WORKING PAPERS

N° 131

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

Pablo Munoz and Mounu Prem

IAST
INSTITUTE for ADVANCED STUDY in TOULOUSE

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

Pablo Muñoz  Mounu Prem

**Abstract**

Governments face many constraints in attracting talented managers to the public sector, where high-powered incentives are often absent. In this paper, we study how a civil service reform in Chile changed the effectiveness of a vital group of public sector managers, school principals. We measure principals' effectiveness using an extension of the canonical teacher value-added model and we evaluate the effect of the reform using a difference-in-differences approach. We find that public schools appoint more effective managers after increasing the competitiveness and transparency of their selection process. Our result shows that better recruitment policies can enhance service provision in the public sector, despite rigid wage schemes.

KEYWORDS: Managers, Public sector, Recruitment, School principals

*This version: February, 2022. 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 Political Economy seminars, the World Bank Research Group, the CAF Development Bank of Latin America, the 2nd Bank of Italy Human Capital workshop, the CEA - UChile economics seminar, NEUDC, the PUC applied microeconomics seminar, and the U. of Edinburgh economics seminar, for their comments and 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. Prem acknowledges IAST funding from the French National Research Agency (ANR) under the grant ANR-17-EURE-0010 (Investissements d'Avenir program). José Niño and Luis Serrano provided excellent research assistance. Pablo Muñoz, FGV EPGE Brazilian School of Economics and Finance, Brazil. Email: pablo.munoz@fgv.br. Mounu Prem, Universidad del Rosario, Department of Economics, Colombia. Email: francisco.munoz@urosario.edu.co

# 1 Introduction

Management is a key resource of both private and public enterprises (Bloom et al., 2013, 2015), 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 (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). Empirical progress in this area has faced at least two important hurdles. First, the dearth of data makes it difficult to objectively measure managers' performance in the public sector. Second, it is hard to find quasi-experimental variation in the allocation of public sector positions.

In this article, we overcome previous limitations by focusing on "street-level" bureaucrats (Besley et al., 2021) and by leveraging the institutional setting and the rich administrative data of Chile. We study the case of school principals, managers of high practical relevance for the delivery of public services, and thus for state capacity. Studying managers in the Chilean educational context is appealing for three main reasons. First, we can use administrative data on student performance in a "value added" framework to construct an objective measure of principals' effectiveness. Second, the largely publicly funded but privately run school system in Chile (Hsieh and Urquiola, 2006; Epple et al., 2017) provides a natural benchmark for evaluating the public system. Third, and most importantly, variation from a civil service reform allows us to assess the impact of more competitive and transparent personnel selection policies on the effectiveness of public schools' managers.

Our setting offers a clear testing ground to evaluate Max Weber's hypothesis that the separation of bureaucracy and politics leads to a more effective public service (Weber, 1922). 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 students' outcomes (Akhtari et al., 2020). 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' appointments 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, 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.

In order to assess the impact of this reform, we first develop a novel extension of the teacher value-added model to disentangle the effectiveness 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 effectiveness. Then, we use a difference-in-differences approach to compare the change in principal effectiveness arising from a new appointment under the reformed selection system to the change in principal effectiveness arising from a new appointment at private schools. We find that, despite having rigid wage schemes, public schools were able to attract more effective managers (a 0.06 standard deviation increase) after changing their selection process. We find similar results when we only keep public schools for estimation, therefore identifying the effect of the reform from variation in the timing of adoption of the new selection system.

A key feature of our analysis is the use of rich administrative data to estimate an output-based measure of principals' effectiveness. Our 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 standardized tests every year, which prevents us from using test scores to measure teachers' and principals' effectiveness. Instead, we use students' course grades in Mathematics and Spanish. In the Chilean setting—with a national curriculum—course grades are both highly relevant and informative. Indeed, course grades determine both grade retention and high school graduation, and are also an important determinant of college admission and access to financial aid (Hastings et al., 2013).[1] Moreover, course grades are strongly correlated with contemporaneous standardized test scores (typically administered in fourth and eighth grades), supporting the view of using them interchangeably (Borghans et al., 2016);[2] and they are also known to capture non-cognitive

[1]In 2017, course grades and contextual course grades had an average weighting of 40% in the college admission score. As shown by Panel A of Figure A.1, a one standard deviation increase in course grades was associated with a 0.8 standard deviation increase in college admission scores.

[2]Panel B of Figure A.1 shows a strong relationship between test scores and course grades. In a sample of students for whom we observe SIMCE test scores and course grades contemporaneously for Math and

traits that play an essential role in academic and labor market performance (Bowles and Gintis, 1976; Bowles et al., 2001; Heckman et al., 2006; Jackson, 2018).[3]

Accommodating principals' effects into the canonical teacher value-added model is challenging. On the one hand, it is necessary to distinguish principals' effectiveness from that of their teaching staff. On the other hand, it is necessary to disentangle the effect of the principal from other school-level factors. To tackle the first challenge, we follow the seminal work of Abowd et al. (1999), and estimate a two-way fixed effects model that leverages teachers' and students' switches across principals (*within* the largest connected set) to separately identify principals' and teachers' effectiveness. To address the second problem, we follow the seminal work of 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 allows us to obtain an estimate of effectiveness for all school principals, while avoiding problems of weak identification that could arise if we added school fixed effects in this setting (Jochmans and Weidner, 2019). Our estimates of principal effectiveness, adjusted by its "reliability" (Krueger and Summers, 1988; Aaronson et al., 2007), imply that a one standard deviation increase in principal effectiveness raises students' course grades by 0.29 standard deviations.

We perform several exercises to show that our objective measure of principal effectiveness is sensible. First, in the spirit of Chetty et al. (2014) and Angrist et al. (2017), we use event studies of principals' arrivals and departures, and show that student achievement changes sharply after event time *as predicted* by our measure of principal effectiveness. Second, since our model considers additive teacher and principal effects, we perform a specification check to show that our results are consistent with the symmetry implications of an additive two-way fixed effects model with exogenous mobility (Card et al., 2013). Third, we leverage soft data from teachers' surveys and show that more effective principals are associated with a larger fraction of their teaching staff highly agreeing with positive statements about them. Finally, we also study *how* principals matter. Using administrative data from different sources we show that principals' effectiveness is associated with fewer parents' complaints about bullying and denying enrollment, and with lower rates of teachers' turnover, especially among high value-added teachers.

Spanish, we find that a one standard deviation increase in course grades is associated with a 0.6 standard deviation increase in test scores.

[3] A concern about grades, however, is that some teachers may have an incentive to inflate their grades. We address this issue following Peteck and Pope (2019) and relating exposure to a teacher at time $t$ to the change in course grades from the past year to the year ahead (i.e., $\text{GPA}_{t+1} - \text{GPA}_{t-1}$).

After validating our measure of principal effectiveness, we focus on the labor market of school principals. A descriptive analysis reveals 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 weaker in public than in private schools. With little room for rewarding performance, the public sector must rely on alternative strategies to attract and retain effective workers (Khan et al., 2019; Bertrand et al., 2020). In this paper, we focus on one of such strategies: personnel selection. To motivate our analysis, we develop a two-sided selection model which highlights that higher wages might not suffice nor be the only relevant variable if workers' choice also depends on their idiosyncratic preferences and the labor demand that they face.[4] An important take away from the model is that poor selection policies may cause public schools to lose talented managers who would take a position despite a lower payment because of strong idiosyncratic preferences for the public sector (e.g., pro-social behavior).

Although identifying and curating talent is of theoretical and practical relevance for the public sector, evidence on this issue remains elusive. To study this empirically, we implement a difference-in-differences research design that compares the change in principal effectiveness arising from a new appointment under the reformed selection system to the change in principal effectiveness arising from a new appointment at private schools. We show that limiting the discretion of local politicians over the appointment of public school managers increases principal effectiveness by 0.06 standard deviations. Reassuringly, the dynamic version of our difference-in-differences approach provides visual support to our identification strategy; and we further complement this using a sensitivity test for violations of the parallel trends assumption (Rambachan and Roth, 2021) and a correction for potential bias coming from pre-treatment differential trends and pre-testing (Roth, 2021).

A plausible concern is that other changes brought about by the 2011 reform could confound our previous estimates comparing public to private schools. Reassuringly, we find similar results when we only keep public schools for estimation, and therefore we identify the effect of the reform from variation in the timing of adoption; a result that is robust to the recent developments in the literature of staggered difference-in-differences models under the presence of heterogeneous treatment effects (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2020). Moreover, placebo exercises looking at principal turnover

[4] As underscored by economists and sociologists in the past (Abowd and Farber, 1982; Logan, 1996), when choice is constrained by opportunity, personnel selection can accentuate or counteract the sorting of workers based on wages.

at private schools during the *post*-reform period as well as principal turnover at public schools during the *pre*-reform period fail to detect positive effects on principal effectiveness. In like manner, an event study around the time of new appointments of school principals shows that—before the civil service reform was enacted—principal turnover did *not* affect principal effectiveness deferentially among public and private schools.

We conclude with a brief discussion on the impacts of the reform on school personnel, students' long-run outcomes, and equity. Our analysis suggests that the appointment of principals elected under the new selection system increased the likelihood of firing poorly evaluated teachers and it also increased students' college admissions test scores. In light of existing research documenting positive sorting induced by admission scores into enrollment at more selective institutions (Rodríguez et al., 2016) and the positive returns associated with more selective degrees in Chile (Hastings et al., 2013; Zimmerman, 2019), we conjecture positive long-run effects on students' outcomes. Finally, in terms of equity, a back-of-the-envelope calculation suggests that the impact of the reform on principals' effectiveness was enough to reduce the public-private course grade gap by half in 5 years.

## 1.1 Related literature

Our paper contributes to different branches of economic research. First, it contributes to the literature on state capacity and personnel economics (Lazear and Shaw, 2007; Finan et al., 2015; Besley et al., 2021). Recent studies have shown that patronage is a common feature in public sector appointments (Xu, 2018; Akhtari et al., 2020; Colonnelli et al., 2020; Voth and Xu, 2020), with mixed findings regarding its effects on state capacity. Likewise, evidence on the role of performance-based hiring is still limited (Ornaghi, 2019; Scot et al., 2021; Moreira and Pérez, 2021). More broadly, our paper relates to research showing that different selection policies and incentive methods can attract different types of public sector workers (Ferraz and Finan, 2011; Dal Bó et al., 2013; Finan et al., 2017; Deserranno, 2019; Deserranno et al., 2019; Ashraf et al., 2020). Closer 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. We contribute to this literature by focusing on managers and by showing that a more transparent selection system based on third-party screening leads to the appointment of more effective workers.

Second, our work speaks to a growing literature on the importance of management and

management practices in the private sector (Bloom and Van Reenen, 2007; Bloom et al., 2013; Bender et al., 2018) and public sector organizations (McCormack et al., 2014; Bloom et al., 2015; Lavy and Boiko, 2017; Rasul and Rogger, 2018). Closer to our study, Fenizia (2021) uses an objective measure of managers' performance and provides sound and novel evidence on the importance of managers in the Italian bureaucracy. Like her, we construct an objective measure of managers' effectiveness (albeit in a different context). In contrast to her, we are able to use quasi-experimental variation in the allocation of managers' positions.

Third, our paper also contributes to the economics of education literature. Related research on school principals has focused on the effect of principal attributes on students' performance (Eberts and Stone, 1988; Clark et al., 2009; Béteille et al., 2012) or on the measurement of principal value-added in isolation (Branch et al., 2012; Coelli and Green, 2012; Dhuey and Smith, 2014; Grissom et al., 2015). We add to this literature by providing an estimation framework that accounts for key inputs in the education production function to disentangle principal effectiveness. In this regard, our paper also adds to the literature on value-added models (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). Finally, our work complements studies on school personnel (Rothstein, 2015; Biasi, 2021; Loyalka et al., 2019; Brown and Andrabi, 2020; Leaver et al., 2021) and the labor market of school principals (Cullen et al., 2016). In contrast to previous research in this area, we study a setting where high-powered incentives are hard to implement, and thus our focus is on personnel selection instead of pecuniary incentives.

# 2 Background and Data

This section describes the main educational reforms implemented in Chile since the eighties, with a special focus on the 2011 reform. It also describes the data used in our analysis and presents some descriptive statistics.

## 2.1 Background

In 1981, under a dictatorship, Chile implemented an educational reform that privatized and decentralized primary and secondary education. Publicly funded school vouchers were

created with flat voucher funds following any children either to public schools or to the private schools that agreed to accept the voucher as payment of tuition.[5] 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. 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.[6] 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.

Alongside with 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. 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 individually negotiated with the schools.

**The "Quality and Equity in Education" Reform:** In 2011, the country enacted a law aimed at improving quality and equity in education (Law N. 20.501). The law recognized school principals as key agents to improve quality in public schools and created a new system to appoint them. Before the reform, the appointment of public school principals

[5] 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).

[6] 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., 2020).

was the exclusive responsibility of the municipalities; this process was unsupervised by the central government and consequently, it was prone to patronage. After the reform, principals are elected through public, competitive, and transparent contests.

The new selection process is overseen by the Civil Service,[7] but municipalities are still responsible for the processes. The contests must be disseminated on newspapers, the website of the municipalities, and the website of the Civil Service *www.directoresparachile.cl*. As shown by Appendix Figure A.2, calls are widely advertised and information on open, ongoing, and finished contests is publicly available to all potential candidates. After having received the background information and having defined the admissibility of the applicants, an external human resources company is hired to perform the pre-selection of candidates. This process includes a curricular analysis and psycho-labor evaluations. The external company specialized in the selection of candidates is chosen by the Civil Service from a pool of registered companies. Then, an independent qualifying commission, integrated by a representative of the Civil Service, the head of education of the municipality, and an outstanding teacher (chosen by lottery) conducts the interviews with the candidates shortlisted by the human resources company. After these interviews, the qualifying commission defines a list of 3 to 5 candidates; and this list is then sent to the mayor who makes the final decision.

The duration of a contest, from the announcement to the definition of the short-list of candidates takes approximately 100 days. Each contest must comply with the principles of i) *non-discrimination:* the process cannot discriminate based on sex, age, religion, politics, sexual orientation, gender identity, civil status, disability, or others. The selection must only consider merit, training, experience, and skills required for good job performance; ii) *confidentiality:* the actors involved in the selection process must keep the identity of the candidates confidential, and all data must be protected under legal regulations; iii) *public-private participation:* external consultancies from expert companies or natural persons, registered with the Civil Service, are part of the process.[8]

The adoption of the new selection system was staggered over time. The authorities established that new appointments would be required to go through the new system once the director who was in office (as of 2011) had completed a period of five years (Ruiz-Tagle,

[7] 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."

[8] For more details, see Silva (2014) and "Alta Dirección Pública y Reforma Educacional" (available at: https://www.serviciocivil.cl/sistema-de-alta-direccion-publica-2/adp-educacion/).

2019). However, no clear sanctions were established in the event that this provision was not complied with (Errázuriz et al., 2016). Moreover, the Civil Service and the municipalities had to incur costs associated with running the contests (Silva, 2014), and not all processes concluded with an appointment.[9]

The reform also established that public school principals i) would be 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; ii) can fire up to 5% of teachers with a bad or regular teacher evaluation; and finally, iii) they get a bonus consistent with their responsibilities, 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. The reform included other measures as well, such as a new retirement plan for teachers, bonuses for teachers with good evaluations, adjustments to the severance payments, the introduction of public contests for the position of educational superintendent, more rights for teachers and teacher assistants, and more funding for both public and subsidized private schools. For more details about the reform, see "Ley 20.501 Calidad Y Equidad de la Educación."

**The school system:** Like many countries, Chile has a nationwide standardized curricula.[10] The Chilean curriculum 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.4 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

[9] According to the records from the Civil Service, of the 4,305 competitions called between 2011 and 2017, 63.7% resulted in appointments, 23.4% were resolved as deserts, and 11.3 % were overridden. The rest did not conclude with an appointment for admissibility problems or because it was not possible to carry out the external evaluation stage (Ruiz-Tagle, 2019).

[10] Countries that have a national curriculum include: France, Hungary, Ireland, Italy, Japan, Korea, the Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, and the United Kingdom. While most states in the U.S. follow common guidelines for a core curriculum, there is no national curriculum as such.

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). 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.

## 2.2 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 and Spanish.[11] We match this data set with a nationwide census of teachers containing rich information on the specific subjects and classrooms taught by them every year, as well as their characteristics (e.g., gender, age, type of degree, 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 the 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.

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

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 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 model used to measure principals' effectiveness and we document the importance of school principals for students' outcomes. We also perform several exercises to validate our model, and study what principals do within schools.

We consider a specification that relates academic achievement to student characteristics, school characteristics, and to the teachers and school principal, as follows:

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

where $Y_{it+1}$ is the course grade obtained by student $i$ in year $t+1$, $\gamma_t$ and $\rho_g$ stand for year and grade fixed effects, and $f(Y_{it-1}, \bar{Y}_{it-1}, \rho_{g(i,t)})$ is a standard third degree polynomial at the student and classroom level in the lagged dependent variable interacted with students' grade level (Kane and Staiger, 2008; Chetty et al., 2014). Importantly, we 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 $t+1$. We do so, to remove systematic bias from teachers evaluating their own students (Peteck and Pope, 2019).[12] We also control by students' age, principal's tenure $X_{it}$, and a set of time-varying and across-time averages of school-level characteristics $X_{s(i,t)t}$ and $\bar{X}_{s(i,t)t}$.

Since we only observe a subset of principals switching between schools, we do not include school fixed effects. Instead, we use correlated random effects (Mundlak, 1978; Chamberlain, 1980) to account for school heterogeneity. Specifically, we include a combination time-varying and across-time averaged characteristics of each school $s$ (i.e., $X_{s(i,t)}$ and

[12] 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.

$\bar{X}_{s(i,t)}$). 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 plus indicators for whether the school is public, subsidized private, or private, and for whether it is located in a rural area. Following Wooldridge (2010), we also add the across-time average of year fixed effects dummies to account for our unbalanced panel. As in Altonji and Mansfield (2018), this approach attempts to absorb the across-schools variation in unobservable school characteristics by controlling for the school averages of its observed characteristics.

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. 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. Since the teacher $\mu_{j(i,t)}$ and principal $\theta_{p(i,t)}$ fixed effects in (1) are identified by movers and can only be compared within connected sets (Abowd et al., 1999), we estimate our model within the largest connected set of teachers and principals. Our final estimation sample includes 7,735,683 student-subject-year observations, corresponding to 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 empirical distributions of our principals' and teachers' fixed effects estimates are plotted in Appendix Figure A.3. 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. Likewise, the standard deviation of teacher fixed effects is 0.41. 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 bias can be more serious for quadratic forms of teachers' fixed effects because of the small populations of students used to identify their value-added. Moreover, in our two-way fixed

effects setting, quadratic forms might also be biased due to "bottlenecks" in the connected set (Jochmans and Weidner, 2019; Kline et al., 2020).

To account for this issue—and since our focus is on school principals—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 corresponds to 0.29.[13] 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 (Aaronson et al., 2007), we shrink our estimates of principal effectiveness to obtain an adjusted measure $\theta_p^*$, 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 $p$, $\sigma_p^2$ 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 $\theta_p^*$ 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.

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

## 3.1 On the validity of our estimates

**Identification assumptions:** 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 (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.

This quasi-experimental design rests on the identification assumption that principal turnover within a school is uncorrelated with student and school characteristics.[14] We begin with event studies looking at the evolution of course grades around the events of entry and exit of low and high value added principals (Figure 1). Naturally, for this exercise, we restrict the sample to the subset of principals who switched schools between 2011 and 2016 (the period for which course grade data is available), and who belong to the top or bottom of the principal effectiveness distribution. 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 course grade gains is very similar to the change in mean principal effectiveness, which

[14] Although untestable, this assumption is plausible insofar as teachers and students are unlikely to immediately switch to a different school because the principal changed.

is 0.09. Consistent with the idea that our estimates of principal effectiveness are forecast unbiased, the null hypothesis that the observed impact on mean gains equals the increase in mean value-added cannot be rejected (p = 0.35). The remaining panels of Figure 1 repeat the event study in Panel A for other types of arrivals and departures. In all but the last panel, the change in course grade gains is significantly different from 0 with $p < 0.05$ and is not significantly different from what one would forecast based on the change in mean principal effectiveness. These event studies show 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.

To complement the previous exercise, we also implement a falsification test similar to that in Rothstein (2010). 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. For this test, we compute "jackknife" estimates of principal effectiveness and consider both a model of gains and a model with lagged dependent variable. As shown by Appendix Table A.3, we do not find much evidence of a 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). It is worth noticing that failing to reject the null hypothesis that future principals have an impact on current achievement does *not* guarantee that there is no sorting. Thus, we take this evidence only as suggestive.

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).[15] Since our model also considers additive teacher and principal effects, one might be worried about the bias in our measure of principal effectiveness. 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

[15]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 considered some specific models of wage determination (e.g., Gibbons et al., 2005).

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, to address concerns related to politically motivated appointment of school principals (Colonnelli et al., 2020; Akhtari et al., 2020), we estimate our model again but now including party turnover, vote margin, and a measure of the concentration of vote shares (Herfindhal-Hirschman Index) as municipal level controls. Reassuringly, the correlation between our main measure of principal effectiveness and the one obtained from this alternative model is 0.997.

**Teachers' perceptions:** To assess how sensible is our measure of principal effectiveness, we contrast it against the perceptions of teachers. We examine 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 2 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), 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 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 this result as *prima facie* evidence that our estimates of principal effectiveness are sensible.

## 3.2 What do school principals do?

Before concluding this section, we study the relationship between our measure of principals' effectiveness and different outcomes related to management practices (Bloom et al., 2015; Di Liberto et al., 2015; Lemos et al., 2021). We leverage data from different sources to evaluate whether more effective principals: i) rely more on the sorting of students across classrooms, ii) receive fewer complaints from parents, and iii) are associated with lower levels of teachers' turnover.

**Students' Tracking:** School principals can impact schools' outcomes through different margins, one of them is the sorting of students with different achievement levels to different classroom (Duflo et al., 2011; Card and Giuliano, 2016). To assess the extent of tracking in Chile, and to explore how principal's effectiveness is related to it, we construct a sorting index at the school-year level à la Kremer and Maskin (1996).[16] Intuitively, perfect sorting is a case in which all variation in classrooms' average course grades comes from variation between instead of within classrooms.

[16]We measure how students sort across classrooms by estimating: $\bar{y}_{cg} = \alpha_g + \tau_{st} y_{i(c,g)} + \varepsilon_{cg}$, where $\bar{y}_{cg}$ stands for the average course grade of classroom $c$ of grade $g$, and $y_{i(c,g)}$ represents the course grade of student $i$. We estimate this specification by school and year, including grade fixed effects. The school-year sorting index is thus given by: $\tau_{st}$.

Figure 4, Panel A, plots a bin scatter with the relationship between the average sorting index (aggregated at the principal level) and principal effectiveness. We cannot reject the null of *no* association between principal effectiveness and tracking at conventional levels. This may not be surprising in light of the low prevalence of sorting in Chilean schools. According to our index, sorting is 0.05 on average. To benchmark this number, we rank students within a school-grade and sort them across all classrooms (in a given grade) to obtain an average upper bound of 0.7. Thus, sorting based on course grades represents only a 7% of our empirical upper bound (0.05/0.7).

**Parents' Complaints**: We explore how principal effectiveness relates to the number of parents' complaints. To do so, we leverage administrative data used to monitor that schools comply with the laws, regulations, and instructions issued by the Superintendency of Education. Using these data, we calculate the number of complaints (per 100 students) filed against the schools in which each principal works. Our data includes complaints related to: bullying/discrimination, denial of enrollment, poor infrastructure, teacher absenteeism, and school accidents.

Figure 4, Panel B, plots a bin scatter with the relationship between a Z-score of the number of different complaints (per 100 students) and principal effectiveness. As shown by Appendix Table A.4, we find that—including year and municipality fixed effects—a one standard deviation increase in principal effectiveness is associated with a 0.024 decrease in the Z-score. Principal effectiveness has a more salient impact on complaints related to "Denied Enrollment" and "Bullying or Discrimination". A one standard deviation increase in principal effectiveness is associated with a 6.1% decrease relative to the sample mean in complaints related to the former cause and a 5.6% decrease in complaints related to the latter.

**Teachers' Turnover**: We now study the relationship between principal effectiveness and the turnover rate of teachers working for them.[17] More specifically, we define our outcome variable as the share of teachers that leave the school run by principal $p$, either because of a job-to-job or a job-to-unemployment transition.

Figure 4, Panel C, plots a bin scatter with the relationship between teachers' turnover and principal effectiveness. To further explore this dimension, in Panel D, we plot the

[17]Teacher shortages and high turnover rates have recently received considerable attention from policymakers, as they impose financial costs on schools and may affect students outcomes (Ronfeldt et al., 2013; Hanushek et al., 2016).

association between principal effectiveness and the turnover of high-value-added teachers (i.e., the share of teachers whose value-added is above the median and who leave the school). Appendix Table A.4 shows our estimates after including municipality and year fixed effects. Overall, we find that principal effectiveness is associated with a decrease in teachers' turnover. Considering that the average turnover in our sample is 0.098, our estimates imply an effect size of 2%. Interestingly, the figure suggests that more effective principals are strongly associated with a decrease in the likelihood that *high*-value-added teachers leave the school. The effect size is larger in this case and corresponds to 23%. This finding is consistent with the idea that principals can recognize good teachers (Jacob and Lefgren, 2008), but this should be interpreted with caution as the covariance between principal and teacher fixed effects might suffer from estimation bias (Kline et al., 2020).

# 4 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, and we discuss the consequences of this type of policy.

## 4.1 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,[18] 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 (Rothstein, 2015; Cullen et al., 2016; Biasi, 2021) and teachers' firing policies (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.5 in the Appendix presents some

[18]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).

features of our data. Perhaps not surprisingly, we find that hourly wages (residualized with respect to year and municipality 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 $t$, $\text{Voucher}_{pt}$ is an indicator that equals one if the principal works at a voucher-private school (and zero otherwise), $\gamma_t$ 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 $\beta_2$, 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 payments described in section 2, and $\text{base}_{pt}$ corresponds to the total wage minus the bonuses.

Table 3 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 component. 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 reveal other 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 (2021) 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 (Dal Bó et al., 2013; 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 (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.

## 4.2 Selection and recruitment of public school principals

We exploit the non-eligibility of private schools and the timing of adoption of this new selection system within public schools to study the impact of the 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.6 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, with around 370 new principals elected every year since 2012.

To formally 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:

$$\begin{aligned}\hat{\theta}_{st} = \alpha_s + \alpha_t + \beta_1 \times ADP_{st} \times \text{ Principal Turnover}_{st} \\ + \beta_2 \times \text{Principal Turnover}_{st} + \sum_t \Phi_t' X_s I[year = t] + \epsilon_{st},\end{aligned} \tag{4}$$

where $s$ and $t$ stand for school and year, and the dependent variable $\theta_{st}$ 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.[19] $X_s$ 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, $\alpha_s$ and $\alpha_t$ are school and year fixed effects, and $\epsilon_{st}$ is an error term robust to heteroscedasticity and clustered at the school level. The parameter of interest is $\beta_1$, 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.[20]

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

[19] In Chile, this new selection system is known as ADP, the acronym of *Alta Dirección Pública*.

[20] 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.

that correlated with the timing of adoption of this new system and other determinants of principal's effectiveness. To partially address this concern, we estimate a variation of model (4) with a dynamic treatment. Specifically, we estimate the following regression:

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

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 schools independent of whether they elected a principal via the new selection process or not.

Figure 5, panel A, presents our estimates of the $\beta_j$s coefficients, 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.[21] The figure shows that the ADP system increases principal effectiveness by 0.06 standard deviations, on average. It also provides visual support for our identification strategy, as 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. Importantly, 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 complement this test by computing the pre-trend that has a 80% power of being detected given the precision of the estimates in the pre-period; and also the adjusted pre-trend that takes into account the pre-testing bias that arises from the fact that the analysis shown is conditional on passing a pre-test (Roth, 2021). In Panel A of Figure A.7, we present the same figure as before but now adding these two trends. For both trends, the average bias that they can create represent less than 50% of our baseline coefficient, and if the trend is negative as is suggested by the pre-treatment coefficients then the estimated parameter is a

[21] Note that standard errors tend to be larger in the post-period as we observe fewer treated schools for several years after a principal's turnover. Our results are robust to only keeping in the sample those schools that we observe for the 8-years window.

lower bound. Finally, we follow Rambachan and Roth (2021) and estimate the confidence set for our parameter of interest allowing for linear and non-linear deviations from the parallel trends assumption.[22] In the case of non-linear deviations, we allow the change in trend from consecutive periods ($M$) to be as large as the size of the pre-trend that has a 80% power of being detected given the precision of the estimates in the pre-period (Roth, 2021), which is 0.013. In Figure A.7 Panel B, we present the results. For both linear and non-linear deviations, we find that the confidence set at 90% does not include the zero, which suggests that our results are robust to moderate deviations from the parallel trends assumption.

As an additional robustness check, we estimate equation (5) again, but now we only consider observations from the pre-reform period: 2008-2011. In this case, we replace $ADP_{st}$ by an indicator that equals one if the school is public, and zero if it is private. Intuitively, this placebo specification allows us to assess whether a new principal's appointment at public schools had an impact on principal effectiveness (relative to new appointments at private schools) during the period when the new selection process was *not* in place. Naturally, since we only consider data from the pre-reform period, we must use a shorter window around principals' turnover in this case. Panel C of Figure 5 plots the point estimates and 95% confidence intervals obtained from this exercise. Reassuringly, we find that—before the civil service reform was enacted—principals' turnover at public schools had no statistically significant impact on principal effectiveness. If anything, when compared to principals' turnover at private schools, turnovers at public schools lead to a decrease in principal effectiveness.

We now turn to our parametric *difference-in-differences* estimates obtained from equation (4). Table 4 presents these results. Column 1 suggests 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. Moreover, 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. (2009) and truncate our analysis sample based on a propensity score that estimates the probability that a school appoints a principal under

[22] We estimate the confidence set for the coefficient in the year that there was a change in the school principal (year=0).

the new system. We also estimate an effect of 0.06 standard deviations using this truncated sample.

In column 6, we 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. In this case, never treated units are public schools that changed their principal after the ADP reform was enacted, but did *not* appoint their new principal using the new selection system. This could happen, for instance, if the call for the contest was left deserted.[23] We find similar results in this case: 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 *late* adopters.

Recent literature on this type of two-way fixed effects estimation have shown that estimates from this model can substantially differ from the group's ATT in the presence of treatment heterogeneity (Borusyak and Jaravel, 2017; De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021). 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 effects estimate following Goodman-Bacon (2021). 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.9). 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 4 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.[24]

As additional robustness checks, we perform two placebo exercises. First, we consider any principal turnover that happened before the reform (2010-2011) 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

[23] Appendix Table A.5 compares public schools that did not adopt the new selection system to ever, early, and late adopters (as well as to private schools). We document that adoption is positively associated to school size and urban status.

[24] In Figure A.8, we present the dynamic versions of these two-way fixed effects models, that leverage variation within public schools.

placebo treatment takes the value one if there is a change in the principal and stays as a one afterwards. 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 in the private sector. As shown in Table A.7, these findings also remain unchanged if we consider the models proposed by De Chaisemartin and d'Haultfoeuille (2020) and Callaway and Sant'Anna (2020).

Finally, we study principals' characteristics, Table 5 summarizes the differences between public school principals who were selected by the new system and those who were not. We find that principals selected under the reformed process are less likely to have worked as teachers and are more likely to have worked at administrative positions; they are also more likely to have worked in the private schooling sector in the past. Principals appointed under the new system are also slightly younger and more likely to have a college degree. In the same vein of the previous statistics, Appendix Table A.6 compares the characteristics of the schools of origin with those of the school of destination for principals appointed with the new system. We find no differences in terms of school wages, but in line with a preference for amenities hypothesis, these principals are arriving at municipalities that have higher income and more years of schooling.

## 4.3 Discussion

Before concluding, in this subsection we discuss the impact of the new system of principals' appointment on the school personnel, the system's equity, and students' long-run outcomes.

**Impact on School Personnel:** Leveraging records on school staff and teachers' evaluations, we can study the impact of the "ADP" selection on the churn of school personnel. We present estimates from our preferred specifications (4) and (5) in Table 6 and Appendix Figure A.10. In this case, the dependent variables are indicators for any hiring or firing within the school. Our results reveal that the appointments using the new selection system increased the likelihood of firing and hiring personnel of the principal's support team by 5 and 6 percentage points, respectively. This result is consistent with the fact that public school principals are allowed to form their own management teams (deputy director, inspector general, and chief technician). Perhaps more interestingly, column 5 of Table 6 shows that principals appointed under the new selection system increased the firing of teachers whose performance was classified as "basic" or "unsatisfactory" by 12 percentage

points.[25] Nonetheless, as shown by columns 3 and 4, the overall likelihood of hiring and firing personnel from the teaching body did not change as a consequence of ADP selection.

**Impact on Equity:** 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.

**Impact on Students' Long-term Outcomes:** 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,[26] 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), 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 ed-

[25] For this exercise, we use records from the teachers' evaluations from 2007 to 2016. The teachers' evaluation system operates on the basis of four sources of evidence: a portfolio, an interview by a peer teacher, a written report of two school authorities, and a self-evaluation report. The evaluation system classifies teachers into four groups: "outstanding", "competent", "basic," or "unsatisfactory". See Appendix C for details.

[26] 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).

ucation via the centralized college admissions system. We present our findings in Table 7 and Appendix Figure A.11. 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.[27] The latter results, coupled with i) the positive sorting induced by admission scores into enrollment at more selective institutions (Rodríguez et al., 2016), and ii) the large positive returns associated to more selective institution-mayors (Hastings et al., 2013; Zimmerman, 2019), leads us to expect positive long-run impacts of this policy on income and other non-pecuniary outcomes (Oreopoulos and Salvanes, 2011; Bautista et al., 2021).

# 5 Conclusion

As states raise their level of ambition in delivering a wide range of public services to their citizens, the need for an effective body of public servants has increased (Besley et al., 2021). In this article, we studied the effectiveness of managers in education, an essential public service and a landmark of state capacity, often recognized as an important determinant of individual earnings, macroeconomic growth, and equity (Barro, 1991; Card, 2001; Chetty et al., 2020). We study the case of Chile, a country where government expenditure in education represents more than one-fifth of the budget, and the efficiency and equity of educational policy are often at the center of the political debate.

Like in many other countries, the implementation of incentive contracts is challenging in Chilean public schools. Indeed, as we document in this paper—while public schools reward tenure more and penalize women less in terms of wages—they do not seem to compensate principals' effectiveness. This absence of high-powered incentives underscores the role that personnel selection plays to enhance the allocation of talent in the public sector.[28] We contribute with novel evidence on this dimension of the labor market. Leveraging detailed

[27]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.

[28]In this paper, we developed a simple two-sided matching model to make this point. A principal-agent model would deliver similar results since selection provides a way of influencing the "type" of agent that the public sector gets, which in turn affects performance.

administrative data and quasi-experimental variation from a civil service reform, our results show that even in the absence of high- powered incentives, simple rules such as limiting the discretion of local politicians over the appointment of school principals can be a powerful way of improving public sector performance.

Finally, and even though a welfare analysis is beyond the scope of this paper, a number of factors suggest positive and potentially large effects of this policy 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 strengthen the positive effects of a more competitive and transparent recruitment.

# 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 (2020): "Political turnover, bureaucratic turnover, and the quality of public services," *Conditionally Accepted, American Economic Review*.

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.

Bautista, M. A., F. González, L. R. Martinez, P. Munoz, and M. Prem (2021): "Does Higher Education Reduce Mortality? Evidence from a Natural Experiment in Chile," *Working Paper*.

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.

Besley, T. J., R. Burgess, A. Khan, and G. Xu (2021): "Bureaucracy and Development," .

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

Biasi, B. (2021): "The labor market for teachers under different pay schemes," *American Economic Journal: Economic Policy*, 13, 63–102.

Biasi, B. and H. Sarsons (2021): "Flexible Wages, Bargaining, and the Gender Gap,"

Tech. rep., Quarterly Journal of Economics.

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

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

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

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

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.

Brown, C. and T. Andrabi (2020): "Inducing positive sorting through performance pay: Experimental evidence from Pakistani schools," *University of California at Berkeley Working Paper*.

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

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 (2020): "Income

segregation and intergenerational mobility across colleges in the United States," *The Quarterly Journal of Economics*, 135, 1567–1633.

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 (2020): "Is Private Education Worth it? Evidence from School-to-Work Transitions," *Working Paper*.

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

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

Cullen, J. B., E. A. Hanushek, G. Phelan, and S. G. Rivkin (2016): "Performance information and personnel decisions in the public sector: The case of school principals," Tech. rep., National Bureau of Economic Research.

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

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.

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

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

Epple, D., R. E. Romano, and M. Urquiola (2017): "School vouchers: A survey of the economics literature," *Journal of Economic Literature*, 55, 441–92.

Errázuriz, M., M. Kutscher, and C. Williamson (2016): "La Ley N° 20.501 sobre

Calidad y Equidad en los Colegios Públicos: Efectos de la Selección de Directores por la Alta Dirección Pública (ADP)," *Documento de Trabajo, Clapes UC*, 28.

Estrada, R. (2019): "Rules versus discretion in public service: Teacher hiring in Mexico," *Journal of Labor Economics*, 37, 545–579.

Fenizia, A. (2021): "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. (2021): "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*.

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

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

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

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.

Jackson, C. K. (2018): "What do test scores miss? The importance of teacher effects on non–test score outcomes," *Journal of Political Economy*, 126, 2072–2107.

Jacob, B. A. and L. Lefgren (2008): "Can principals identify effective teachers? Evi-

dence on subjective performance evaluation in education," *Journal of Labor Economics*, 26, 101–136.

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

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.

Khan, A. Q., A. I. Khwaja, and B. A. Olken (2019): "Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings," *American Economic Review*, 109, 237–70.

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

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

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.

Leaver, C., O. Ozier, P. Serneels, and A. Zeitlin (2021): "Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools," *American Economic Review*.

Lemos, R., K. Muralidharan, and D. Scur (2021): "Personnel management and school productivity: Evidence from india," 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.

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.

Ornaghi, A. (2019): "Civil service reforms: Evidence from US police departments," .

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

Rambachan, A. and J. Roth (2021): "An honest approach to parallel trends," .

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.

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

Roth, J. (2021): "Pre-test with caution: Event-study estimates after testing for parallel trends," .

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.

Ruiz-Tagle, C. (2019): "Selección de directivos escolares sobre la base de procesos competitivos: evidencia de una política para Chile," *Calidad en La Educación*, 51, 85–130.

Scot, T., R. Dahis, and L. Schiavon (2021): "Selecting Top Bureaucrats: Admission Exams and Performance in Brazil," *Available at SSRN*.

Silva, F. (2014): "Alta dirección pública y educación municipal: Ley de calidad y equidad de la educación," .

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 (2020): "Discretion and Destruction: Promotions, Performance, and Patronage in the Royal Navy," *Working Paper*.

Weber, M. (1922): "Economy and Society," .

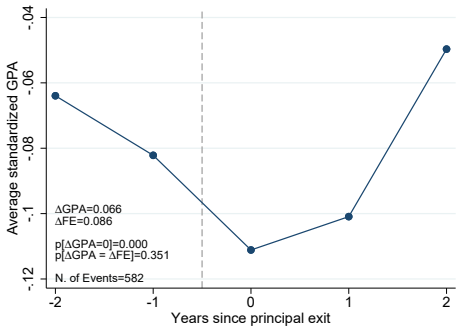
Wooldridge, J. M. (2010): "Correlated random effects models with unbalanced panels," *Michigan State University, Department of Economics*.

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

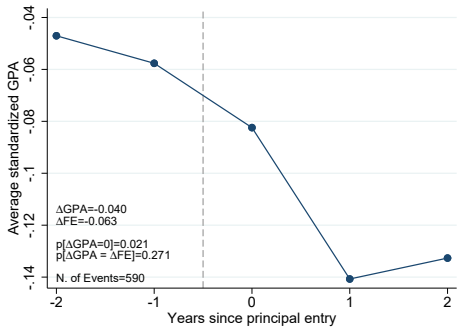
Zimmerman, S. D. (2019): "Elite colleges and upward mobility to top jobs and top incomes," *American Economic Review*, 109, 1–47.

# Figures and Tables

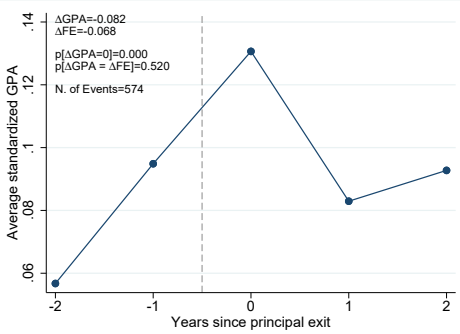
Figure 1: Impacts of Principal Entry and Exit on Student's Performance



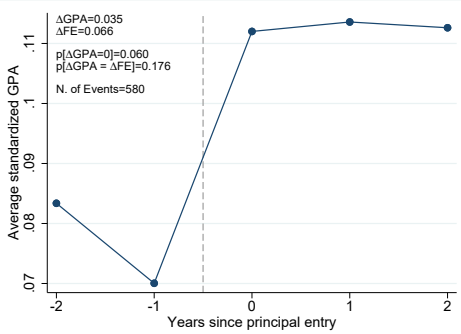

**A.** Low Value-Added Principal Exit




**B.** Low Value-Added Principal Entry



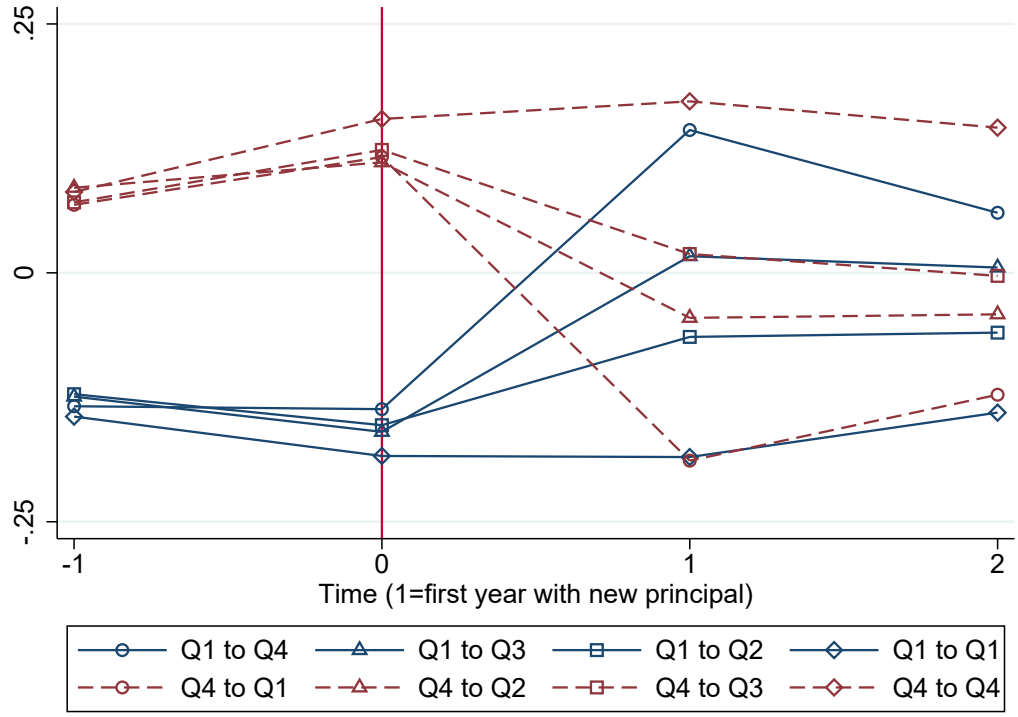

**C.** High Value-Added Principal Exit




**D.** High Value-Added Principal Entry

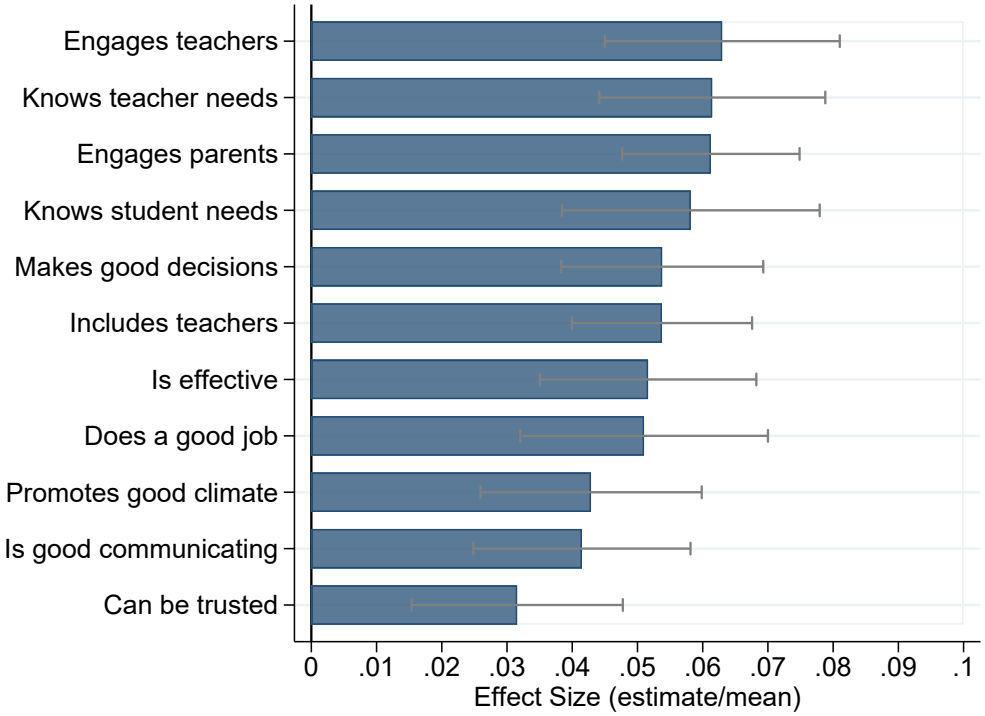
**Notes:** These figures plot event studies of of standardized course grades as principals arrive at or leave a school at year t=0. Panels A and B plot the impact of the exit and entry of a low value-added principal (principals with VA in the bottom 20% of the distribution) on mean course grades. Likewise, 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



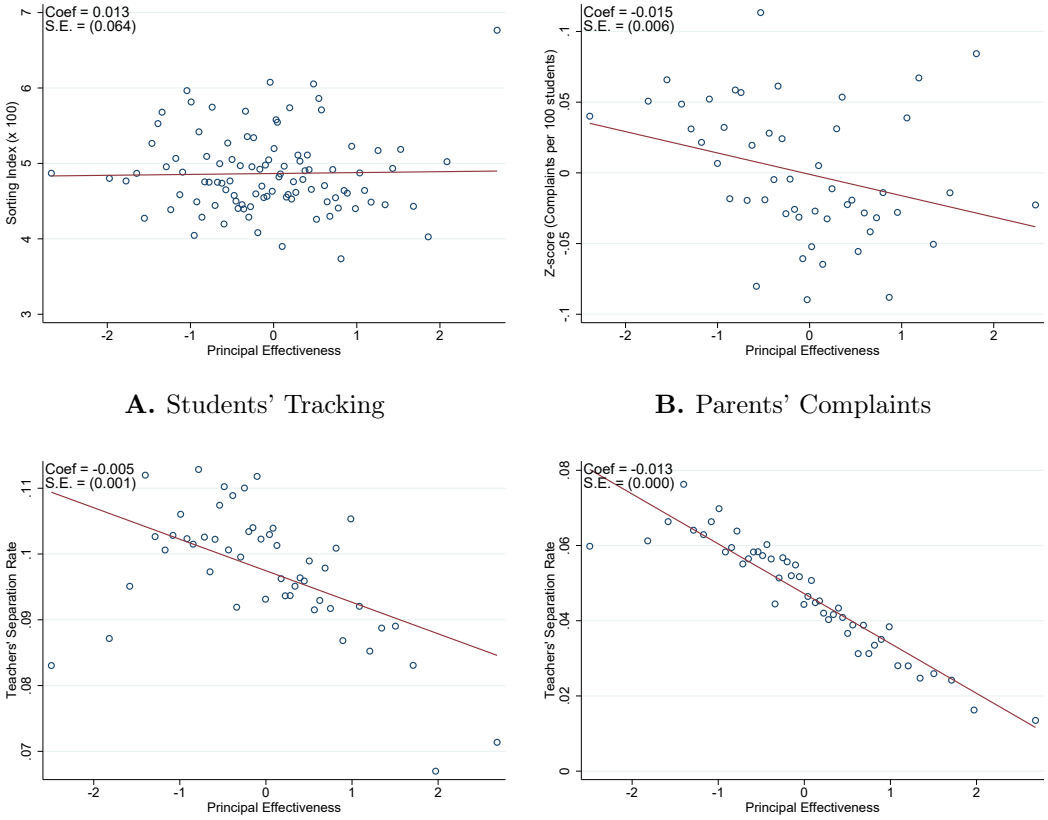

**Notes:** This figure plots the mean residualized course grades of teachers who changed principal in 2011-2016. We consider the first time a teacher switches to work under a new principal but we do not condition on holding the old or new job relationship for a minimum number of years. Each principal 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: What do School Principals Do?




**A.** Students' Tracking

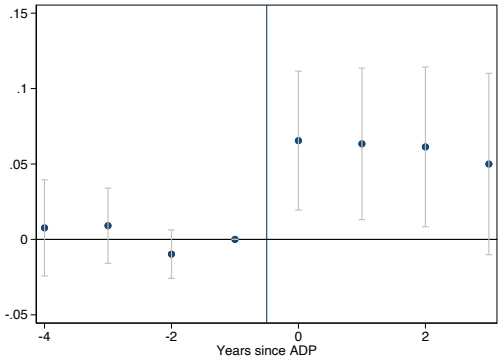
**B.** Parents' Complaints

**C.** Teachers' Turnover (All)

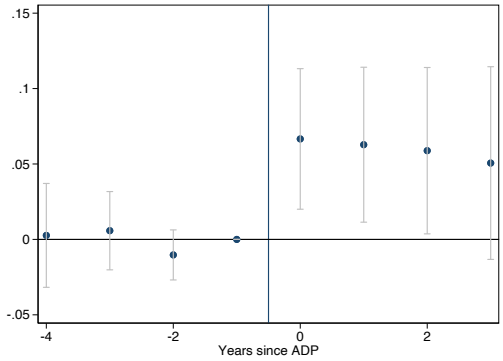
**D.** Teachers' Turnover (High VA)

**Notes:** This figure shows the association between principal effectiveness and different dimensions on which school principals can have an impact. Specifically, we plot bin-scatters with the projection of different outcome variables (at the school-year level) on principal effectiveness. Panel A considers students' tracking (defined à la Kremer and Maskin, 1996). Panel B considers a Z-score of parents' complaints (per 100 students). Panel C considers the separation hazard (job to job or job to unemployment) of teachers; and panel D replicates this analysis but focusing on the separation hazard of high value-added teachers. Each panel reports the coefficient from a simple regression of the outcome variable on principal effectiveness and its bootstrapped standard errors (100 replications) clustered at the school principal level.

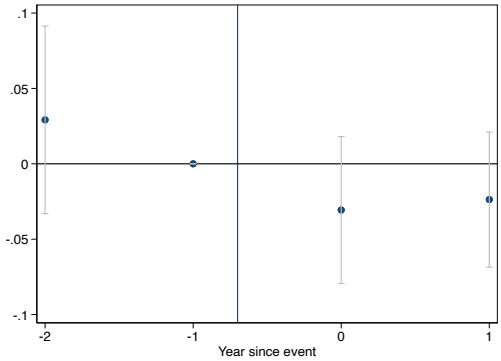
Figure 5: Principal Selection and Principal Effectiveness




**A.** No Controls




**B.** With Controls




**C.** Pre-Reform

**Notes:** Panels A and B show the impact of appointments under the new selection system on the effectiveness of public schools' principals. Specifically, we plot the point estimates and 95% confidence intervals estimated from equation (5) only considering school and year fixed effects (A) and further including controls by school and municipality characteristics during the pre-reform period (measured in 2010) interacted with year dummies (B). Panel C shows the impact of principal turnover on the effectiveness of public schools' principals before the reform was enacted. Specifically, we plot the point estimates and 95% confidence intervals obtained from equation (5) but only considering the pre-reform period (2008-2011) for estimation and replacing $ADP_s$ by an indicator of whether the school is public.

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: 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 3: 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) |
| $\text{Age}^2$ | | -0.000*** | | -0.000*** | | 0.000** |
| | | (0.000) | | (0.000) | | (0.000) |
| $\text{Age}^2$ 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 4: 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 | | | | 0.001 | -0.031* |
| | (0.018) | (0.018) | (0.018) | (0.018) | (0.019) | | | | (0.088) | (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 | 2389 | 2389 | 1666 | 2802 |
| R-squared | 0.931 | 0.931 | 0.931 | 0.931 | 0.930 | 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 20010-2011 (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 5: Characteristics of Principals by ADP Status

| | Public Schools | | | |
|---|---|---|---|---|
| | Not ADP | ADP | Difference | Private Schools |
| | (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 School Staff

| | Principal's Support Team | | Teaching Staff | | | |
|---|---|---|---|---|---|---|
| | | | All | | Bad quality | |
| | Firing | Hiring | Firing | Hiring | Firing | Hiring |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Principal Turnover × ADP | 0.047*** | 0.056*** | -0.002 | 0.006 | 0.116*** | -0.024* |
| | (0.012) | (0.012) | (0.009) | (0.008) | (0.014) | (0.014) |
| Principal Turnover | 0.010 | 0.038*** | 0.005 | 0.008 | -0.022** | 0.011 |
| | (0.009) | (0.008) | (0.007) | (0.006) | (0.009) | (0.008) |
| Observations | 30,714 | 30,714 | 30,714 | 30,413 | 30,714 | 30,413 |
| R-squared | 0.279 | 0.272 | 0.346 | 0.333 | 0.274 | 0.310 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| School FE | Yes | Yes | Yes | Yes | Yes | Yes |
| School controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Municipality controls | Yes | Yes | Yes | Yes | Yes | Yes |
| # of Schools | 4934 | 4934 | 4934 | 4908 | 4934 | 4908 |
| Mean Dep Var | 0.118 | 0.094 | 0.911 | 0.931 | 0.177 | 0.192 |
| pvalue a + b = 0 | 0.000 | 0.000 | 0.742 | 0.022 | 0.000 | 0.296 |

**Notes:** This table presents the effects of the new selection system (ADP) on the churn of the school staff. "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 and 2 show the results when we use a dummy for any firing or hiring of personnel in the principal's support team (deputy director, inspector general, chief technician). Likewise, columns 3 and 4, show the results when we use a dummy for any firing or hiring of personnel in the teaching body. Finally, columns 5 and 6, show the estimates when we use a dummy for any firing or hiring of teachers who: i) took their teacher evaluation and ii) obtained a regular or bad classification. 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.

Table 7: Principal Selection and College Admissions Scores

| | College Entrance Exams score | | | |
|---|---|---|---|---|
| | Math | Spanish | Average | Application score |
| | (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 |

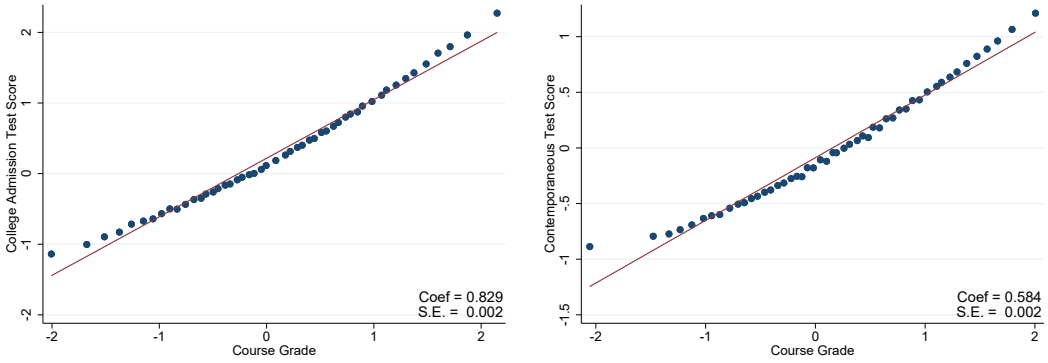
**Notes:** This table presents the effects of the new selection system (ADP) on college admission scores. "ADP" is a dummy that takes the value one after the first time a school selects a principal under the ADP system. "Principal Turnover" is a dummy that takes the value one after the first time a school 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

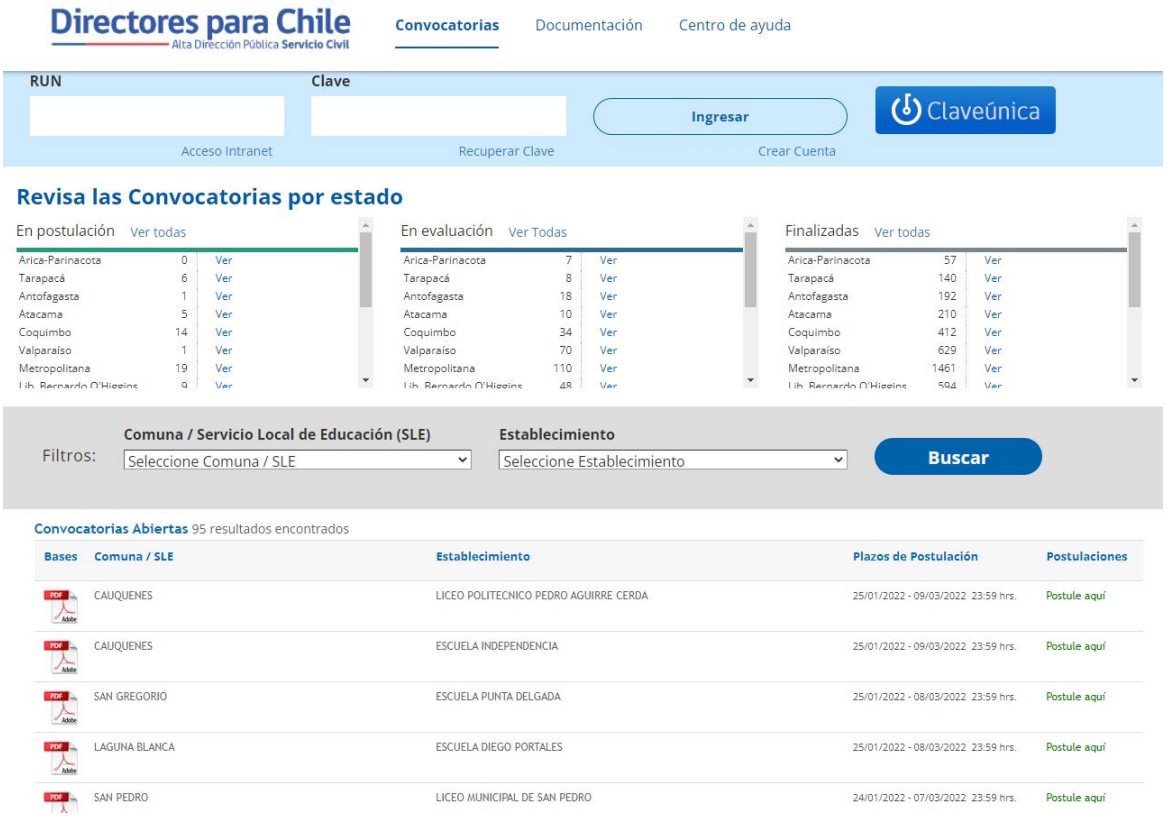
Figure A.1: Course Grades and Test Scores




**A.** PSU Test Score **B.** SIMCE Test Score

**Notes:** Panel A considers a sample of 132,585 students accepted into college and for whom we can compute college admission scores from the *Prueba de Selection Universitaria* (PSU), 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. Panel B considers a sample of 1,061,231 students for whom we observe test scores from the *Sistema de Medición de la Calidad de la Educación* (SIMCE) and course grades contemporaneously for Math and Spanish between 2011 and 2016. We report the coefficient and robust standard error from a linear regression of test score on course grades.

Figure A.2: Broad Calls



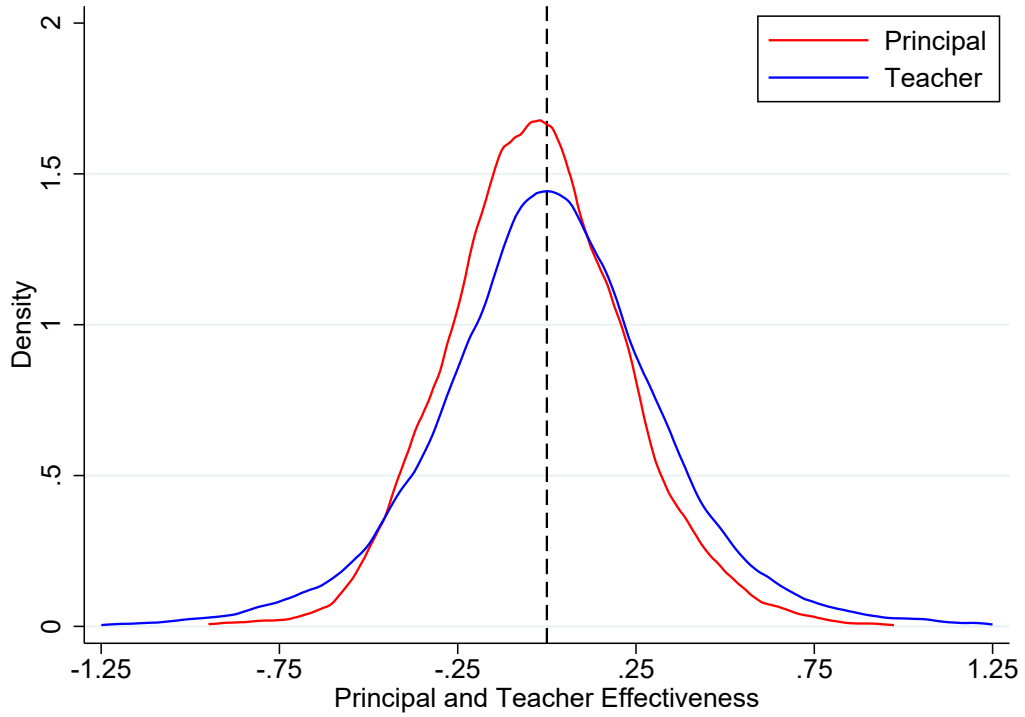

**A.** Webpage


**B.** Newspapers


**C.** Media Salience

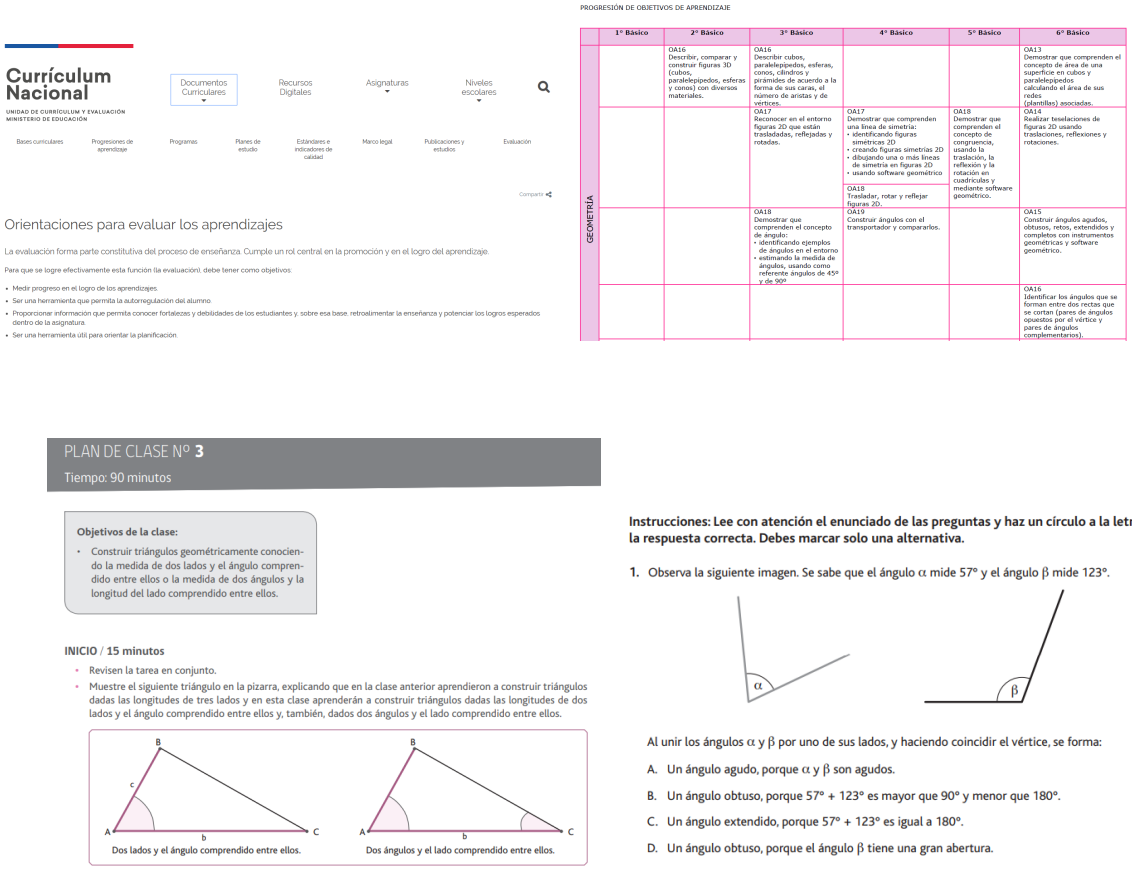
**Notes:** Panel A shows an screenshot from the webpage *directoresparachile.cl*. Information on open, ongoing and finished contests is publicly available to candidates. Panel B shows some advertisements of new positions in the newspaper, and Panel C presents cuts from different newspapers to exemplify the salience of the reform in the media. Source: webpage *directoresparachile.cl* and downloadable documents available at: https://www.serviciocivil.cl/sistema-de-alta-direccion-publica-2/adp-educacion/.

Figure A.3: Empirical Distributions of Teacher and Principal FE Estimates



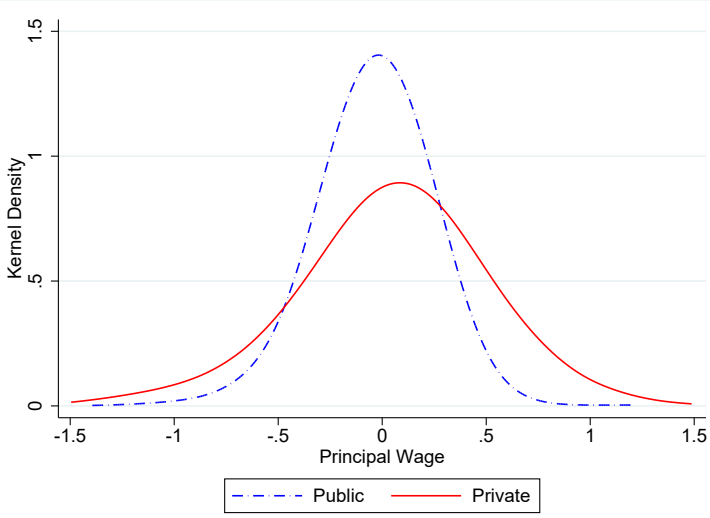

**Notes:** This figure plots kernel densities of the empirical distribution of teacher and principal fixed effects. Fixed effects are normalized using sum to zero constraints, and the densities are weighted by the number of students' course grade observations used to estimate each of these fixed effects. Note that these standard deviations are larger than the "true" standard deviations because of estimation error (see the main text for more details).

Figure A.4: Standardized National Curriculum

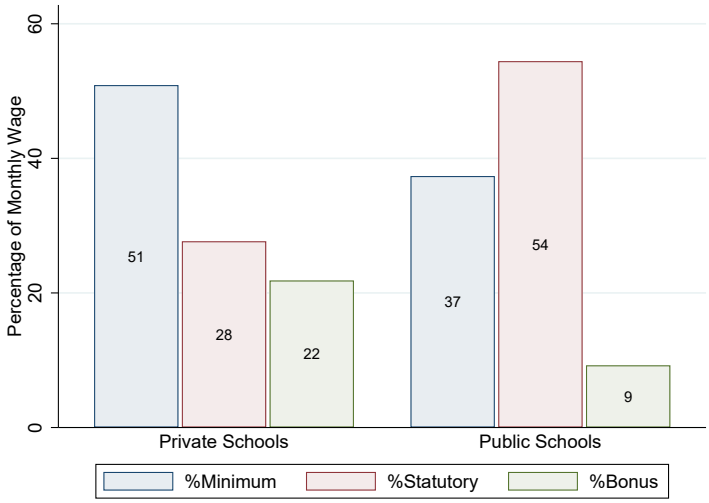



**Notes:** This figure displays different materials that are available to teachers as part of the national curriculum in Chile. Through the National Curriculum webpage (top left), teachers can access the specific topics that must be covered by grade and year (top right) and specific lesson plans and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

Figure A.5: Principals' Wages



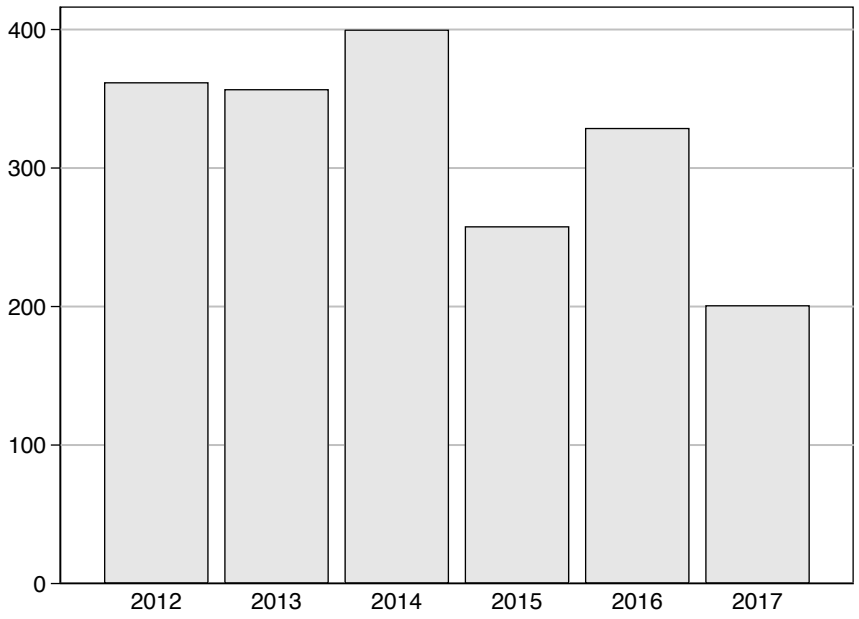

**A.** Residualized Log Wage




**B.** Wage Components

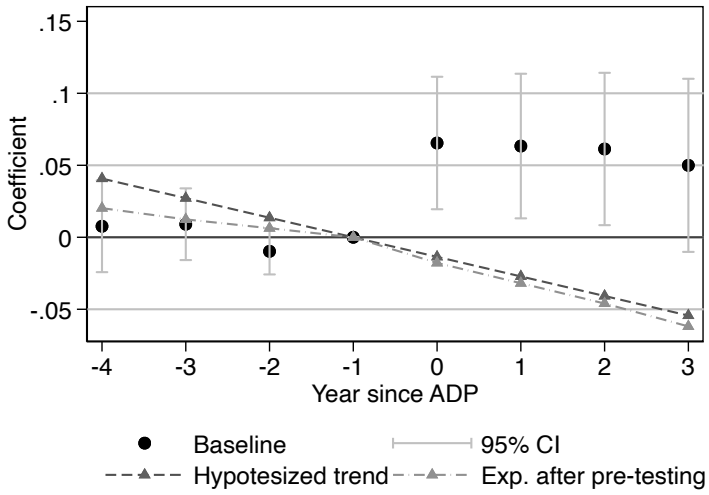
**Notes:** Panel A presents the distributions of log principals' wages in both public and subsidized private schools. Log principals' wages are residualized with respect to year and municipality fixed effects. Panel B decomposes the average monthly wage of school principals into the three components 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.6: Number of Newly Elected Principals by ADP, per Year

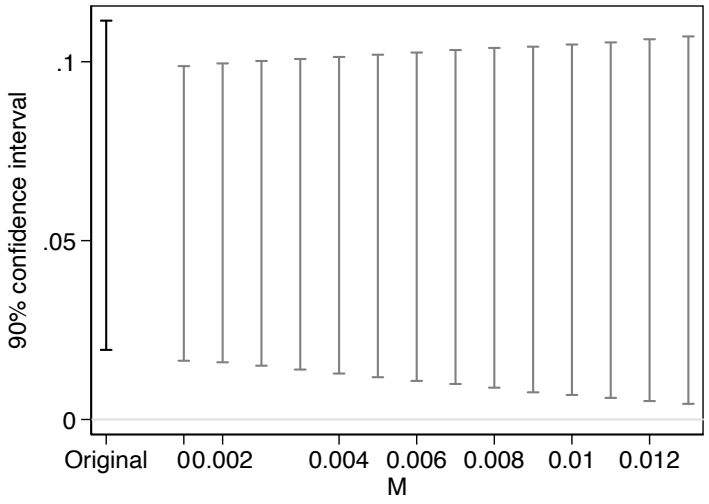



**Notes:** This figure shows the number of schools that elected a principal through the new ADP selection system for the first time, by year.

Figure A.7: Principal Selection and Principal Effectiveness: Parallel trends violations



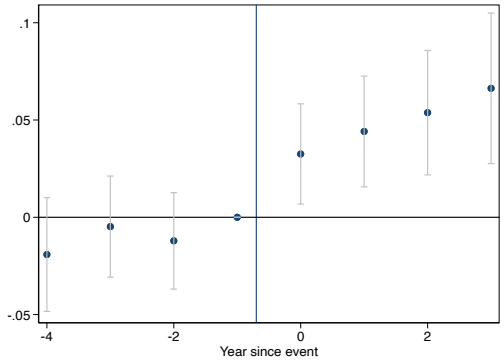

**A.** Pre-trend



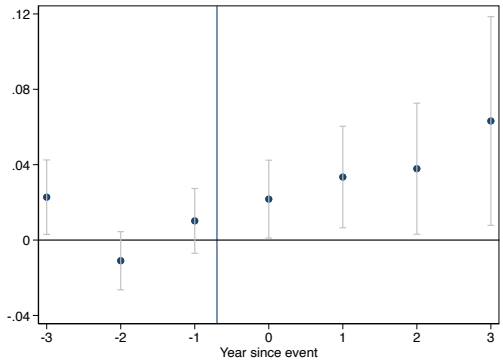

**B.** Violations from parallel trends

**Notes:** This figure presents two exercises related to the parallel trends assumption. In Panel A, we present the baseline dynamic figure but we add the pre-trend that has a 80% power of being detected given the precision of the estimates in the pre-period and the adjusted pre-trend that takes into account the pre-testing bias that arises from the fact that the analysis shown is conditional on passing a pre-test (Roth, 2021). In Panel B, we follow Rambachan and Roth (2021) and estimate the confidence set at 90% for our parameter of interest allowing for linear and non-linear deviations from the parallel trends assumption. We estimate the confidence set for the coefficient in the year that there was a change in the school principal (year=0). In the case of non-linear deviations, we allow the change in trend from consecutive periods ($M$) to be as large as the size of the pre-trend that has a 80% power of being detected given the precision of the estimates in the pre-period (Roth, 2021), which is 0.013. In Figure A.7 Panel B, we present the results.

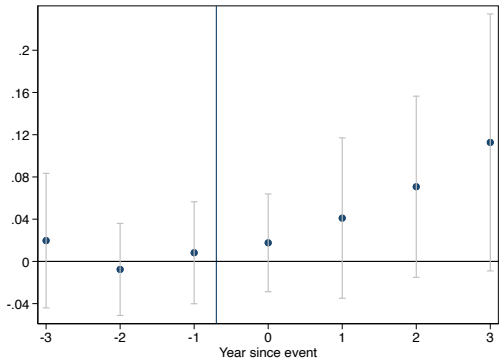
Figure A.8: Principal Selection and Principal Effectiveness within Public Schools




**A.** Two-way fixed effect



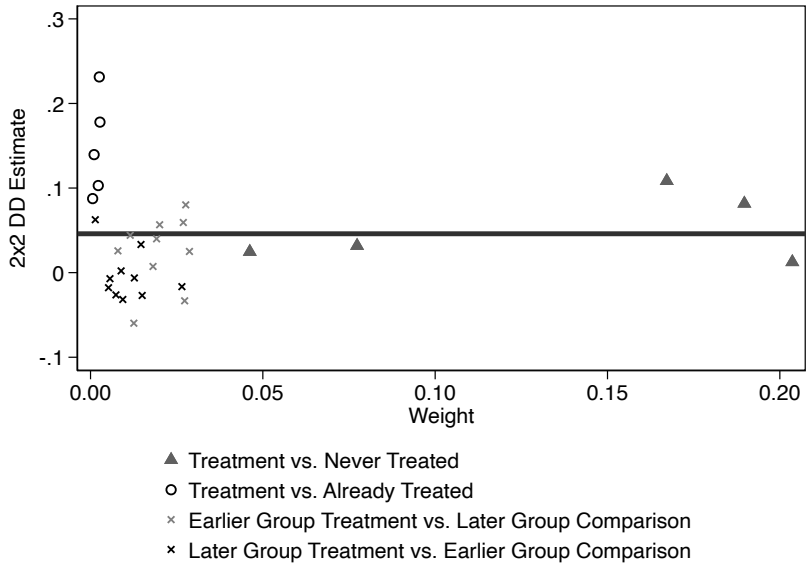

**B.** De Chaisemartin and d'Haultfoeuille (2020)




**C.** Callaway and Sant'Anna (2020)

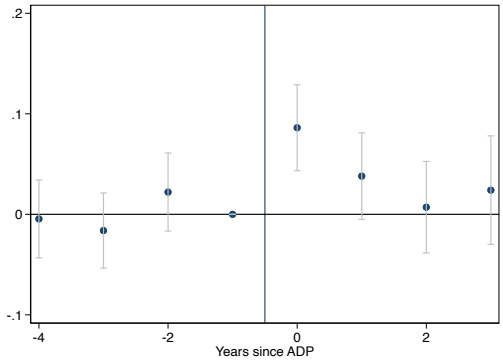
**Notes:** This figure presents the dynamic version of our *staggered difference-in-differences* approach in the sample of public schools. Panel A presents the estimates from a version of equation (5) for the sample of public schools. Panel B presents the dynamic version of the staggered difference-in-differences 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.9: Goodman-Bacon (2021) Decomposition

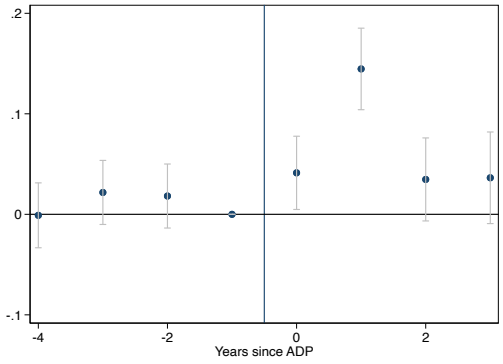



**Notes:** This figure presents the decomposition of the two-way fixed effect estimator suggested by Goodman-Bacon (2021).

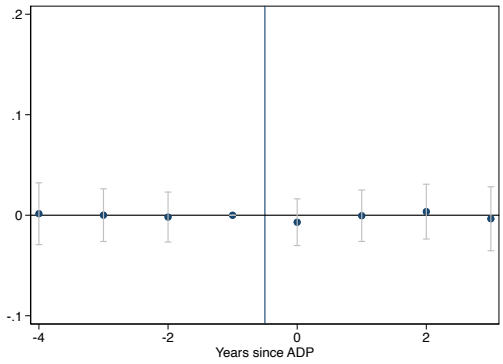
Figure A.10: Principal Selection and School Staff



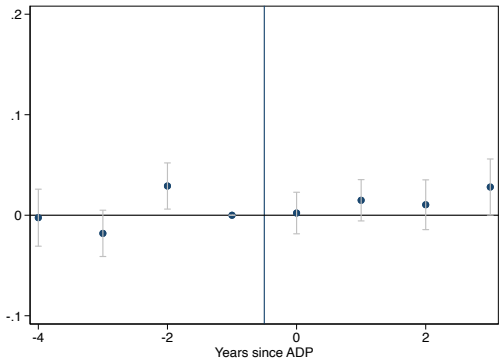

**A.** Firing: Principal's Team



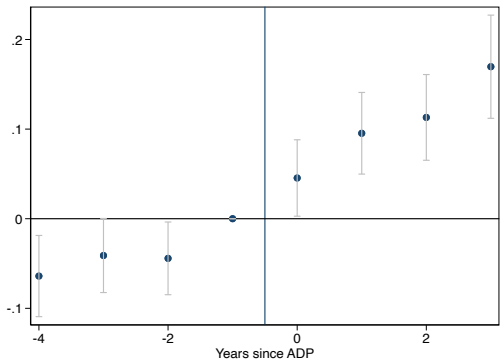

**B.** Hiring: Principal's Team



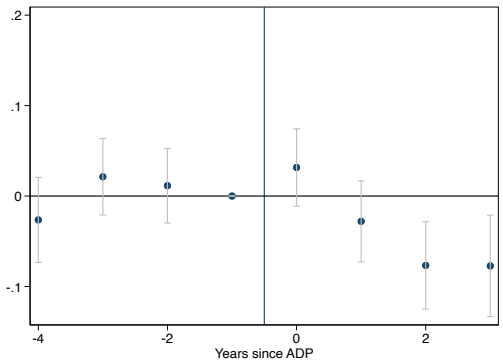

**C.** Firing: All Teachers




**D.** Hiring: All Teachers



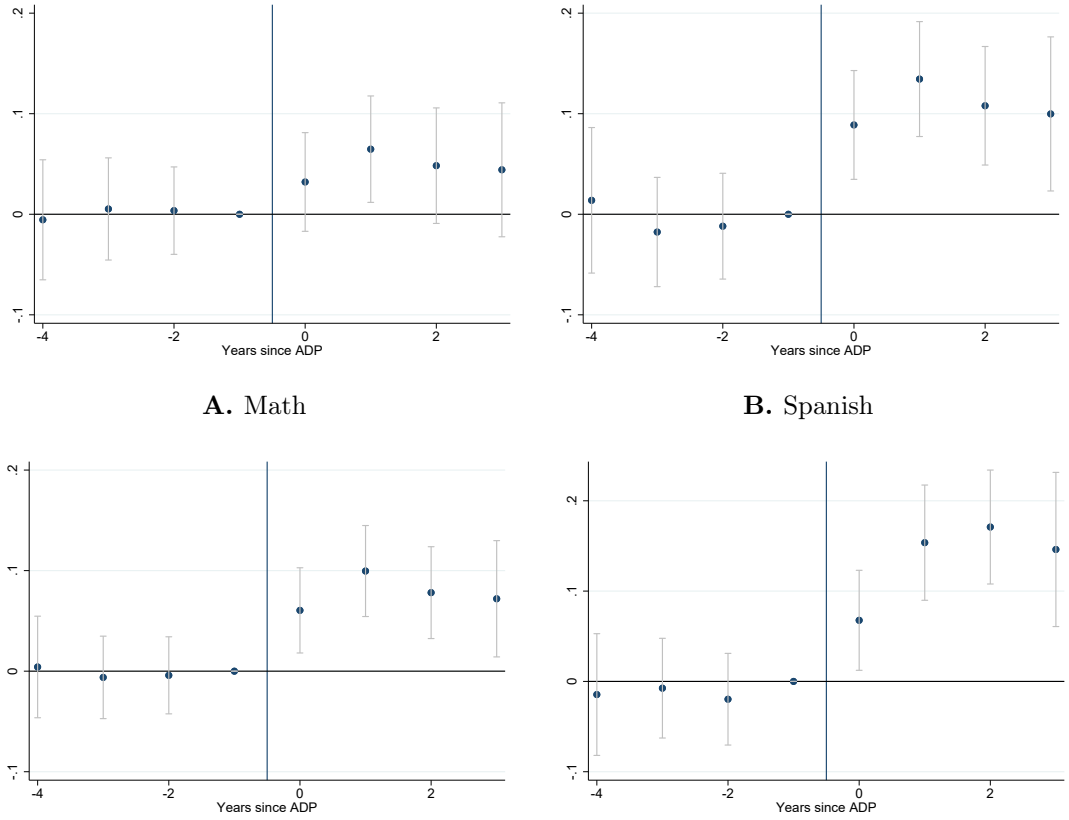

**E.** Firing: Teachers with Bad Evaluation




**F.** Hiring: Teachers with Bad Evaluation

**Notes:** In this figure, we plot the point estimates and 95% confidence intervals estimated from equation (5). The dependent variable is a dummy that takes the value one if there was any firing (Panels A, C, and E) or any hiring (Panels B, D, and F). In Panels A and B, this dummy is based on the principal's team (deputy director, inspector general, and the chief technician), in Panels C and D is based on all the teachers body, while in Panels E and F is based on teachers with poor performance according to teachers evaluations. 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.

Figure A.11: 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 A.1: Descriptive Statistics in Different Samples

| | **Full Sample** | | **$\Delta$ 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. "$\Delta$Teacher = 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: 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 A.4: What do School Principals Do?

| | | Parents' Complaints | | | | | | Teachers' Turnover | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sorting Index | Z-score | Accidents | Infrastructure | Teachers' Absenteeism | Bullying Discrimination | Denied Enrollment | All | High-VA | Low-VA |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Principal Effectiveness | -0.091 | -0.024*** | -0.001 | -0.002 | -0.001 | -0.019*** | -0.007** | -0.002** | -0.011*** | 0.009*** |
| | (0.063) | (0.007) | (0.001) | (0.001) | (0.001) | (0.006) | (0.003) | (0.001) | (0.001) | (0.001) |
| Observations | 13,470 | 10,153 | 10,153 | 10,153 | 10,153 | 10,153 | 10,153 | 41,312 | 41,312 | 41,312 |
| Mean Dep Var | 4.868 | -0.001 | 0.044 | 0.027 | 0.021 | 0.338 | 0.114 | 0.098 | 0.0472 | 0.050 |
| R-squared | 0.100 | 0.098 | 0.099 | 0.060 | 0.068 | 0.103 | 0.072 | 0.129 | 0.092 | 0.088 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Municipality FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** This table shows the results from a set of regressions of different outcome variables on principal effectiveness. Sorting Index is defined à la Kremer and Maskin (1996) and reflects the amount of variation in classrooms' average course grades that comes from variation between instead of within classrooms. Parents' complaints refer to the number of complaints per 100 students issued by parents for different causes related to the management of the schools. Teacher turnover corresponds to the share of teachers who will leave the school the next year. All regressions include year and municipality fixed effects. Bootstrapped standard errors (100 replications) are clustered at the school principal level.

Table A.5: 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*** | 466.428 |
| | (186.517) | (351.519) | (0.000) | (352.286) | (349.506) | (0.001) | (482.246) |
| Δ Total Enrollment | -4.870 | -14.698 | -9.827*** | -16.734 | -13.428 | 3.306 | -5.057 |
| | (21.439) | (45.673) | (0.000) | (51.932) | (41.269) | (0.136) | (38.760) |
| Rural School | 0.838 | 0.256 | -0.582*** | 0.224 | 0.276 | 0.051** | 0.235 |
| | (0.369) | (0.437) | (0.000) | (0.418) | (0.447) | (0.015) | (0.424) |
| Income per student | 19.631 | 6.455 | -13.175*** | 6.471 | 6.446 | -0.025 | 8.094 |
| | (25.551) | (1.859) | (0.000) | (1.798) | (1.896) | (0.784) | (11.538) |
| Δ Income per student | 0.595 | -0.055 | -0.650* | -0.099 | -0.028 | 0.072 | 0.183 |
| | (15.703) | (1.173) | (0.080) | (1.739) | (0.590) | (0.209) | (10.365) |
| Share of disadvantaged students | 0.702 | 0.508 | -0.193*** | 0.493 | 0.518 | 0.026** | 0.410 |
| | (0.220) | (0.225) | (0.000) | (0.240) | (0.214) | (0.018) | (0.284) |
| Δ Share of disadvantaged students | -0.037 | -0.014 | 0.023*** | -0.010 | -0.016 | -0.006 | -0.018 |
| | (0.175) | (0.114) | (0.000) | (0.121) | (0.108) | (0.249) | (0.138) |
| 4rd grade test scores (Spanish) | 255.565 | 255.610 | 0.045 | 256.303 | 255.203 | -1.100 | 267.873 |
| | (30.287) | (22.027) | (0.961) | (21.944) | (22.077) | (0.344) | (26.145) |
| Δ 4rd grade test scores (Spanish) | -1.590 | -2.637 | -1.047 | -1.792 | -3.140 | -1.348 | -3.453 |
| | (33.854) | (22.127) | (0.308) | (21.959) | (22.223) | (0.251) | (25.680) |
| 4rd grade test scores (Math) | 238.975 | 245.459 | 6.484*** | 245.793 | 245.262 | -0.531 | 256.762 |
| | (33.034) | (24.510) | (0.000) | (24.177) | (24.715) | (0.682) | (33.265) |
| Δ 4rd grade test scores (Math) | 9.237 | 8.194 | -1.043 | 8.549 | 7.982 | -0.567 | 6.489 |
| | (34.392) | (23.280) | (0.324) | (22.527) | (23.727) | (0.646) | (26.613) |
| Graduation test score (Spanish) | 414.049 | 436.469 | 22.420*** | 439.418 | 434.231 | -5.188 | 513.328 |
| | (43.775) | (57.979) | (0.000) | (55.863) | (59.561) | (0.380) | (76.775) |
| Δ Graduation test score (Spanish) | -4.119 | -4.079 | 0.040 | -3.011 | -4.896 | -1.885 | -2.621 |
| | (29.886) | (25.310) | (0.989) | (21.255) | (28.042) | (0.475) | (25.492) |
| Graduation test score (Math) | 418.480 | 441.136 | 22.657*** | 441.347 | 440.977 | -0.370 | 516.988 |
| | (41.965) | (55.204) | (0.000) | (54.580) | (55.794) | (0.948) | (80.752) |
| Δ Graduation test score (Math) | -7.798 | -3.738 | 4.061 | -3.929 | -3.591 | 0.337 | -3.612 |
| | (31.307) | (25.322) | (0.155) | (21.378) | (28.017) | (0.898) | (25.072) |
| **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). Δ represents the first difference of the predetermined (pre-reform) school characteristic. 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.6: 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.7: 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).[29] 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 $\theta_i$ 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), $\beta_j$ 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, $\epsilon_{1ij}$ and $\epsilon_{0ij}$ represent factors that are not known to the observer but that influence the utility of school $j$ of hiring or not hiring worker $i$.

[29] 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 $\epsilon_{1ij}$ and $\epsilon_{0ij}$ 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 $\theta_i$, 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:

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

## Simulation

We are interested in the allocation of worker quality in the public and private sectors. More specifically, we seek to understand how the allocation of principal effectiveness in a given sector depends on the *selection* parameter $\beta_j$ 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:

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

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

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

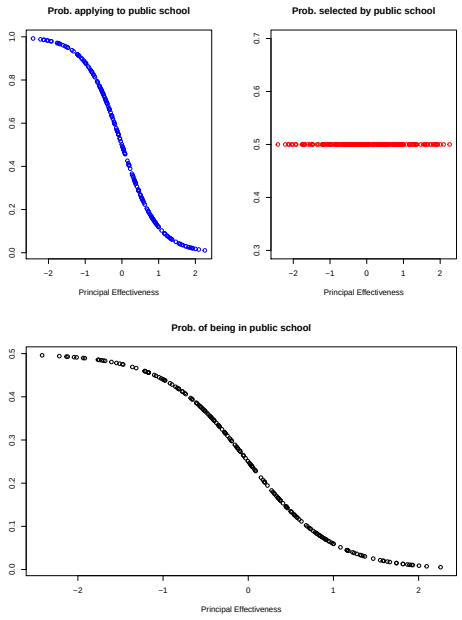
From Bayes' rule, we know that:

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

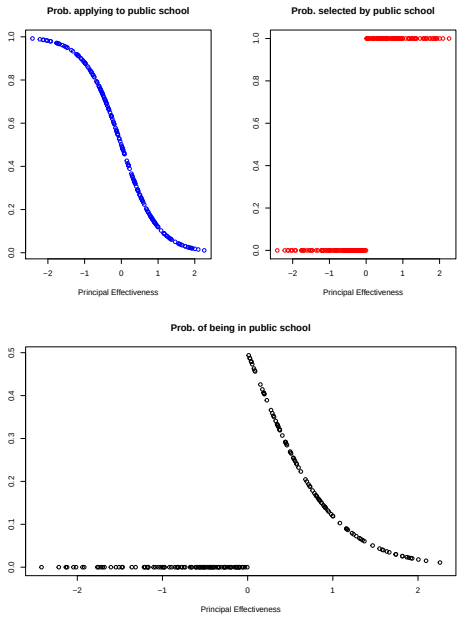
where $Pr(A_{ip} \mid \theta_i)$ is given by (12) and $Pr(\text{Public})$ is a scale factor equal to the fraction of public schools (0.5 in this case). Assuming that $f_\theta(\theta_i)$ is a standard normal, we can compute $\mathrm{E}[\theta_i \mid \text{Public}]$ using numerical integration. More importantly, we can study how this object depends on $\beta_p$ 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 $\beta_p$ and $w_p$ from 1 to 10, and compute $\mathrm{E}[\theta_p \mid \text{school type: Public}]$ for each cell of this grid.

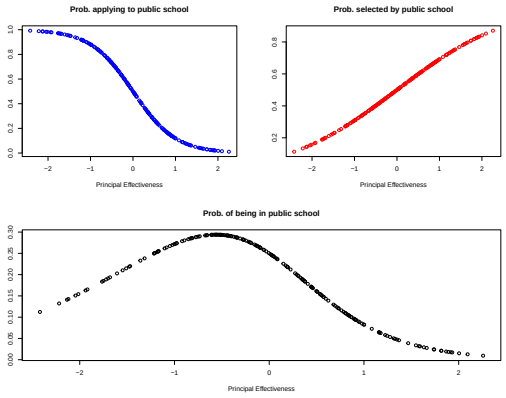
Figure 1: Simulation of a two-sided matching model



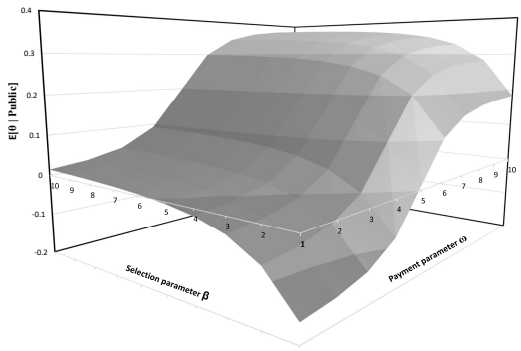

**A.** Selection at Random




**B.** Selection Above Cutoff




**C.** Selection on Quality




**D.** Allocation of $\theta_p$ 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 $\omega$ 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 $\beta_p$ and $w_p$ from 1 to 10, and compute $\mathrm{E}[\theta_p \mid \text{school type: Public}]$ for each cell of this grid.

# 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 school workers between 2008 and 2017. These records include 13,693 unique schools and 331,167 unique workers. 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).

We complement this data with records from the teachers' evaluation system. The Chilean evaluation system operates on the basis of four sources of evidence: a portfolio, an interview by a peer teacher with at least five years of experience, a written report of two school authorities on the basis of a set framework and a self-evaluation report by the teacher

following a given structure. Among the instruments the portfolio has the highest weighting in the process of establishing the competence level of the teacher being evaluated (60%), followed by the peer interview with 20% and the other two sources of evidence with 10% each. Based on this information teachers are classified in four performance categories: "outstanding", "competent", "basic" or "unsatisfactory". For details see Avalos-Bevan (2018).

**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.

**Complaints against the schools:** The Superintendency of Education provided access to all complaints filed against the school between 2014 and September 2018. These records have the number of complaints by category. The categories include: i) bullying and discrimination (also includes behaviors of sexual connotation against students or teachers), ii) denied enrollment (for instance because of disciplinary measures), iii) poor infrastructure (includes lack of furniture), iv) teacher absenteeism (or lack of teachers), v) school

accidents, vi) charge of extra fees (or ask for extra materials), vii) resource accountability (irregularities in the use of vouchers or misreporting of attendance).

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

# References


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

Avalos-Bevan, B. (2018): "Teacher evaluation in Chile: highlights and complexities in 13 years of experience," *Teachers and Teaching*, 24, 297–311.

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. (2021): "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*.

Kremer, M. and E. Maskin (1996): "Wage inequality and segregation by skill," 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.

Rambachan, A. and J. Roth (2021): "An honest approach to parallel trends," .

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

Roth, J. (2021): "Pre-test with caution: Event-study estimates after testing for parallel trends," .