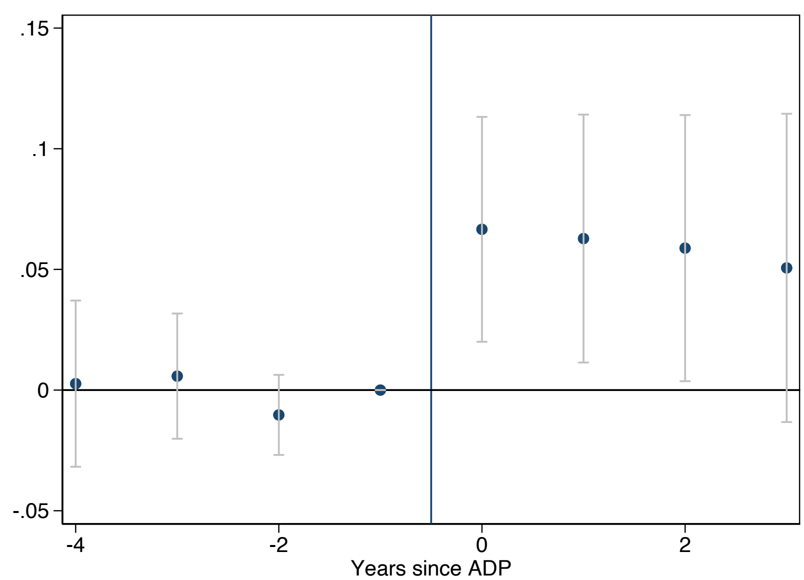


Notes: This figure shows the association between our measure of principal effectiveness and the likelihood that the teaching staff agrees with positive statements about the principal. Using teachers' surveys we create an indicator if a teacher "highly agrees" with a given statement. In cases when the survey had 5 or 4 options we always use the highest number to create this indicator. We take the average across respondents at the school-year level and assign this to a principal. Then, using a data set at the principal-level, we estimate a simple regression of the fraction of teachers highly agreeing with a given statement about the school principal on principal effectiveness.

Figure 4: Principal Selection and Principal Effectiveness



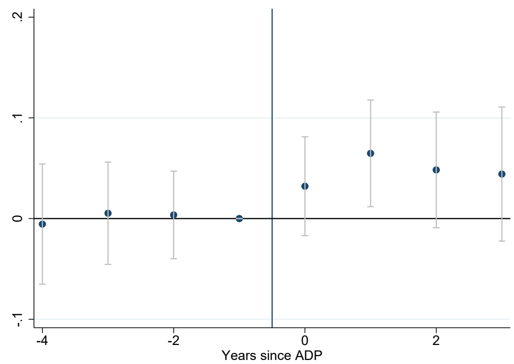
A. No Controls



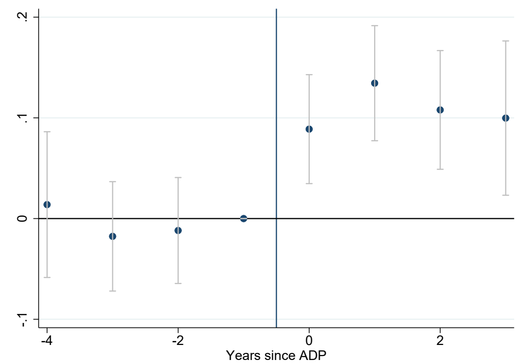
B. With Controls

Notes: This figure shows the impact of appointments under the new selection system on the effectiveness of public schools' principals. Specifically, it plots the point estimates and 95% confidence intervals estimated from equation (5). 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.

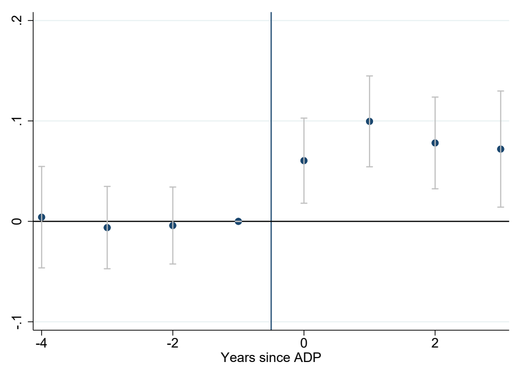
Figure 5: Principal Selection and College Admission Scores



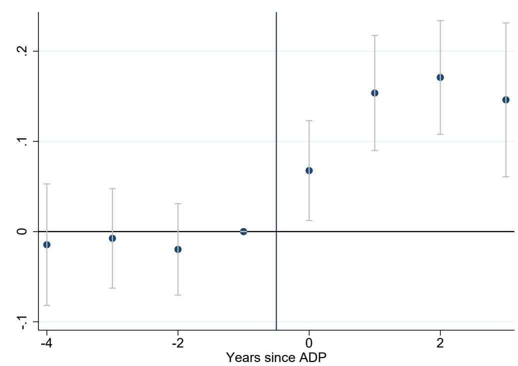
A. Math



B. Spanish



C. Average Math and Spanish



D. Application score

Notes: This figure shows the impact of appointments under the new selection system on college admission scores. The figure plots the point estimates and 95% confidence intervals estimated from equation (5). Panels A and B show the impact on the mandatory exams of Math and Spanish, while Panel C shows the impact of the average between Math and Spanish. Panel D plots the impact on the composite score used for admissions. This score is a weighted average of entry exam scores and course grades, with weights defined by each degree (institution-major pair). We consider the weights of the most preferred degree of a student (as revealed by her preferences in the application process) to construct this score. All panels include 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

	Summary Statistics					By Type of School		
	Mean	Std. dev.	Median	10th pctile	90th pctile	Public	Private	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Student characteristics								
Math course grade	5.3	0.8	5.2	4.2	6.4	5.16	5.31	-0.15***
						(0.80)	(0.85)	(0.00)
Spanish course grade	5.3	0.7	5.3	4.3	6.3	5.19	5.38	-0.19***
						(0.74)	(0.73)	(0.00)
Math test scores	263.5	50.6	264.7	196.1	328.3	244.83	274.45	-29.61***
						(47.21)	(49.26)	(0.00)
Spanish test scores	258.3	50.4	261.4	188.4	322.7	243.54	266.35	-22.81***
						(49.72)	(48.97)	(0.00)
Ever Grade retention (%)	8.4	27.8	0.0	0.0	0.0	10.78	7.06	3.72***
						(31.02)	(25.62)	(0.00)
Panel B: School characteristics								
Avg. College Admission Score	592.1	38.5	588.0	548.2	645.1	576.19	596.50	-20.31
						(33.47)	(38.64)	(0.00)***
Enrollment	306.0	404.6	156.0	2.0	838.0	215.90	413.15	-197.25***
						(303.61)	(477.15)	(0.00)
Annual subsidy per student (USD)	2423.5	3704.1	1445.8	840.3	4086.9	2977.05	1627.09	1349.96***
						(4302.89)	(2397.34)	(0.00)
Share of disadvantaged students	46.1	36.0	55.0	0.0	92.0	57.30	31.70	25.60***
						(32.88)	(34.64)	(0.00)
Teachers per hundred students	8.1	18.3	5.8	2.8	14.3	9.15	6.84	2.30***
						(12.44)	(23.65)	(0.00)
Rural school	43.5	49.6	0.0	0.0	100.0	62.40	21.07	41.32***
						(48.44)	(40.78)	(0.00)
School attendance	86.1	9.5	87.8	76.2	95.1	85.78	86.50	-0.72***
						(9.23)	(9.94)	(0.00)
Panel C: Principal characteristics								
Wage (USD)	2846.5	2139.9	2597.8	1609.8	4159.0	2648.59	3030.74	-382.15***
						(2582.42)	(1601.50)	(0.00)
% Base salary	43.0	19.8	35.8	23.3	75.4	34.87	50.55	-15.68***
						(13.68)	(21.47)	(0.00)
% Bonus	17.4	19.8	8.4	0.8	50.9	10.64	23.63	-12.99
						(15.45)	(21.30)	(0.00)***
% Statutory	40.5	27.9	39.9	9.8	69.7	55.75	26.39	29.36***
						(27.62)	(19.54)	(0.00)
Permanent contract	90.9	28.7	100.0	100.0	100.0	88.32	93.37	-5.04***
						(32.12)	(24.89)	(0.00)
Age	54.3	9.9	55.0	40.0	65.0	54.83	53.74	1.09***
						(8.57)	(10.99)	(0.00)
Female	54.4	49.8	100.0	0.0	100.0	47.08	61.20	-14.12***
						(49.92)	(48.74)	(0.00)

Notes: Columns 1 to 5 present summary statistics for students, schools, and principals. Columns 6 to 8 show the differences between private and public schools in terms of students', schools', and principals' characteristics. Columns 6 and 7 present the average and standard deviation (in parentheses), and column 8 presents the difference between both columns and the p-value of this difference (in parentheses). These descriptive statistics consider students, schools, and principals in our main estimation sample. Principals' wages are only available for public and subsidized private schools from 2015 to 2017.

Table 2: Falsification Test

	Gain Achievement Model		Lagged Achievement Model	
	(1)	(2)	(3)	(4)
$\hat{\theta}_{p(-i)}$ at school s_0	0.112**	0.110**	0.181***	0.174***
	(0.053)	(0.054)	(0.051)	(0.051)
$\hat{\theta}_{p(-i)}$ at school s_1		0.017		0.056
		(0.042)		(0.037)
Course Grade			0.600***	0.600***
			(0.015)	(0.015)
Course Grade^2			0.112***	0.112***
			(0.009)	(0.009)
Course Grade^3			-0.001	-0.001
			(0.007)	(0.007)
N. of Events	11,542	11,542	11,542	11,542

Notes: This table shows the results from the validation exercise discussed in section 3. We consider a sample of students who switched schools at the end of primary because their school did not offer secondary education at the time. We find 11,542 of these events where students were exposed to different principals because they had to switch schools. For this exercise, we use "jackknife" estimates of principal effectiveness, i.e., estimates of principal effectiveness in a sample that leaves out all observations of the students who switched schools. Bootstrapped standard errors (100 replications) clustered by school of origin.

Table 3: Teachers' Survey Responses

% Teachers highly agreeing that the principal:	$\hat{\beta}$ (1)	Std error (2)	Mean Dep Var (3)	Obs (4)	Placebo p-value (5)	RW p-value (6)
Does a good job	0.023***	(0.004)	0.460	5351	0.000	0.001
Can be trusted	0.016***	(0.004)	0.521	5349	0.000	0.001
Makes good decisions	0.025***	(0.004)	0.459	6384	0.000	0.001
Is effective	0.023***	(0.004)	0.448	6380	0.000	0.001
Is good at communicating	0.022***	(0.004)	0.529	5355	0.000	0.001
Engages teachers	0.028***	(0.004)	0.444	6365	0.000	0.001
Engages parents	0.028***	(0.003)	0.464	6384	0.000	0.001
Knows teacher needs	0.027***	(0.004)	0.439	6387	0.000	0.001
Knows student needs	0.029***	(0.005)	0.502	5351	0.000	0.001
Includes teachers	0.025***	(0.003)	0.469	7228	0.000	0.001
Promotes good work climate	0.023***	(0.005)	0.525	5272	0.000	0.001

Notes: To construct this table, we first create an indicator variable at the survey respondent level which takes a value of one if the survey respondent 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. Columns 1 and 2 report the estimated coefficients and bootstrapped standard errors from a regression on the fraction of the teaching staff highly agreeing with a given statement and our measure of principal effectiveness. To gauge effect sizes, we report the mean of the dependent variable in column 3. Column 5 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 s where the value of $\hat{\beta}_s$s was greater than or equal to our estimate $\hat{\beta}$. Finally, column 6 presents p-values adjusted for multiple hypothesis testing using the step-down procedure of Romano and Wolf (2005).

Table 4: Principal Compensation and Principal Effectiveness

	ln(Wage)		ln(Base)		$\ln\left(\frac{\text{Wage}}{\text{Base}}\right)$	
	(1)	(2)	(3)	(4)	(5)	(6)
Private	0.105***	0.148***	-0.171***	-0.124***	0.276***	0.272***
	(0.007)	(0.007)	(0.010)	(0.011)	(0.007)	(0.011)
Principal Effectiveness	-0.004	-0.005	0.001	-0.000	-0.005	-0.005
	(0.005)	(0.004)	(0.006)	(0.006)	(0.004)	(0.004)
Principal Effectiveness x Private	0.023**	0.020**	-0.001	-0.003	0.024***	0.022***
	(0.009)	(0.008)	(0.011)	(0.010)	(0.007)	(0.007)
Female		-0.003		0.010		-0.014*
		(0.008)		(0.010)		(0.007)
Female x Private		-0.114***		-0.078***		-0.036**
		(0.015)		(0.020)		(0.015)
Age		0.032***		0.043***		-0.011**
		(0.005)		(0.007)		(0.005)
Age x Private		0.005		-0.020**		0.025***
		(0.007)		(0.009)		(0.006)
Age^2		-0.000***		-0.000***		0.000**
		(0.000)		(0.000)		(0.000)
Age^2 x Private		-0.000		0.000**		-0.000***
		(0.000)		(0.000)		(0.000)
Perm. Contract		0.067***		0.078***		-0.010
		(0.016)		(0.018)		(0.011)
Perm. Contract x Private		0.136***		0.063*		0.074***
		(0.032)		(0.034)		(0.021)
Hours Contract		0.005		0.012**		-0.008
		(0.005)		(0.005)		(0.008)
Hours Contract x Private		0.026***		0.017***		0.009
		(0.006)		(0.005)		(0.008)
College Degree		0.032*		0.040**		-0.008
		(0.017)		(0.019)		(0.012)
College Degree x Private		-0.019		-0.003		-0.016
		(0.025)		(0.032)		(0.019)
Observations	9,898	9,898	9,898	9,898	9,898	9,898
R-squared	0.181	0.303	0.163	0.241	0.227	0.234

Notes: This table presents the estimates from specification (3). We focus on a sample of principals for whom we have an standardized measure of effectiveness and detailed wage data from 2015 to 2017. Wage data is only available for public and subsidized private (voucher) schools. All specifications include year and municipality fixed effects. Bootstrapped standard errors (100 replications) clustered at the principal level are in parenthesis.

Table 5: Characteristics of Principals by ADP Status

	Public Schools			Private Schools
	Not ADP	ADP	Difference	
	(1)	(2)	(3)	(4)
Panel A: Ever worked				
As teacher	0.541	0.442	-0.099***	0.430
	(0.498)	(0.497)	(0.000)	(0.495)
As admin. support worker	0.280	0.355	0.075***	0.226
	(0.449)	(0.479)	(0.000)	(0.418)
As administrative worker	0.929	0.950	0.022**	0.916
	(0.258)	(0.217)	(0.011)	(0.277)
In a private school	0.009	0.035	0.025***	0.229
	(0.096)	(0.183)	(0.000)	(0.420)
Panel B: Principal characteristics				
College degree	0.838	0.901	0.063***	0.893
	(0.368)	(0.299)	(0.000)	(0.309)
Age	57.217	55.781	-1.435***	54.299
	(8.760)	(8.943)	(0.000)	(11.977)
Female	0.490	0.490	-0.000	0.615
	(0.500)	(0.500)	(0.988)	0.489
Observations	2,058	1,769	3,827	4,433

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). Finally, column 4 present the average and standard deviation for school principals at private schools.

Table 6: Principal Selection and Principal Effectiveness

	All Schools					Public Schools				Private Schools
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(a) Principal Turnover × ADP	0.060***	0.065***	0.064***	0.063***	0.062***	0.045***	0.035**	0.061**		
	(0.023)	(0.024)	(0.023)	(0.024)	(0.023)	(0.017)	(0.012)	(0.030)		
(b) Principal Turnover	-0.025	-0.028	-0.028	-0.028	-0.027				-0.042	-0.031*
	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)				(0.104)	(0.018)
Observations	30,714	30,714	30,714	30,714	29,508	14,168	14,168	14,168	5,303	17,498
# of Schools	4934	4934	4934	4934	4732	2389	1666	2802		
R-squared	0.931	0.931	0.931	0.931	0.930	0.925	0.925	0.925	0.958	0.935
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School controls	No	Yes	No	Yes	No	No	No	No	No	No
Municipality controls	No	No	Yes	Yes	No	No	No	No	No	No
p-value $a + b = 0$	0.018	0.016	0.018	0.026	0.020					

Notes: This table presents the effects of the new selection system (ADP) on the standardized measure of principal effectiveness discussed in section 3. "ADP" is a dummy that takes the value one after the first time a school selects a principal under the ADP system. "Principal Turnover" is a dummy that takes the value one after the first time a school selects a new principal (after 2012). Columns 1 to 4 estimate the regressions described by equation (4). 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 8.2%. Column 6 estimates the main regression only within public schools that selected a principal under the ADP system. Column 7 implements the model suggested by De Chaisemartin and d'Haultfoeuille (2020), while column 8 shows the result for the model suggested by Callaway and Sant'Anna (2020). Column 9 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 292. Column 10 shows a similar placebo exercise where we focus only on principal turnover after 2012 but in private schools. The number of private schools that had a turnover after 2012 is 1,590. Robust standard errors clustered at school level in parenthesis.

Table 7: Principal Selection and College Admissions Scores

	College Entrance Exams score			Application score
	Math	Spanish	Average	
	(1)	(2)	(3)	(4)
(a) Principal Turnover × ADP	0.045**	0.115***	0.080***	0.132***
	(0.018)	(0.018)	(0.015)	(0.019)
(b) Principal Turnover	-0.040***	-0.050***	-0.045***	-0.057***
	(0.010)	(0.011)	(0.008)	(0.010)
Observations	13,556	13,556	13,556	13,556
# of Schools	2313	2313	2313	2313
R-squared	0.866	0.774	0.870	0.758
Year FE	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes
Municipality controls	Yes	Yes	Yes	Yes
pvalue $a + b = 0$	0.768	0.000	0.014	0.000

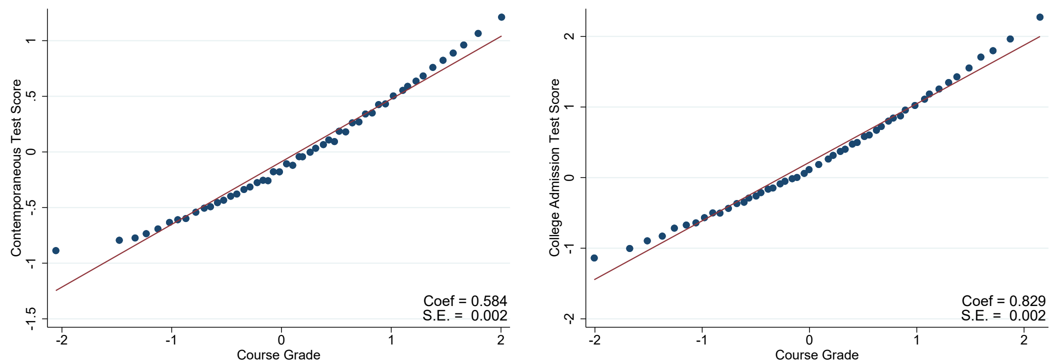
Notes: This table presents the effects of the new selection system (ADP) on college admission scores. "ADP" is a dummy that takes the value one after the first time a school selects a principal under the ADP system. "Principal Turnover" is a dummy that takes the value one after the first time a school selects a new principal (after 2012). Column 1 shows the results for Math test scores, column 2 shows them for Spanish test scores, column 3 for the average of both, and column 4 for the composite score used for admissions. This score is a weighted average of entry exam scores and course grades, with weights defined by each degree (institution-major pair). We consider the weights of the most preferred degree of a student (as revealed by her preferences in the application process) to construct this score. All columns include 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. Robust standard errors clustered at school level in parenthesis.

APPENDIX (For Online Publication)

Managers' Productivity and Recruitment in the Public Sector: The case of school principals

A Additional Figures and Tables

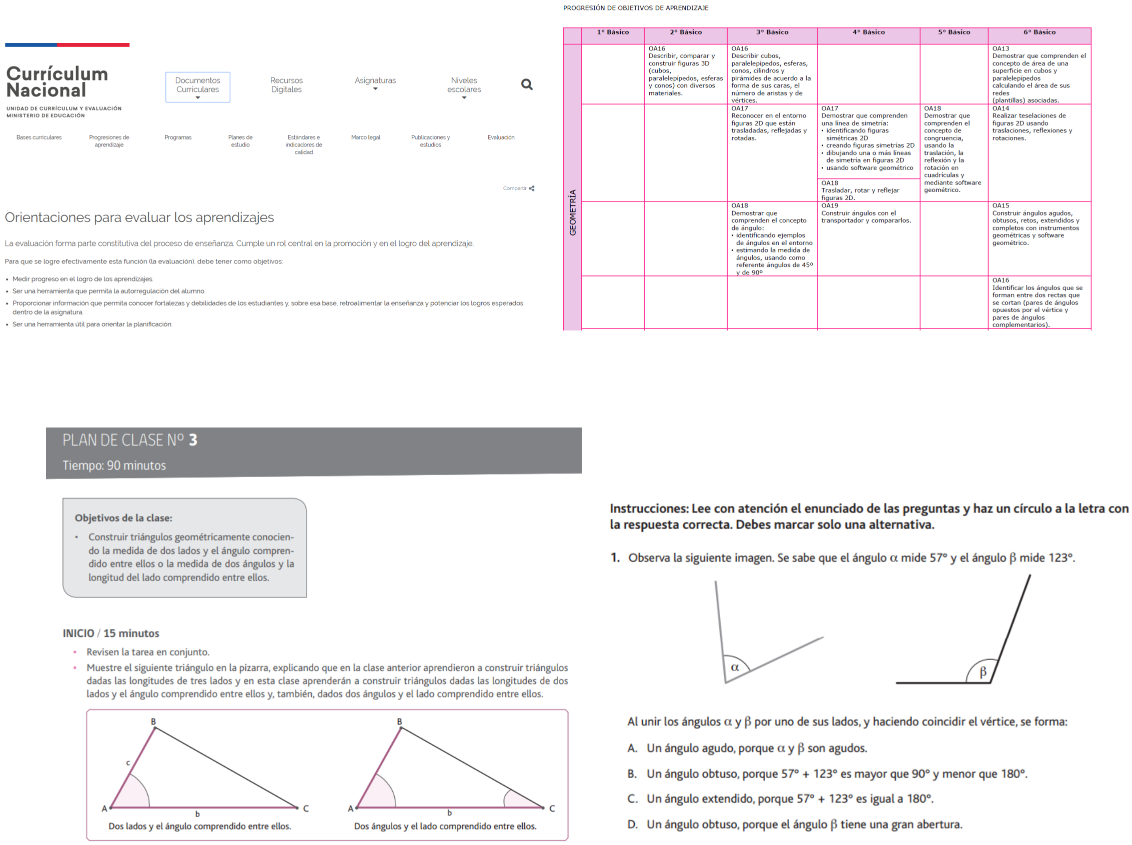
Figure A.1: Course Grades and Test Scores



A. SIMCE **B.** PSU

Notes: Panel A considers a sample of 1,061,231 students for whom we observe test score and course grades contemporaneously for Math and Spanish (between 2011 and 2016). Panel B considers a sample of 132,585 students accepted into college and for whom we can compute college admission scores (in the 2017 process). The college admission score is an institution-major specific weighted average of applicants' high-school course grades and entrance exam scores. We report the coefficient and robust standard error from a linear regression of test score on course grades.

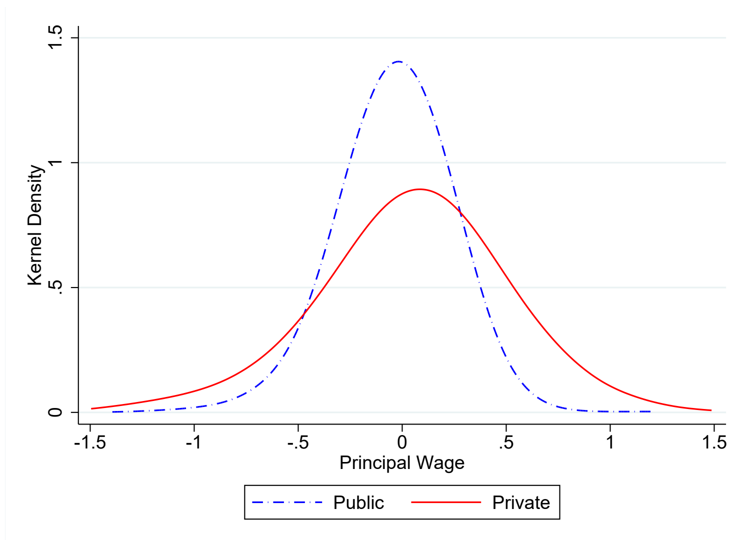
Figure A.2: Standardized National Curriculum



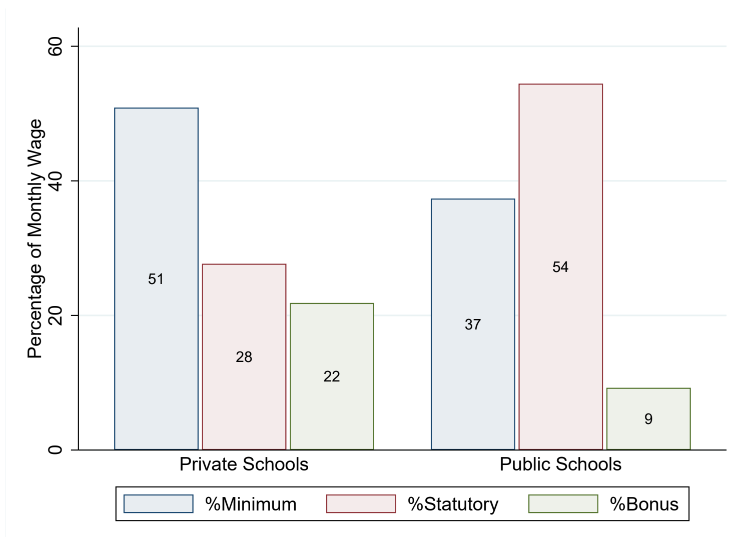
	1° Básico	2° Básico	3° Básico	4° Básico	5° Básico	6° Básico
GEOMETRÍA		OA16 Describir, comparar y construir figuras 3D (cubos, paralelepípedos, esferas y conos) con diversos materiales.	OA16 Describir cubos, paralelepípedos, esferas, conos, cilindros y pirámides de acuerdo a la forma de sus caras, el número de aristas y de vértices.			OA13 Demostrar que comprenden el concepto de área de una superficie en cubos y paralelepípedos calculando el área de sus redes (plantillas) asociadas.
			OA17 Reconocer en el entorno figuras 2D que están trasladadas, reflejadas y rotadas.	OA17 Demostrar que comprenden una línea de simetría: • identificando figuras simétricas 2D • creando figuras simetrías 2D • dibujando una o más líneas de simetría en figuras 2D • usando software geométrico OA18 Trasladar, rotar y reflejar figuras 2D.	OA18 Demostrar que comprenden el concepto de congruencia, usando la traslación, la reflexión y la rotación en cuadrículas y mediante software geométrico.	OA14 Realizar teselaciones de figuras 2D usando traslaciones, reflexiones y rotaciones.
			OA18 Demostrar que comprenden el concepto de ángulo: • identificando ejemplos de ángulos en el entorno • estimando la medida de ángulos, usando como referente ángulos de 45º y de 90º	OA19 Construir ángulos con el transportador y compararlos.		OA15 Construir ángulos agudos, obtusos, retos, extendidos y completos con instrumentos geométricas y software geométrico.
						OA16 Identificar los ángulos que se forman entre dos rectas que se cortan (pares de ángulos opuestos por el vértice y pares de ángulos complementarios).

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 A.3: Principals' Wages



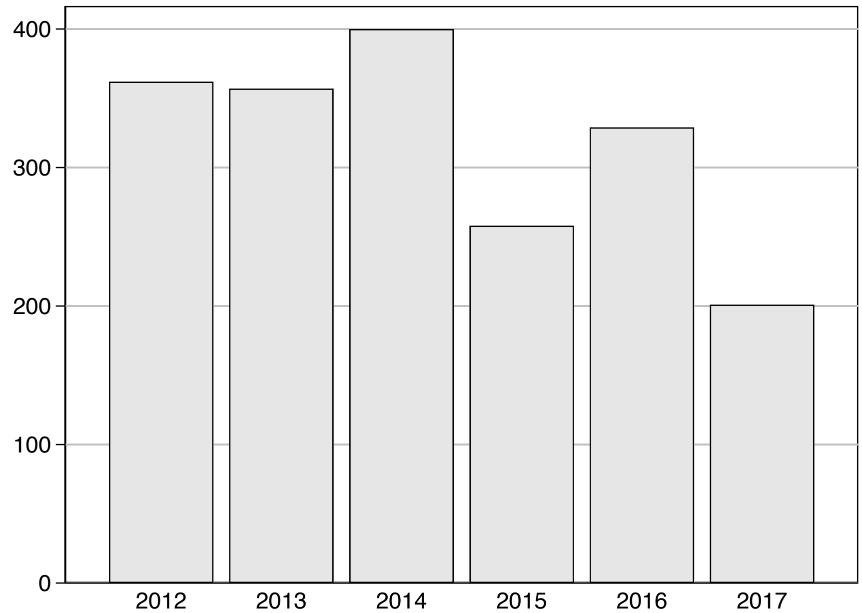
A. Residualized Log Wage



B. Wage Components

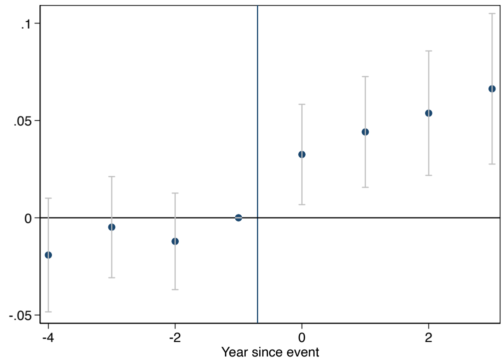
Notes: Panel A presents the distributions of log principals' wages in both public and subsidized private schools. Log principals' wages are residualized with respect to year and municipality fixed effects. Panel B decomposes the average monthly wage of school principals into the three components discussed 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 A.4: Number of Newly Elected Principals by ADP, per Year

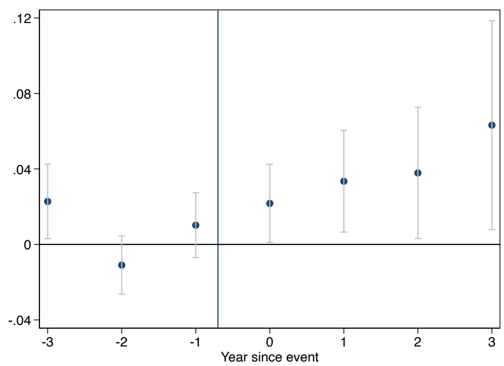


Notes: This figure shows the number of schools that elected a principal through the new ADP selection system for the first time, by year.

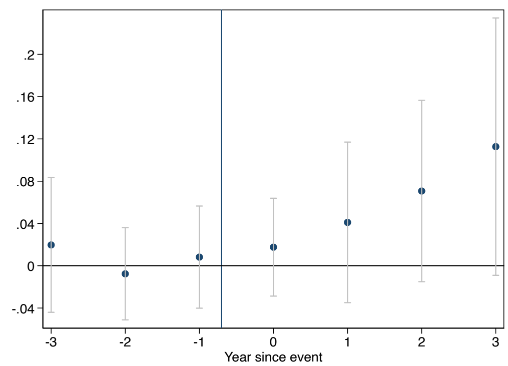
Figure A.5: Principal Selection and Principal Effectiveness within Public Schools



A. Two-way fixed effect



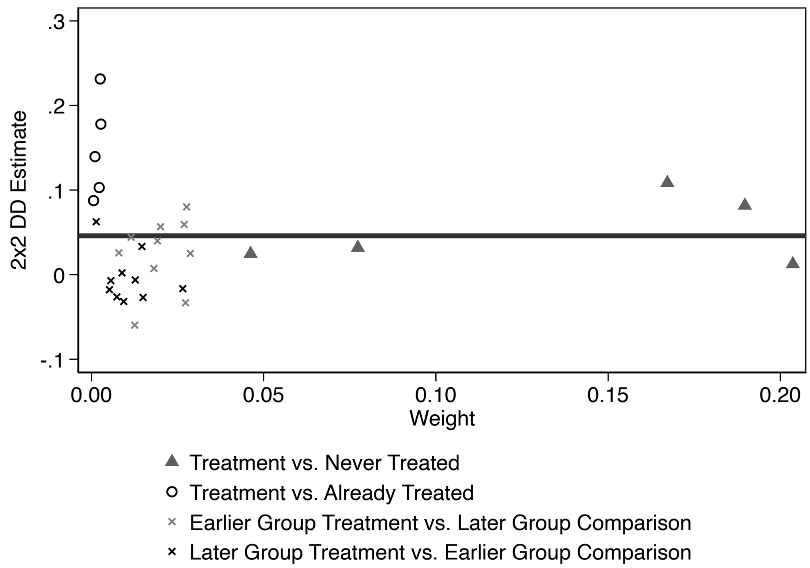
B. De Chaisemartin and d'Haultfoeuille (2020)



C. Callaway and Sant'Anna (2020)

Notes: This figure presents the dynamic version of our *staggered difference-in-differences* approach in the sample of public schools. Panel A presents the estimates from a version of equation (5) for the sample of public schools. Panel B presents the dynamic version of the staggered difference-in-differences model suggested by De Chaisemartin and d'Haultfoeuille (2020). Panel C presents the dynamic version of the staggered difference-in-differences suggested by Callaway and Sant'Anna (2020). All panels include confidence intervals at the 95%. In panels A and C we cannot reject the null hypothesis of all the coefficients being equal to zero at conventional levels (in the pre-period). The p-value of this test is > 0.09 in Panel B.

Figure A.6: Goodman-Bacon (2018) Decomposition



Notes: This figure presents the decomposition of the two-way fixed effect estimator suggested by Goodman-Bacon (2018).

Table A.1: Descriptive Statistics in Different Samples

	Full Sample		**Δ Teacher=1**		**LCS=1**	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
	(1)	(2)	(3)	(4)	(5)	(6)
Primary (2-8)	0.82	0.39	0.83	0.38	0.90	0.30
Secondary (9-11)	0.18	0.39	0.17	0.38	0.10	0.30
Subject = Math	0.50	0.50	0.50	0.50	0.50	0.50
Course Grade	5.66	0.63	5.66	0.63	5.69	0.62
% Attendance	92.21	7.21	92.03	7.39	92.17	7.20
% Rural School	0.09	0.29	0.08	0.28	0.07	0.26
% Public School	0.38	0.48	0.36	0.48	0.33	0.47
School Size	794.30	596.63	825.96	618.68	847.43	624.29
Sample Size	12,709,699		9,120,301		7,735,683	

Notes: This table presents descriptive statistics of students in three different samples. "Full Sample" includes all students in our dataset after excluding 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. "ΔTeacher = 1" corresponds to the restricted sample of students for whom the teacher, in a given subject, changed between t and $t+1$. Finally, "LCS" includes all students within the largest connected of teachers and principals.

Table A.2: Manager Effectiveness and Observable Characteristics

	Principal Effectiveness $\hat{\theta}_p$			
	All		Public	Private
	(1)	(2)	(3)	(4)
Age	0.010***	0.010***	0.030***	-0.000
	(0.003)	(0.003)	(0.008)	(0.004)
Age^2	-0.000***	-0.000***	-0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.052***	0.052***	0.069***	0.044***
	(0.009)	(0.010)	(0.015)	(0.014)
Perm. Contract	0.017	0.020	-0.004	0.038
	(0.017)	(0.019)	(0.028)	(0.025)
Hours Contract	-0.004***	-0.003**	-0.008*	-0.003**
	(0.001)	(0.001)	(0.005)	(0.001)
College Degree	-0.015	-0.034**	0.018	-0.097***
	(0.015)	(0.016)	(0.023)	(0.022)
Ever Teacher		0.019	-0.008	0.030*
		(0.012)	(0.017)	(0.016)
Ever Admin. Supp. Worker		-0.016	-0.040**	-0.022
		(0.013)	(0.017)	(0.018)
Ever Admin. Worker		-0.035	0.043	-0.031
		(0.033)	(0.056)	(0.040)
Observations	42,013	35,333	15,829	19,497

Notes: This table presents the correlation between the principal effectiveness estimated from equation (1) and principal characteristics. These characteristics include age, gender, experience, type and hours of contract, and indicators for holding a college degree, and for their experience in previous "schooling type" of positions. All specifications include year and municipality fixed effects. Robust standard errors in parentheses.

Table A.3: School and Municipality Characteristics, by ADP Adoption

	Never ADP	Ever ADP	Difference	Early ADP	Late ADP	Difference	Private Schools
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: School characteristics							
Total Enrollment	100.330	454.654	354.325***	489.115	433.365	-55.750***	396.351
	(186.517)	(351.519)	(0.000)	(352.286)	(349.506)	(0.001)	(474.969)
Rural School	0.838	0.256	-0.582***	0.224	0.276	0.051**	0.204
	(0.369)	(0.437)	(0.000)	(0.418)	(0.447)	(0.015)	(0.403)
Income per student	19.631	6.455	-13.175***	6.471	6.446	-0.025	10.198
	(25.551)	(1.859)	(0.000)	(1.798)	(1.896)	(0.784)	(15.485)
Share of disadvantaged students	0.702	0.508	-0.193***	0.493	0.518	0.026**	0.369
	(0.220)	(0.225)	(0.000)	(0.240)	(0.214)	(0.018)	(0.321)
4rd grade test scores (Spanish)	255.565	255.610	0.045	256.303	255.203	-1.100	266.919
	(30.287)	(22.027)	(0.961)	(21.944)	(22.077)	(0.344)	(27.432)
4rd grade test scores (Math)	238.975	245.459	6.484***	245.793	245.262	-0.531	255.535
	(33.034)	(24.510)	(0.000)	(24.177)	(24.715)	(0.682)	(33.246)
Graduation test score (Spanish)	414.049	436.469	22.420***	439.418	434.231	-5.188	514.169
	(43.775)	(57.979)	(0.000)	(55.863)	(59.561)	(0.380)	(72.710)
Graduation test score (Math)	418.480	441.136	22.657***	441.347	440.977	-0.370	518.640
	(41.965)	(55.204)	(0.000)	(54.580)	(55.794)	(0.948)	(78.207)
Panel B: Municipality characteristics							
Share of households in poverty	0.124	0.082	-0.042***	0.082	0.082	-0.000	0.059
	(0.075)	(0.056)	(0.000)	(0.057)	(0.055)	(0.952)	(0.061)
Income per capita	1.699	2.151	0.453***	2.223	2.107	-0.115**	3.216
	(0.489)	(1.115)	(0.000)	(1.400)	(0.892)	(0.033)	(2.189)
Unemployment rate	0.080	0.080	0.001	0.083	0.079	-0.004*	0.080
	(0.047)	(0.047)	(0.626)	(0.050)	(0.045)	(0.064)	(0.032)
Average years of schooling	8.974	9.998	1.024***	9.930	10.041	0.110*	10.785
	(1.124)	(1.315)	(0.000)	(1.385)	(1.269)	(0.083)	(1.554)
Observations	3,029	1,820	4,849	695	1,125	1,820	6,415

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. Finally, column 7 presents summary statistics for all private schools.

Table A.4: Characteristics of Origin and Destination Schools of ADP principals

	School of Origin	School of Destination	Mean Difference
	(1)	(2)	(3)
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).

Table A.5: Principal Selection and Principal Effectiveness: Placebos

	Public Schools		Private Schools	
	(1)	(2)	(3)	(4)
Principal Turnover	-0.043	0.019	-0.032*	-0.026
	(0.111)	(0.116)	(0.019)	(0.039)
Observations	5,303	5,303	17,498	17,498
# of Schools	1666	1666	2802	2802
Year FE	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes

Notes: This table presents the estimate from our placebo exercise looking at the impact of non-ADP principal turnovers on the standardized measure of principal effectiveness discussed in section 3. In columns 1 and 2, "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 292. In columns 3 and 4, "Principal turnover" is a dummy that takes the value one after the first time a private school selects a new principal (after 2012). Columns 1 and 3 show the estimates from the model suggested by De Chaisemartin and d'Haultfoeuille (2020), while columns 2 and 4 show the estimates from the model suggested by Callaway and Sant'Anna (2020). Robust standard errors clustered at school level in parentheses.

B 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).[20] 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:

$$U_j(i) = m_j + \beta_j \theta_i + \epsilon_{1ij}, \tag{6}$$

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

$$U_j(\neg i) = s_j + \epsilon_{0ij}, \tag{7}$$

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.

[20] 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.

When expression (6) is greater in value than expression (7), employer j makes a job available: $o_{ij} = 1$, zero otherwise. Thus, the exact probability that school j will make an offer depends on the distribution of the differences between the two error terms, as well as on the non-stochastic parts of (6) and (7). 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:

$$Pr(o_{ij}) = \frac{\exp(\beta_{0j} + \beta_j \theta_i)}{1 + \exp(\beta_{0j} + \beta_j \theta_i)}, \tag{8}$$

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:

$$Pr(S_{ik}) = \prod_{m \in O_k} Pr(O_{im} = 1) \prod_{n \in \bar{O}_k} Pr(O_{in} = 0), \tag{9}$$

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:

$$V_{i(j)} = h_j + w_j \theta_i + v_{ij}, \tag{10}$$

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:

$$Pr(A_{ij} \mid O_k) = \begin{cases} \frac{\exp(h_j + w_j\theta_i)}{\sum_{h\in O_k}\exp(h_h + w_h\theta_i)} & , \quad j \in O_k \\ 0 & , \quad j \notin O_k. \end{cases} \tag{11}$$

Given our assumptions about the distribution of the random components in (6), (7), and (10), 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:

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

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:

$$
\begin{aligned}
Pr(A_{ip} \mid \theta_i) = & \left(\frac{\exp(h_p + w_p\theta_i)}{\exp(h_p + w_p\theta_i) + \exp(h_v + w_v\theta_i)} \right. \\
& \left. \times \frac{\exp(\beta_{0p} + \beta_p\theta_i)}{1 + \exp(\beta_{0p} + \beta_p\theta_i)} \times \frac{\exp(\beta_{0v} + \beta_v\theta_i)}{1 + \exp(\beta_{0v} + \beta_v\theta_i)} \right) \\
& + \left(1 \times \frac{\exp(\beta_{0p} + \beta_p\theta_i)}{1 + \exp(\beta_{0p} + \beta_p\theta_i)} \times \frac{1}{1 + \exp(\beta_{0v} + \beta_v\theta_i)} \right). \qquad (12)
\end{aligned}
$$

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

$$
\mathrm{E}[\theta_i \mid \text{Public}] = \int_\theta \theta_i f_{\theta|\text{Public}}(\theta_i \mid \text{Public}) d\theta \qquad (13)
$$

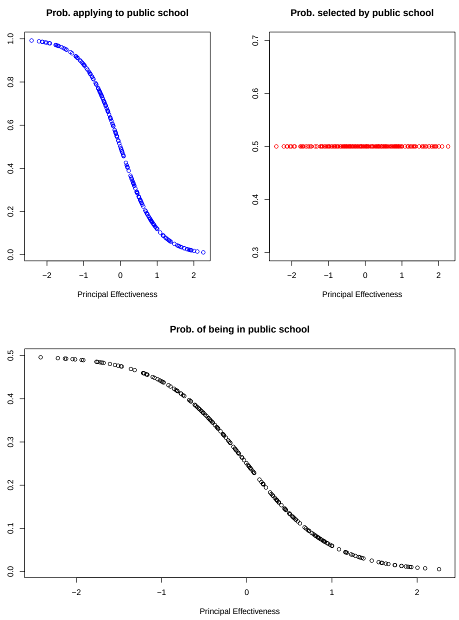
From Bayes' rule, we know that:

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

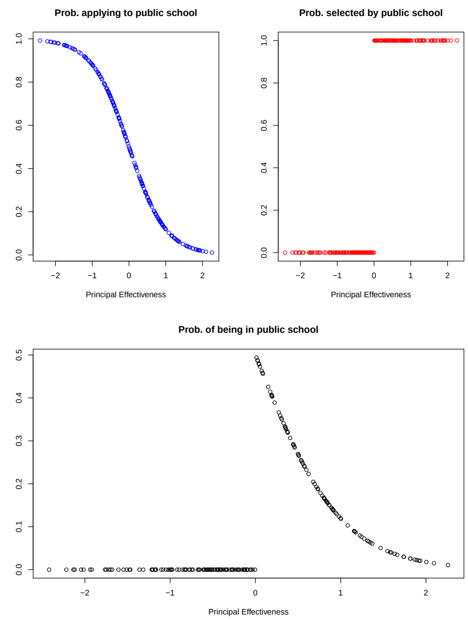
where $Pr(A_{ip} \mid \theta_i)$ is given by (12) and $Pr(\text{Public})$ is a scale factor equal to the fraction of public schools (0.5 in this case). Assuming that $f_\theta(\theta_i)$ is a standard normal, we can compute $\mathrm{E}[\theta_i \mid \text{Public}]$ using numerical integration. More importantly, we can study how this object depends on β_p and w_p, the two relevant parameters related to selection and payment policies in public schools, respectively.

Our simulation is presented in Figure 1. Panel A, B, and C consider different personnel selection rules. Panel A shows a case where personnel selection is independent of worker quality. Panel B shows a case where a worker is selected if and only if her quality is above some threshold. Panel C shows the case where the likelihood of selecting a worker is increasing in proportion to her quality. Finally, Panel D shows the allocation of principal effectiveness given by equation (13). To construct this figure, we created a grid for β_p and w_p from 1 to 10, and compute $\mathrm{E}[\theta_p \mid \text{school type: Public}]$ for each cell of this grid.

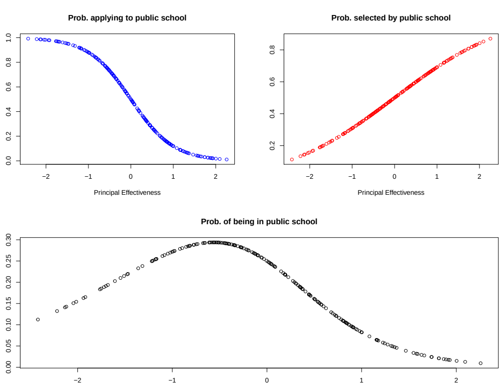
Figure 1: Simulation of a two-sided matching model



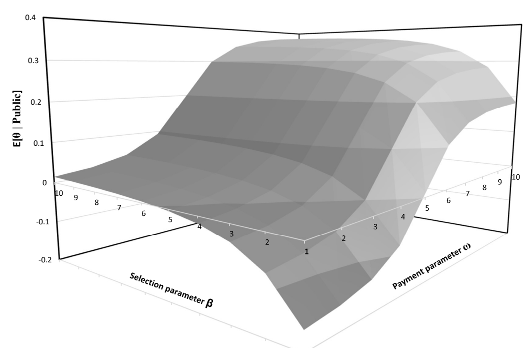
A. Selection at Random



B. Selection Above Cutoff



C. Selection on Quality



D. Allocation of θ_p if Selection on Quality

Notes: Panel A, B, and C show simple simulations that exemplify how personnel selection rules can counteract the self-selection component of labor markets. For this, we assume that the idiosyncratic preferences of principals follow a type I extreme value distribution, that principals do not anticipate the schools' selection rule, and that private schools have a larger pay-for-performance component ω than public schools. Panel D shows the allocation of principal effectiveness as a function of the selection and payment parameters. To construct this figure, we created a grid for β_p and w_p from 1 to 10, and compute E[θ_p | school type: Public] for each cell of this grid.

C Data Files

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.

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. We complement these data with records from the centralized admission system. Specifically, we add the average (at the school level) of the students' composite score used for college admission and the average (at the school level) of the students' score in the college entrance exams of Math and Spanish.

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.

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.

References

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

Callaway, B. and P. H. Sant'Anna (2020): "Difference-in-differences with multiple time periods," *Journal of Econometrics.*

De Chaisemartin, C. and X. d'Haultfoeuille (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 110, 2964–96.

Goodman-Bacon, A. (2018): "Difference-in-differences with variation in treatment timing," Tech. rep., National Bureau of Economic Research.

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

Roth, A. E. and M. Sotomayor (1990): "A study in game-theoretic modeling and analysis," *Econometric Society Monographs*, 18.