EIEF WORKING PAPER SERIES

IEF

EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 23/03

May 2023

Managers' Productivity and Recruitment in the Public Sector

By

Pablo Muñoz
(Universidad de Chile)

Mounu Prem
(EIEF)

Managers' Productivity and Recruitment in the Public Sector*

Pablo Muñoz   Mounu Prem

Abstract

Governments face many constraints in attracting talented managers to the public sector, which often lacks high-powered incentives. 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. First, we estimate principal effectiveness by using an extension of the canonical teacher value-added model. Then, we evaluate the effect of the reform on principal effectiveness using a difference-in-differences approach. We find that public schools appointed more effective managers and improved their students' outcomes after increasing the competitiveness and transparency of their selection process.

Keywords: Managers, Public sector, Recruitment, School principals

*This version: April, 2023. 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, MINEDUC, and DEMRE on for their help with the data. 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, Department of Industrial Engineering, Universidad de Chile. The views expressed here are those of the authors and do not necessarily reflect the views of these individuals or institutions. Email: pablomh@uchile.cl. Mounu Prem, Einaudi Institute for Economics and Finance. Email: francisco.munoz@eief.it

1 Introduction

Effective management is a key resource in both private and public enterprises (Bloom et al., 2013, 2015b), but identifying and recruiting effective managers remains challenging in the public sector.[1] Empirical research in this area has faced at least two important hurdles. First, the dearth of data makes it difficult to objectively measure public sector managers' performance. Second, it is hard to find quasi-experimental variation in the allocation of public sector positions.

We overcome these limitations by focusing on "street-level" bureaucrats (Besley et al., 2022) and by leveraging the institutional setting of Chile, which features rich administrative data. Specifically, we study the case of school principals, managers with high practical relevance for the delivery of public services, and thus for state capacity. The Chilean educational context is appealing for three reasons. First, we can construct an objective measure of principals' effectiveness based on students' performance. Second, the publicly funded but privately run schools provide a benchmark for evaluating the country's public system (Hsieh and Urquiola, 2006). Third, variation from a recent civil service reform allows us to assess the impact of more competitive and transparent personnel selection policies on the effectiveness of public school principals.

To reduce politicians' discretion over the appointment of school principals, Chile enacted a reform in 2011 that modified the selection of school leaders.[2] Under the new selection system, local politicians still have a say in principals' appointments, but only after a competitive and transparent competition creates a shortlist of candidates based on merit and suitability. We assess the impact of this reform in two steps. First, we estimate a course grades value-added model that disentangles the contribution of a school principal from that of her teaching staff and other school-related factors. For brevity, we refer to the principal's course grade value-added level as principal effectiveness. We find that a one standard deviation increase

[1]Many public settings feature pervasive discretionary appointments and patronage (Xu, 2018; Colonnelli et al., 2020), while incentive schemes are hard to define and mostly absent (Lazear and Shaw, 2007; Finan et al., 2017).

[2]Since the 1980s, the recruitment of public school principals in Chile had been the exclusive responsibility of the municipalities. This gave local politicians a significant degree of discretion over the appointments of school personnel, a feature that, in a similar context, has been associated with negative effects on student outcomes (Akhtari et al., 2022).

in principal effectiveness raises students' course grades by 0.27 standard deviations. Then, we use a difference-in-differences approach to compare the change in principal effectiveness arising from new appointments in public schools under the reformed selection system to that arising from new appointments at private schools. We find that public schools attracted more effective managers—a 0.06 standard deviation increase—after changing their selection process, leading to course grade gains of 0.016 (0.27×0.06) standard deviations per year. We also document that these schools increased their likelihood of firing poorly evaluated teachers and experienced positive impacts on students' test scores, yearly attendance, and college admission scores.

To incorporate principals' effects in the teacher value-added model, we estimate a two-way fixed effects specification (Abowd et al., 1999) that leverages teachers' and students' switches (i.e., between schools) across principals, to identify principals' and teachers' effectiveness separately. To distinguish the effect of the principal from other school-level factors, we also include school-level controls in a correlated random effects fashion (e.g., Altonji and Mansfield, 2018). We perform several exercises to show that our measure of principal effectiveness is sensible. In the spirit of Chetty et al. (2014), we show that student achievement changes sharply after principal turnover, *as predicted* by the change in our measure of effectiveness. We also provide evidence in favor of our additive principal and teacher effect specification (Card et al., 2013), and we show that our correlated random effects approach works well at removing selection coming from students who deliberately sort into school principals (Jackson et al., 2022). Building on the conceptual framework of Leaver et al. (2019), we also study *how* principals matter, demonstrating that more effective principals have: i) a larger fraction of their teaching staff agreeing with positive statements about them, ii) fewer parental complaints about bullying and denying enrollment, and iii) lower rates of teacher turnover, especially among high value-added teachers.

To assess the impact of Chile's recruitment reform, we estimate a difference-in-differences model that compares changes in principal effectiveness arising from new appointments in public schools under the reformed selection system to similar changes arising from new appointments at private schools. Overall, we find that limiting the discretion of local politicians over the appointment of public school managers increases principal effectiveness by 0.06 standard deviations. We complement our analysis with a battery of robustness checks. First, we show that the dynamic version of our difference-in-differences approach, alongside recent developments in the literature (e.g., Rambachan and Roth, 2021; Roth, 2022), supports our identification strategy. Second, we show that our results are quantitatively similar—and robust to recent theoretical advances (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2020)—if we only keep public schools in the estimation. Third, placebo event studies show that before the civil service reform was enacted, principal turnover did *not* affect principal effectiveness deferentially among public and private schools.[3]

To address the concern that course grades might be susceptible to manipulation, we show that the school-level correlation between test scores and grades does not change upon a new principal's arrival. Moreover, since no single measure effectively captures every skill students need in adulthood (Jackson et al., 2020), we show that the impact of the reform is similar when using other measures of principals' effectiveness obtained from value-added models of students' test scores, absenteeism, contemporaneous (instead of future) course grades, and from a model that excludes teachers' fixed effects. Finally, we leverage our difference-in-differences approach and show that the reform led to *direct* improvements in course grades, test scores, and yearly attendance. It also increased the turnover of the management team, the likelihood of firing poorly evaluated teachers, and led to better college admission scores among applicants, suggesting positive long-run effects of the selection policy.

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., 2017; Besley et al., 2022). Recent studies have shown that patronage is a common feature in public sector appointments (Xu, 2018; Akhtari et al., 2022; 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., 2022; Moreira and Pérez, 2021).[4] We contribute to this literature by focusing on managers and showing that a more transparent selection system leads to appointing more effective school principals. Second, our work complements a growing literature on the importance of effective management and management practices in public sector organizations (McCormack et al., 2014; Bloom et al., 2015b; Lavy and Boiko, 2019; Rasul and Rogger, 2018). In the same vein as Janke et al. (2019), Fenizia (2022), and Muñoz and Otero (2022), we construct an objective measure of managers' performance, but in the context of schools. By analyzing data on teachers' perceptions and school management, our work also relates to research on management practices and students' learning (Bloom et al., 2015a; Leaver et al., 2019). Finally, our paper contributes to the economics of education literature (Eberts and Stone, 1988; Clark et al., 2009; Béteille et al., 2012; Branch et al., 2012; Coelli and Green, 2012;

[3]To partially address the concern about spillovers between public and private schools, we also show that our results are similar when comparing schools within the same school market, as defined in Cuesta et al. (2020) and Neilson (2019).

[4]In this line, 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.

Grissom et al., 2015), and complements studies on school personnel (Rothstein, 2015; Cullen et al., 2016; Biasi, 2021; Loyalka et al., 2019; Brown and Andrabi, 2020; Leaver et al., 2021) by analyzing a setting where high-powered incentives are hard to implement.

The rest of the paper proceeds as follows. Section 2 describes the recruitment reform and data. Section 3 presents the main value-added model used to measure principal effectiveness and discusses its validity. Section 4 examines the effects of the recruitment reform on principal effectiveness and school outcomes; it also offers several robustness checks. Finally, Section 5 concludes with a discussion of our findings.

2 Background and Data

In 1981, Chile implemented an educational reform that privatized and decentralized both primary and secondary education. Publicly funded school vouchers were created, with flat voucher funds following children to any public school or private school that agreed to accept the voucher as tuition payment. 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 expenses covered by the voucher.[5] Private, subsidized private, and public schools represented 7, 53, and 40 percent of enrollment, respectively, in 2018. Along with privatizing the educational system, this reform also decentralized it by transferring control of public schools from the central government to municipal authorities. Departments of Education were created to manage the public schools; the efficacy and probity of these departments were strongly related to those of their local governments (Guerra and Arcos, 2012).

The Recruitment Reform: In 2011, the country enacted a law to improve quality and equity in education (Law Nº 20,501). The law recognized school principals as key agents in improving quality in public schools and created a new system to appoint them. Before the reform, the appointment of public school principals was the exclusive responsibility of the schools' municipalities, and the process was unsupervised by the central government. Since the reform, principals have been elected through public, competitive, and transparent contests overseen by the Civil Service.[6]

[5]A major change to the voucher policy program was introduced in 2008 when a new voucher targeting low-income students was created (Neilson, 2019). Since the share of poor students is larger in public schools than in private schools, this led to larger subsidies for public schools.

[6]The creation of the Civil Service in Chile responded to a large corruption scandal (known as the MOP-Gate case) that exposed illegal payments to top government officials in 2003. Law Nº 19,882 regulated the public sector's personnel policies and created the service, with the aim "to provide government

In the context of the 2011 education reform, the Civil Service was 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." These contests must be disseminated in newspapers and by the Civil Service. As can be seen in the website "School Principals for Chile" (www.directoresparachile.cl), calls are widely advertised and all the job information is available to potential candidates. After receiving the background information, an external human resources company is hired to pre-select candidates. This process includes a curricular analysis and psycho-labor evaluations. Then an independent qualifying commission, including a representative of the Civil Service, the head of education of the municipality, and an outstanding teacher, conducts interviews with the candidates shortlisted by the company. After these interviews, the qualifying commission defines a list of 3 to 5 finalists; this list is sent to the mayor, who makes the final hiring decision. The whole contest, from the announcement to the definition of the shortlist of candidates, takes approximately 100 days.[7]

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 service period of five years (Ruiz-Tagle, 2019). However, no clear sanctions were established to handle non-compliance with this provision (Errázuriz et al., 2016), and not all processes concluded with an appointment. According to records from the Civil Service, of the 4,305 competitions held between 2011 and 2017, 63.7% resulted in appointments, 23.4% were resolved as abandoned, and 11.3 % were overridden. The rest did not conclude with an appointment for admissibility problems or because it was impossible to carry out the external evaluation stage (Ruiz-Tagle, 2019).[8]

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 teaching evaluation; and finally, iii) get a bonus consistent with their responsibilities, based on the total number of students enrolled and the

institutions—through public and transparent competitions—with executives with proven management and leadership capacity to execute effectively and efficiently the public policies defined by the authority."

[7]For 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/).

[8]The Civil Service and the municipalities have to incur costs associated with running the contests (Silva, 2014). These costs include fees to be paid to the headhunter companies and the expenses of the advertisement. According to a Civil Service report (Civil-Service, 2015), the median cost of a recruitment process similar to those we studied is 8,500 USD. Regarding media expenses, the Civil Service spent 31,000 USD on press coverage related to selection processes in education in 2018. For details, see https://www.serviciocivil.cl/presupuesto-y-otros.

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 superintendents, 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 Nº 20,501 Calidad y Equidad de la Educación."

The school system: Chile has a nationwide standardized curricula.[9] The Chilean curriculum is determined by the Ministry of Education for each grade and subject, directly affecting each school's curricular offerings and instructional resources. It also works as an accountability system (Valverde, 2004). As can be seen in the "National Curriculum" website (https://www.curriculumnacional.cl), the government not only provides teachers with curriculum guides and official textbooks, but also with lesson plans and exams. Curriculum guidelines establish minimum content goals and fundamental educational objectives, ultimately determining 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 to students in the 2nd, 4th, 8th, and 10th grades. However, SIMCE has not been systematically run every year in the country. Finally, to gain admission into higher education, most students take a standardized college entrance exam known as PSU.[10] Students must complete mathematics and language exams, and many take optional tests in history and sciences. 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).

[9] Other countries with a national curriculum include France, Hungary, Ireland, Italy, Japan, Korea, the Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, and the United Kingdom. While most public schools in the U.S. follow "Common Core" guidelines for their core curriculum, there is no national curriculum.

[10] Chilean entrance exams are more like the ACT than the SAT, in that Chilean exams are oriented to measure how much of the school curriculum has been learned rather than mainly measuring cognitive aptitude (González Adonis et al., 2017). As shown by Panel A of Appendix Figure A.1, a one standard deviation increase in course grades was associated with a 0.8 standard deviation increase in college admission scores.

2.1 Data and Descriptive Statistics

To estimate principals' effectiveness, we use a panel of data at the student-year-subject level. This panel spans from 2011 to 2016 and has information on subject-specific course grades for all students from grades 1-12. For cohorts of students that take standardized exams, it is also possible to link our data to their test scores in Math and Spanish. We match this data with a nationwide census of teachers containing information on the specific subjects and classrooms they teach every year, and their characteristics. For a subset of these teachers, we recover their perceptions about the school principal from survey responses. We also leverage data from a yearly school panel that includes several school characteristics such as the type of administration (e.g., public or private), an indicator for schools located in a rural area, the school's total enrollment, the fraction of disadvantaged students, the fraction of parents with a college degree, parents' income level, the identity of the school principal, and whether she was elected through the new selection system. We complement this data with characteristics of the municipalities where the schools are located.

Table 1 presents detailed descriptive statistics of the students, schools, and principals at public and private schools. Panel A 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 percentage points less likely to fail a grade. 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 student attendance. Some of these differences are likely related to the fact that only 21% of private schools are in rural areas versus 62% of public schools. Regarding school finances, public schools receive a larger subsidy, but they also serve more disadvantaged students. The share of students considered poor and eligible for special subsidies is 57% in public schools and 32% in private schools. Finally, Panel C of Table 1 shows that, compared to private schools, public schools pay lower wages to their 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 are 51, 26, and 24 percent, respectively.[11] Regarding demographic characteristics, public school principals have more tenure, and most are male; this is in contrast to private schools, where 60% of principals are female.

[11] We classify compensation items into three categories: basic wage, statutory payments, and bonuses (see Appendix D for details). On average, principals earn around 2,739 USD per month, which corresponds to roughly 11 times the legal minimum wage.

3 Measuring Principals' Effectiveness

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

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

where Y_{it+1} is the course grade obtained by student i in year $t+1$, γ_t and ρ_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) of the lagged dependent variable interacted with students' grade level (Kane and Staiger, 2008; Chetty et al., 2014). X_{it} controls for the time-variant variables students' age and principals' tenure. Since we include teachers, we focus on future 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 (Jackson, 2018).[12] We add teacher fixed effects and principal fixed effects to account for confounders. For instance, a school that tends to attract both high-quality teachers and high-quality principals may lead to an upward-biased estimate of principal quality if teachers are not considered.[13] Since this modeling decision comes at the cost of potentially muting the teachers' incentive and selection channels through which principals can impact schools (Leaver et al., 2019), we also show the robustness of our result to alternative specifications without teachers' fixed effects.

In light of recent research (Jochmans and Weidner, 2019; Bartanen and Husain, 2022), we do not include school fixed effects as we only observe a subset of principals switching between schools.[14] Instead, we use correlated random effects (Mundlak, 1978; Chamberlain, 1980) to account for school heterogeneity. Specifically, we include a combination of time-varying and across-time averaged characteristics for each school s (i.e., $X_{s(i,t)}$ and $\bar{X}_{s(i,t)}$). 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.

[12] Intuitively, our specification gives credit to a math teacher if her students improved their math course grades after having her as a teacher, controlling for the students' past achievement. This avoids confounding easy graders with high-value-added teachers.

[13] This is particularly relevant in public schools, where principals have limited scope to make personnel decisions, since the employer is the municipality and personnel decisions are not subject to the private labor code.

[14] In our main estimation sample, the connected set with the largest number of movers only has 6. This set includes 11 principals and 7 schools (9,301 student-level observations).

Fixed school characteristics include the across-time average of the previous list of time-variant characteristics, plus indicators for whether the school is public or subsidized private, and for whether it is located in a rural area.[15] Following Wooldridge (2010), we also add the across-time average of the year fixed effects dummies to account for our unbalanced panel. This approach is predicated upon the idea that the observed school-level factors we control are correlated with the unobserved school-level factors we cannot control for. As in Altonji and Mansfield (2018), the goal is to absorb the across-school variation in unobservable school characteristics using each school's 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 student's performance in the two subjects for which we observe course grades every year: Math and Spanish. Our analysis excludes preschools, adult education schools, and special education schools. We also excluded classes that had more than one teacher per year and eliminate the bottom and top one percent of classroom size outliers. Since the teacher $\mu_{j(i,t)}$ and principal $\theta_{p(i,t)}$ fixed effects in Equation (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 7,735,653 student-subject-year observations, corresponding to 1,977,203 students, 64,770 teachers, and 8,061 principals.[16]

We disentangle principals' and teachers' effectiveness by leveraging teachers' and students' switches within the largest connected set. Close to the identification argument developed in the seminal work by Abowd et al. (1999), between-principal mobility is essential to identify principal and teacher effectiveness separately from our value-added model. Figure 1 presents the empirical distributions of the principal and teacher fixed effects estimates.[17] The adjusted standard deviation of principal effectiveness corresponds to 0.27,[18] namely, one standard deviation in principals' effectiveness is associated with a 0.27 standard deviation increase in students' course grades. This figure is larger than previous estimates reported for school principals (Grissom et al., 2021). Still, it is not disproportionately larger than those reported

[15]To address concerns related to the politically motivated appointment of school principals (Colonnelli et al., 2020; Akhtari et al., 2022), we add party turnover, vote margin, and a measure of the concentration of vote shares (the Herfindahl-Hirschman Index) as municipal-level controls to our model, and find that this measure of effectiveness is strongly correlated (0.997) with the main one.

[16]As shown in Appendix Table A.1, we do not find evidence of selective sample attrition regarding grades, subject, attendance, student performance, or teachers' characteristics.

[17]The standard deviations of the principal and teacher fixed effects are 0.39 and 0.41, respectively.

[18]The standard deviation of $\hat{\theta}_{p(i,t)}$ can be upward biased due to sampling error. Thus, we compute the adjusted standard deviation of principal effectiveness $\hat{\theta}_{p(i,t)}$ using the formula proposed in Krueger and Summers (1988) and we shrink our estimates of principal effectiveness for the analysis.

for teachers, e.g., Kane and Staiger (2008) documents standard deviations in the range of 0.18 to 0.20.[19] We present the robustness of our main difference-in-difference result to this and other alternative measures of principal effectiveness in Section 4.2.

In Appendix B, we address several concerns related to our specification and the identification of principal effectiveness. We begin with event studies looking at the evolution of course grades around the events of entry and exit of low-value-added and high-value-added principals, following Chetty et al. (2014). We find that, upon principals' turnover, the change in course grade gains is significantly different from 0 and is not significantly different from what one would forecast based on the change in our measure of principal effectiveness. Since our model considers additive teacher and principal effects, we also perform specification checks 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). To assuage concerns related to students sorting into schools based on school principals, we follow Jackson et al. (2022) and show that conditional on our school correlated random effects, predicted course grades based on students' characteristics are unrelated to our measure of principal effectiveness. Moreover, in the spirit of Rothstein (2010), we focus on a subset of students who switched schools and who were consequently exposed to more than one principal, and show that the effectiveness of the students' future principal (the *treatment* variable) is not correlated with the current growth in their course grades (the *pre-assignment* variable). Finally, we document how estimated principal effectiveness correlates with observable characteristics. To do so, we regress the estimated principal fixed effects on age, age squared, gender, an indicator for holding a college degree, and an indicator for having experience in previous "schooling type" positions. As shown in Appendix Table A.2, female principals appear to be, on average, more productive than their male counterparts. Principal effectiveness is also strongly correlated with experience, a feature for which we find a concave profile at public schools.

3.1 Principals' Effectiveness and the Management of Schools

Before concluding this section, we present correlations between our measure of principals' effectiveness and different outcomes related to management practices (Bloom et al., 2015b; Di Liberto et al., 2015; Lemos et al., 2021). This exercise is motivated by the conceptual

[19] It is worth noting that previous estimates reported for school principals consider test score value-added models without teacher effects. In our setting, the adjusted standard deviation of principal effectiveness when using test scores in a model without teachers' fixed effects is 0.15, a magnitude within the range of previous comparable estimates (e.g., Dhuey and Smith, 2018).

framework proposed by Leaver et al. (2019), whereby better managers affect students' learning because different actors such as teachers, students, and parents become more productive (the *incentive channel*), or because new actors join the school (the *selection channel*). To assess the incentive channel, we leverage data from various sources and evaluate whether more effective principals: i) are recognized as such by their teaching staff, ii) rely more on the sorting of students across classrooms, and iii) receive fewer complaints from parents. To assess the selection channel, we study the association between principal effectiveness and teachers' turnover, distinguishing the data by teachers' value-added levels. Finally, to assess the role of operations management, we also explore the relationship between principal effectiveness and school finances.

Teachers' Perceptions: 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 strong disagreement with the statement and 4 (or 5) represents a strong level of agreement. We use their responses—for the years 2010, 2011, 2014, and 2015—to create a dummy variable at the survey-respondent level that equals one if the teacher "strongly 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 regression of the fraction of teachers strongly agreeing with a given statement on our estimated measure of principal effectiveness. Figure 2 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 strongly agreeing with positive statements about their management style. Ordered by effect size, we find that a one standard deviation increase in principal effectiveness increases agreement with the statements *the principal engages teachers*, *the principal knows teacher needs*, *the principal engages parents*, and *the principal knows students needs* by around 6%; *the principal makes good decisions*, *the principal includes teachers*, *the principal is effective*, and *the principal does a good job* by around 5%; and *the principal promotes a good work climate*, *the principal is good at communicating*, and *the principal can be trusted* by around 3 to 4%. Table A.3 in the Appendix presents our point estimates and shows the robustness of these results to accounting for multiple hypothesis testing, following Romano and Wolf (2005).[20]

[20] As an additional robustness check, we perform 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

Student Tracking: School principals can impact schools' outcomes through different margins, including tracking students, i.e., sorting students with different achievement levels into different classrooms (Duflo et al., 2011; Card and Giuliano, 2016). To assess the extent of tracking in Chile, and to explore how principal effectiveness is related to it, we construct a sorting index at the school-year level à la Kremer and Maskin (1996).[21] Intuitively, perfect sorting is a case in which all variation in classrooms' average course grades comes from variation *between* classrooms instead of *within* classrooms.

Figure 3, Panel A, plots a binned scatter showing 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 0.06 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 0.71. Thus, sorting based on course grades represents only 8% of our empirical upper bound (0.06/0.71).

Parents' Complaints: We also explore how principal effectiveness relates to parents' complaints. To do so, we leverage administrative data used to monitor how well schools comply with the laws and regulations 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 about bullying/discrimination, denial of enrollment, poor infrastructure, teacher absenteeism, and school accidents.

Figure 3, Panel B, plots a binned scatter showing the relationship between the Z-score of the number of different complaints (per 100 students) and principal effectiveness. As shown in 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 0.02 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.

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).

[21] 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: τ_{st}.

Teacher Turnover: We next study the relationship between principal effectiveness and the turnover rate of teachers working for them.[22] More specifically, we define our outcome variable as the share of teachers that leave the school run by principal p at time t, either because of a job-to-job or a job-to-unemployment transition at time $t + 1$.

Figure 3, Panel C, plots a binned scatter showing the relationship between teacher turnover and principal effectiveness. To further explore this dimension, in Panel D, we plot the 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 teacher turnover. Considering that the average turnover in our sample is 0.12, 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 21%. This finding is consistent with the idea that principals can recognize good teachers (Jacob and Lefgren, 2008), but should be interpreted with caution as the covariance between principal and teacher fixed effects might suffer from bias (Kline et al., 2020).

School Finances: Finally, we assess the relationship between principal effectiveness and different school finance measures. To do so, we use detailed records on school income and expenditures, which are only available for 2016. As shown in Appendix Table A.5, we cannot reject (at standard confidence levels) the null hypothesis of zero association between principal effectiveness and different school finance measures. This aligns with the fact that school owners, not necessarily principals, are the main ones responsible for schools' budgets and financial decisions in Chile.

4 The Impacts of the Recruitment Reform

Generally, the compensation of public school principals is rigid and based on statutory payments. In Appendix C, we illustrate this point with a two-sided matching model and provide a descriptive analysis of principals' wages at public and subsidized-private schools in Chile. Our estimates reveal a wage premium of 14% in subsidized-private schools, most of which

[22]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).

is driven by the bonus components of wages. A modest (although statistically significant) association between principal effectiveness and wages is also detected at subsidized-private schools, again driven by the bonus components of wages. Our estimates also show that the tenure profile is salient at public schools and that the size of the gender wage gap is almost 11% at subsidized-private schools but close to zero at public schools.[23] 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 what follows, we study the impact of a particular strategy: civil service recruitment.

4.1 Selection and Recruitment of Public School Principals

We exploit the non-eligibility of private schools and the timing of adoption of the new selection system within public schools to study the impact of this 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 when a new principal was appointed under the new selection process. As shown by Figure A.2 in the Appendix, the adoption of this system was staggered. This is because replacing principals was not mandatory and contests did not always succeed at appointing a principal (see Section 2 for details). As expected, the number of principals elected under the new regime increased over time, with around 370 new principals elected yearly since 2012.

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

$$\begin{aligned}\hat{\theta}_{st} = \alpha_s + \alpha_t + \beta_1 \times ADP_{st} \times \text{ Principal Turnover}_{st} \\ + \beta_2 \times \text{Principal Turnover}_{st} + \sum_t \Phi'_t X_s I[year = t] + \epsilon_{st},\end{aligned} \tag{2}$$

where s and t stand for school and year, and the dependent variable $\hat{\theta}_{st}$ corresponds to the standardized version of our measure of principal effectiveness. Principal Turnover$_{st}$ is a dummy variable that equals one for the first year (after 2012) when the school selected a new principal, and ADP_{st} is a dummy variable that takes the value one for the first year a public

[23]This finding is in line with recent evidence by Biasi and Sarsons (2022) showing that flexible pay reforms can increase the gender wage gap.

school appointed a principal using the new selection system.[24] X_s is a vector of predetermined (as of 2010) school characteristics, including income per student, the share of disadvantaged students, total enrollment, and 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 parameterized by these school and municipality characteristics. Finally, α_s and α_t are school and year fixed effects, and ϵ_{st} is an error term robust to heteroskedasticity and clustered at the school level. The parameter of interest is β_1, and it captures the difference between 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.[25]

The key identification concern in our setting is that—conditional on time-invariant school characteristics, year aggregate shocks, and differential trends parameterized by pre-reform school and municipality characteristics—there might still be unobserved confounding factors that correlated with the timing of adoption of this new system and other determinants of principal effectiveness. To partially address this concern, we estimate an extension of model (2) with a dynamic treatment. Specifically, we estimate the following regression:

$$\begin{aligned}\hat{\theta}_{st} &= \alpha_s + \alpha_t + \textstyle\sum_{j=-4}^{-2} \beta_j \times ADP_s \times I[k=j] + \sum_{j=0}^{4} \beta_j \times ADP_s \times I[k=j] \\ &\quad + \textstyle\sum_{j=-4}^{-2} \delta_j \times I[k=j] + \sum_{j=0}^{4} \delta_j \times I[k=j] + \sum_t \gamma_t' X_s I[year=t] + \epsilon_{st},\end{aligned} \tag{3}$$

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.

Figure 4, Panel A presents our estimates of the β_js coefficients, which capture the difference in principal effectiveness in period j (relative to the omitted period -1) for schools that had principal turnover via the new recruitment system, relative to the difference in principal effectiveness (in period j relative to the omitted period -1) for private schools that experienced principal turnover.[26] The figure shows that the new system increased principal effectiveness

[24] In Chile, this new selection system is known as ADP, an acronym for *Alta Dirección Pública*.

[25] For those schools that had 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.

[26] 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 we observe for the 8-year window.

by 0.06 standard deviations, on average. It also provides visual support for our identification strategy, as point estimates are around zero and insignificant 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 suggests that our results are not reflecting reversion to the mean. Indeed, in the post-period, we observe an increase in principal effectiveness that remains stable over time. As shown by Panel B, we find similar patterns when flexibly controlling for pre-reform school and municipality characteristics.

We complement this test by computing the pre-trend that has an 80% power of being detected given the precision of the estimates in the pre-period, along with an 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, 2022). In Panel A of Appendix Figure A.3, we present the same figure as before but add these two trends. For both trends, the average bias that they can create represents 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 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.[27] 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 an 80% power of being detected given the precision of the estimates in the pre-period (Roth, 2022), which is 0.013. In Appendix Figure A.3, Panel B, we present the results. For both linear and non-linear deviations, we find that the confidence set at 90% does not include zero, which suggests that our results are robust to moderate deviations from the parallel trends assumption.

As an additional robustness check, we estimate Equation (3) again, but now we only consider observations from the pre-reform period: 2008-2011. In this case, we compare the effect of principals' turnovers between public and private schools. Intuitively, this placebo specification allows us to assess whether a new principal's appointment at a public school impacted principal effectiveness (relative to new appointments at private schools) when the new selection process was *not* in place. Naturally, since we consider data from the pre-reform period, we must use a shorter window around principal turnover and restrict our sample to schools with principals for whom we estimated a measure of effectiveness. Panel C of Figure 4 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

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

schools had no significant impact on principal effectiveness. If anything, when compared to principals' turnover at private schools, turnover at public schools led to a decrease in principal effectiveness.

We now turn to our parametric difference-in-differences estimates obtained from Equation (2). Table 2 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 principal elected under the new recruitment system, increases principal effectiveness by 0.06 standard deviations.[28] We find a non-significant negative change in effectiveness after principal turnover at private schools. Column 2 shows that controlling flexibly by school and municipality characteristics during the pre-reform period does not affect the significance or the effect size of our estimates.[29] In Column 3, 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 the new system. We also estimate an effect of 0.06 standard deviations using this truncated sample.[30]

In Column 4, we present the results when we only keep public schools for the estimation, therefore identifying the effect of the new selection policy from variation in the timing of adoption. In this case, never-treated units are public schools that changed their principal after the civil service 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 abandoned.[31] We find similar results in this case: schools that selected a principal using the system experienced a statistically significant increase in principal effectiveness of 0.045 standard deviations when compared to *never* and *late* adopters. Panel D of Figure 4 plots the point estimates and 95% confidence intervals obtained from the dynamic model given by Equation (3) when we only consider public schools for the estimation.

[28]We find similar results if we use the un-adjusted principal fixed effects as our dependent variable and weight the regressions using the inverse of the estimation error of each fixed effect, as in Card and Krueger (1992). See Appendix Table A.6 and Figure A.4.

[29]We find similar results if we estimate an instrumental variable version, where we use as an excluded instrument a Post-2011 dummy interacted with an indicator for public schools for the term $ADP_{st} \times Principal\ Turnover_{st}$ in Equation (2). We find a point estimate of 0.043 with a p-value of 0.068 and an F-statistic for the first stage of 6,929.

[30]In Appendix Table A.7, we explore the heterogeneous impacts of the reform by how rural the school area is. We find no strong evidence for heterogeneous effects along this dimension.

[31]Appendix Table A.8 compares public schools that did not adopt the new selection system to never, early, and late adopters, as well as to private schools. We document that adoption is positively associated with school size and urban status, two features we control for in Equation (1).

4.2 Robustness Checks

New Developments in Two-way Fixed Effects Estimation: Recent literature on this type of two-way fixed effects estimation has 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 Appendix Figure A.5). Together, these results suggest that the concerns regarding this staggered difference-in-differences estimation should be minor. However, as a robustness check, in Columns 5 and 6 of Table 2, 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 0.040.[32]

As additional robustness checks, we perform two placebo exercises. First, we consider any principal turnover before the reform (2010-2011) in public schools as a treatment (Column 7). Second, we consider any principal turnover after the reform (post-2011) in private schools as a treatment (Column 8). In both cases, the placebo treatment takes the value one if there is a change in the principal and stays as one afterward. We find that turnover does not increase principal effectiveness in these placebo exercises. If anything, we find that turnover is associated with a decrease in principal effectiveness in the private sector. As shown in Appendix Table A.9, these findings also remain unchanged if we consider the models proposed by De Chaisemartin and d'Haultfoeuille (2020) and Callaway and Sant'Anna (2020).

Alternative Measures of Principal Effectiveness: We estimate alternative measures of principal effectiveness using standardized test scores as the dependent variable in our value-added model (see Section 3). The nationwide standardized exams are only taken by students in specific grades, and they have not been systematically administered every year in the country. Thus, when using test scores, we focus on a sub-sample of students who took the standardized exam in fourth grade and for whom their performance on the second-grade exam is available. Moreover, since teachers may change from one year to another, we do not

[32] In Appendix Figure A.6, we present the dynamic versions of these two-way fixed effects models that leverage variation within public schools.

include teachers' fixed effects in this model. To account for what test scores miss (Jackson, 2018), we also estimate a measure of effectiveness based on students' absenteeism. Since yearly student attendance is not available at the subject level, teachers are also excluded from this value-added model. Finally, we also estimate principals' course grade value-added from a model that excludes teachers and from a model that uses contemporaneous (instead of future) course grades.

Table 3 presents the difference-in-differences estimates obtained from Equation (2) when using these alternative measures of principal effectiveness as a dependent variable. In all cases, we find that—relative to the effect of principal turnover on principal effectiveness at private schools—the turnover at public schools due to appointing a new principal elected under the new recruitment system increased principal effectiveness. The positive and significant effect on principal effectiveness ranges between 0.07 (model with course grades but without teacher effects) to 0.15 (model with test scores but without teacher effects). We also find that there are negative changes in effectiveness after principal turnover at private schools. Figure 5 (Panels A-D) presents the estimates from the dynamic difference-in-differences specification, which provides further visual support for our identification strategy, as the point estimates are around zero and not significant in the pre-reform period.

Spillovers Between Public and Private Schools: To assess concerns related to potential spillovers of the policy, we control for school markets. We proceed in two different ways. First, we consider the municipality in which the school is located. Second, we follow Cuesta et al. (2020) and Neilson (2019) and define a market based on the distance between schools. In Chile, the students' average distance to their chosen schools is 2km, and the 90th percentile of such distribution is 5km (Cuesta et al., 2020). Following these papers, we define a market based on the schools that are close to each other and use a diameter (k) equal to 4, 5, 6, and 7km. We implement this by creating a symmetric adjacency matrix (A), where the element $A(i, j)$ takes the value one if i and j are less than k km away from each other. Then, we construct the set of "connected components" of the matrix, where a component is defined as a set of schools where one can always find a "path" that connects two pairs of schools. Table A.10 presents our results when we interact school markets with year fixed effects. Reassuringly, we find estimates of the same magnitude (0.06) across all specifications.

4.3 Direct Impact of the Reform on School Outcomes

Before concluding, we present evidence of the reform's impact on school-level outcomes. Leveraging our difference-in-differences research design, we quantify the impact of the policy on i) the mapping between test scores and course grades, ii) students' performance (test scores, yearly attendance, and course grades), and iii) staff turnover and perceptions.

Assessing Manipulation of Course Grades: Course grades are highly relevant and informative in Chile, and school principals have little to no room to affect subject-specific course grades. Nonetheless, we empirically assess the possibility of grade manipulation. If new principals are changing the culture or gaming the system, we should find a differential change in the school-level link between test scores and grades upon a new principal's arrival. Our examination of the school-level correlation between test scores and grades suggests that this is not the case. Specifically, we estimate more than 30,000 regressions of 4th-grade test scores on course grades by school and year, and then we use the regression coefficients from these regressions as the dependent variable of our main difference-in-differences specification. Reassuringly, as shown by Table 4 (Columns 1 and 2) and Panel A of Appendix Figure A.7, we find that the mapping between course grades and test scores does not change with the appointment of new principals, a result consistent with the fact that teachers, not principals, define course grades in Chile.[33]

Impact on Students' Performance: Leveraging data from the SIMCE exams and the centralized college admissions system (between 2010 and 2017), we can study the impact of the new recruitment system on test scores and college entrance exams. Table 4 presents the estimates obtained from our preferred specification given by Equation (2), and Appendix Figure A.7 presents the event studies given by Equation (3). Column 3 shows the impact on the average SIMCE test scores. We find that after an appointment of a principal under the new system, there is an increase in test scores of 0.07 standard deviations. Columns 4 to 6 show the estimates of the impact of "ADP" appointments on college entrance exams and application scores. For this analysis, we need to restrict our estimating sample to high schools whose students apply to higher education via the centralized college admissions system. It is worth mentioning that 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 being used for post-secondary admissions. Among the students accepted into college in 2017, the correlation between their standardized course

[33]By Law (Article 16.b of Law 19070), the education professionals who perform the teaching function enjoy autonomy to evaluate their students' learning processes.

grades and college admission score was 0.83 (see Appendix Figure A.1, Panel B).[34] Our estimates show that appointing a principal elected under the new selection system increases the average score (of Math and Spanish) by 0.09 standard deviations. Importantly, the final application score, which determines admission for a given institution-major pair, increases by 0.19 standard deviations after appointing a new principal.[35] Finally, in Appendix Table A.11 and Figure A.8, we use data available from 2011 to 2016 to assess the direct impact of the reform on students' course grades and yearly attendance. Consistent with our main findings, estimates show that course grades increased by 0.046 and 0.072 standard deviations in grades 1-8 and 9-12 (high school), respectively. Attendance, on the other hand, increased by 0.1 standard deviation, but only for high school students. Consistent with Miller (2013), we also find a negative association between principal turnover itself (not $\times$ADP) and most of the student outcomes considered here.

Impact on School Personnel: Leveraging records on school staff and teachers' evaluations, we can study the impact of the new selection system on the churn of school personnel. We present estimates from our preferred specifications following Equations (2) and (3) in Table 5, along with Appendix Figure A.9. In this case, the dependent variables are indicators for hiring or firing within the school. Our results reveal that the appointments using the new selection system increased the likelihood of firing and hiring within the principal's support team personnel by 5 and 6 percentage points, respectively. This result is consistent with the fact that public school principals can form their own management teams (deputy director, inspector general, and chief technician). Perhaps more interestingly, Column 5 of Table 5 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.[36] Nonetheless, as shown by Columns 3 and 4, the overall likelihood of hiring and firing personnel from the teaching body did not change due to the reform. We also leverage the data used in Section 3.1 and assess the direct impact of ADP on teachers' perceptions and students' tracking. Appendix Table A.12 presents our results. Estimates suggest that

[34]This strong correlation is due to two facts. First, course grades and contextual course grades have an average weighting of 40 percent in the admission score. Second, since the entrance exams in Chile are oriented to measure how much of the school curriculum has been learned, course grades correlate with the student's performance in the entrance exams.

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

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

the recruitment policy increased the likelihood that teachers agree with the statements: “the principal is good at communicating,” “the principal engages teachers,” and “the principal knows students’ needs.” However, as shown by Column 1, the policy did not affect student sorting (i.e., tracking).

We also estimate our main staggered difference-in-differences using principal characteristics as a dependent variable. We present those estimates in Appendix Table A.13. The results suggest that after the reform, the principals who were hired under the ADP system tended to have more experience in the private sector (a 19% increase over the average of the dependent variable) and were more likely to hold a college degree (a 7% increase over the average of the dependent variable). We also find that they were marginally less likely to have worked as a principal before. This is consistent with anecdotal evidence suggesting that the removal of discretionary appointments at public schools allowed *outsiders* to compete for the position of the school principal. We find similar patterns if we compare the characteristics between public school principals who were selected by the new system and those who were not (see Appendix Table A.14). We likewise find similar results when comparing schools within the same school market (see Appendix Table A.15).[37] To assess the general equilibrium effects of this policy on the labor market, we also estimate a specification where we only keep private schools and divide them into different markets. Then, we consider that a *market* becomes “treated” if there is a public school that changed its principal using the new system for the first time. Based on a similar two-way fixed effects specification, we find that after a school market gets treated, there are few changes in the characteristics of principals in the private school market. As shown by Appendix Table A.15, we only find a small increase in the probability of the principals holding a college degree (1.8% of the average of the dependent variable). These results suggest few market-level effects on the characteristics of principals in the private sector. Finally, Appendix Table A.16 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. Still, in line with the preference for amenities hypothesis, these principals are arriving at municipalities with higher income and more years of schooling (on average).

[37]For this exercise, we define school markets the same way as in the previous section, based on the municipality where the school is located and with the distance metric using a diameter of 7km, i.e., the middle point between the radii of 2 and 5km computed by Cuesta et al. (2020).

5 Discussion

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., 2022). In this article, we measured and studied the effectiveness of managers in education, an essential public service and a landmark of state capacity— one 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 expenditures on education represent more than one-fifth of the budget, and where the efficiency and equity of educational policy are often at the center of the political debate. Leveraging detailed 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 powerfully improve public sector performance.

Specifically, we found that the new selection system generated a 0.06 standard deviation increase in principals' effectiveness. This implies that *ceteris paribus*, the reform created course grade gains of 0.016 standard deviations per year.[38] Evidence of the positive effects of the policy on students' college admission exams, 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 with more selective institution-mayors (Hastings et al., 2013; Zimmerman, 2019), leads us to also expect positive effects on students' adult lives.

Moreover, several factors suggest that the policy studied here may have positive effects in other settings too. First, our results are consistent with the model by Leaver et al. (2019). Second, school principals are important worldwide and have an impact on *all* the students attending their schools. Thus, policies oriented toward recruiting better principals might be an effective way to boost school quality at a relatively low cost. Third, reforms like the one studied here can also be an alternative method to achieve accountability and flexibility in public education (Abdulkadiroğlu et al., 2011), where politicians or unions often have discretion over the appointment of personnel. Whether complementary 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 effects

[38] Compared to other studies in the economics of education literature, a 0.06 effect is not particularly large. It places just above the 25th percentile in terms of learning impact (see Table 2 in Evans and Yuan, 2022).

of more competitive and transparent recruitment is an interesting question for future work.

References

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., 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 (2022): "Political turnover, bureaucratic turnover, and the quality of public services," *American Economic Review*, 112, 442–93.

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.

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

Bartanen, B. and A. N. Husain (2022): "Connected networks in principal value-added models," *Economics of Education Review*, 90, 102292.

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., R. Burgess, A. Khan, and G. Xu (2022): "Bureaucracy and development," *Annual Review of Economics*, 14, 397–424.

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 (2022): "Flexible wages, bargaining, and the gender gap," *The Quarterly Journal of Economics*, 137, 215–266.

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

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

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

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

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., NBER.

Brown, C. and T. Andrabi (2020): "Inducing positive sorting through performance pay: Experimental evidence from Pakistani schools," *UC 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. 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.

Card, D. and A. B. Krueger (1992): "Does school quality matter? Returns to education and the characteristics of public schools in the United States," *Journal of political Economy*, 100, 1–40.

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.

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.

Civil-Service (2015): "Costos de los procesos de selección de directivos sector público vs privado," *Minuta*.

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," *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.

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

Cuesta, J. I., F. González, and C. L. Philippi (2020): "Distorted quality signals in school markets," *Journal of Development Economics*, 147, 102532.

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

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

Dhuey, E. and J. Smith (2018): "How school principals influence student learning," *Empirical Economics*, 54, 851–882.

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.

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.

Evans, D. K. and F. Yuan (2022): "How big are effect sizes in international education studies?" *Educational Evaluation and Policy Analysis*, 44, 532–540.

Fenizia, A. (2022): "Managers and productivity in the public sector," *Econometrica*, 90, 1063–1084.

Finan, F., B. A. Olken, and R. Pande (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.

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., A. J. Egalite, and C. A. Lindsay (2021): "How principals affect students and schools," *Wallace Foundation*.

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," *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., NBER.

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.

Jackson, C. K., S. C. Porter, J. Q. Easton, A. Blanchard, and S. Kiguel (2020): "School effects on socioemotional development, school-based arrests, and educational attainment," *American Economic Review: Insights*, 2, 491–508.

Jackson, C. K., S. C. Porter, J. Q. Easton, and S. Kiguel (2022): "Who Benefits From Attending Effective High Schools?" .

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

Janke, K., C. Propper, and R. Sadun (2019): "The impact of ceos in the public sector:

Evidence from the english nhs," Tech. rep., NBER.
Jochmans, K. and M. Weidner (2019): "Fixed-effect regressions on network data," *Econometrica*, 87, 1543–1560.
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., NBER.
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 (2019): "Effects and mechanisms of CEOs quality in public education," Tech. rep.
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., R. F. Lemos, and D. Scur (2019): "Measuring and explaining management in schools: New approaches using public data," *World Bank Policy Research WP*.
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.
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.
Miller, A. (2013): "Principal turnover and student achievement," *Economics of Education Review*, 36, 60–72.
Moreira, D. and S. Pérez (2021): "Civil service reform and organizational practices: Evidence from the pendleton act," Tech. rep., National Bureau of Economic Research.
Mundlak, Y. (1978): "On the pooling of time series and cross section data," *Econometrica*, 69–85.
Muñoz, P. and C. Otero (2022): "Managers and public hospital performance," *Working Paper*.
Neilson, C. (2019): "Targeted vouchers, competition among schools and the academic achievement of poor students," *Working Paper*.
Ornaghi, A. (2019): "Civil service reforms: Evidence from US police departments," .
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. (2022): "Pretest with caution: Event-study estimates after testing for parallel trends," *American Economic Review: Insights*, 4, 305–22.

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.

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 (2022): "Selecting top bureaucrats: Admission exams and performance in Brazil," *Review of Economics and Statistics*.

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

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*.

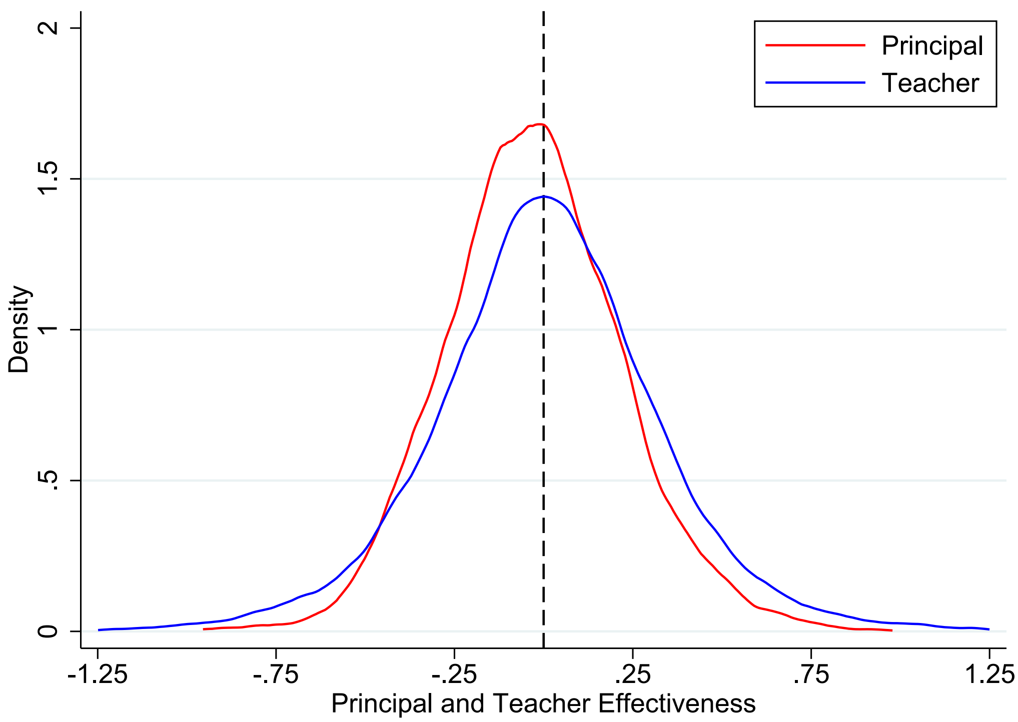
Wooldridge, J. M. (2010): "Correlated random effects models with unbalanced panels," *Michigan State University, Department of Economics*.

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.

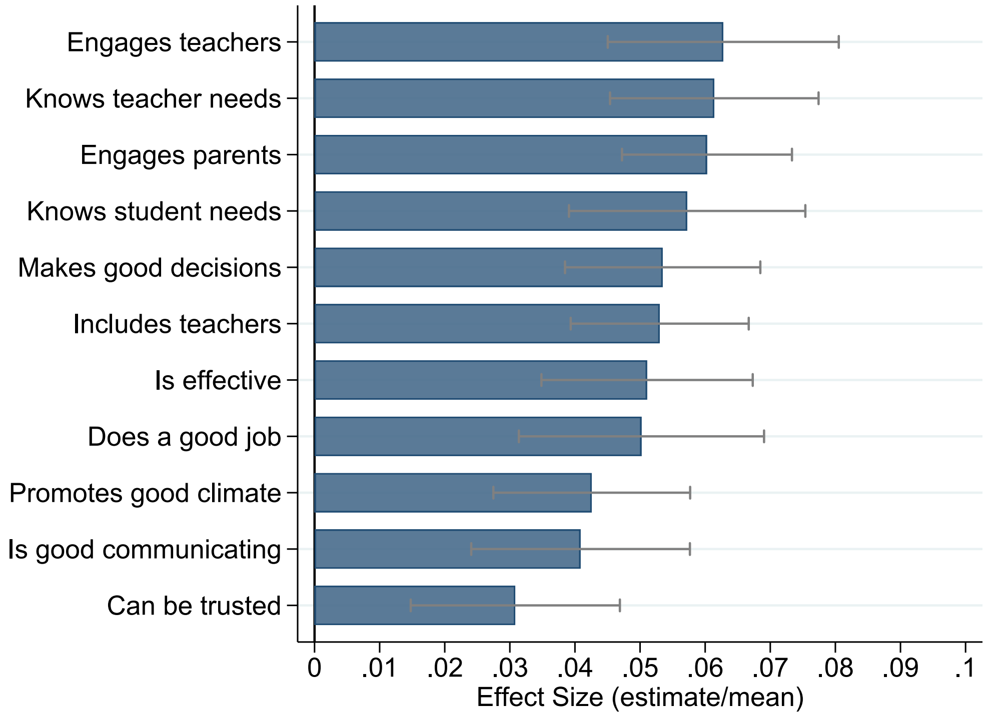
Figures and Tables

Figure 1: Empirical Distributions of Teacher and Principal FE Estimates



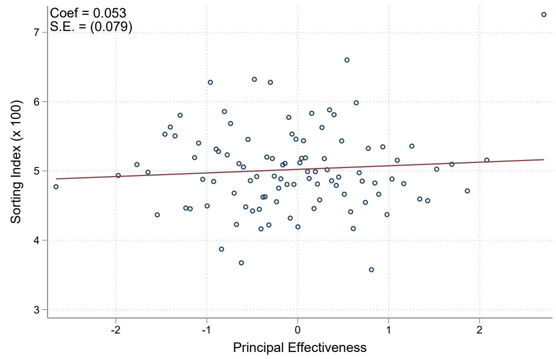
Notes: This figure plots kernel densities of the empirical distribution of teacher and principal fixed effects. Fixed effects are normalized using sum-to-zero constraints, and the densities are weighted by the number of students' course grade observations used to estimate each of these fixed effects.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

Figure 2: Teachers' Survey Responses

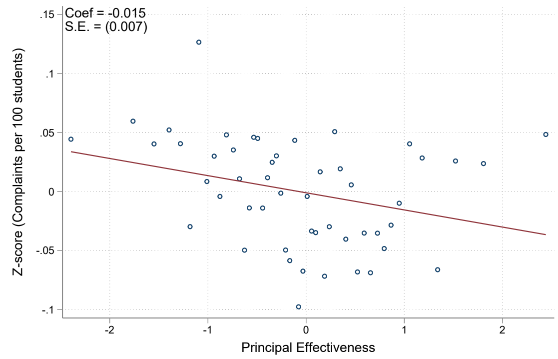


Notes: This figure shows the association between our measure of principal effectiveness and the likelihood that the teaching staff agrees with positive statements about the principal. Using teachers' surveys, we create an indicator if a teacher "highly agrees" with a given statement. In cases when the 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.

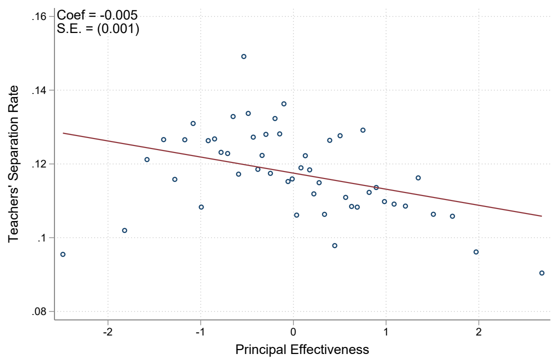
Figure 3: Principals' Effectiveness and the Management of Schools



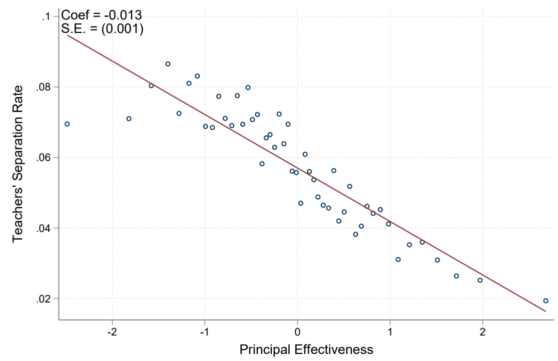
A. Students' Tracking



B. Parents' Complaints



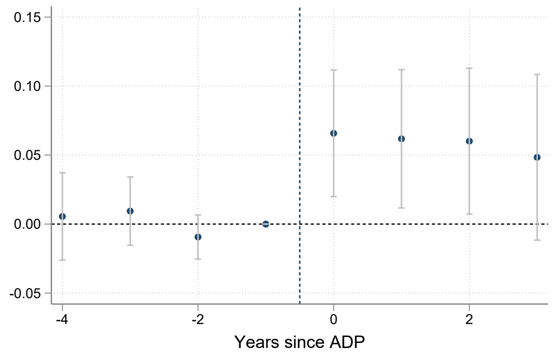
C. Teachers' Turnover (All)



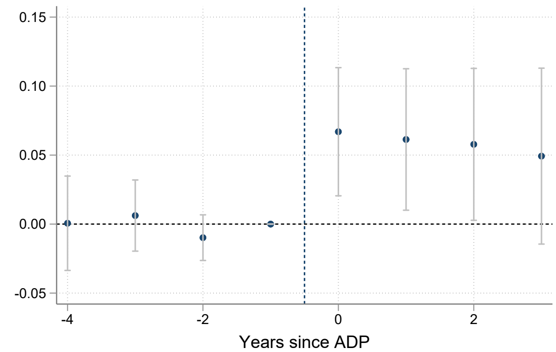
D. Teachers' Turnover (High VA)

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' sorting (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 focuses 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.

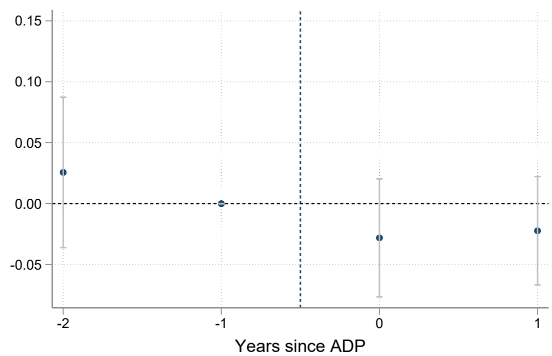
Figure 4: Principal Selection and Principal Effectiveness



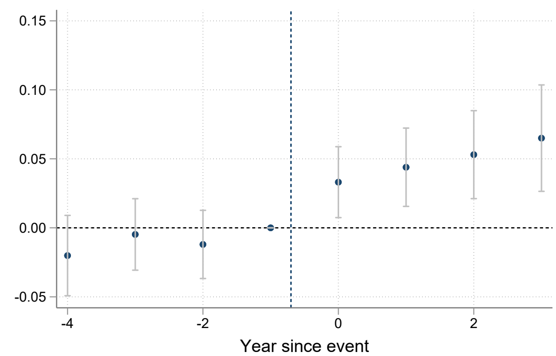
A. No Controls



B. With Controls



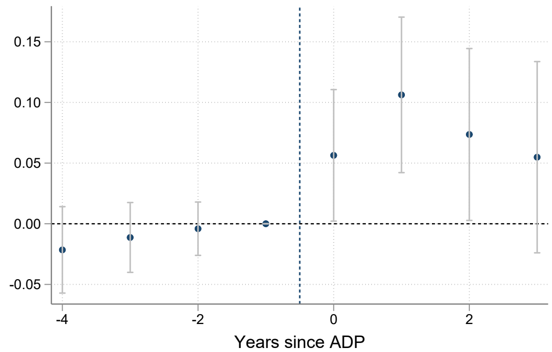
C. Pre-Reform



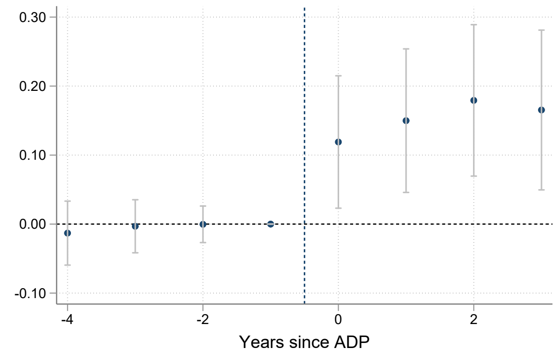
D. Public Schools

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 (3) 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 (3) but only considering turnover during the pre-reform period (2008-2011) and comparing the principal turnover's between public and private schools. Finally, Panel D presents the estimates from a version of equation (3) for the sample of public schools.

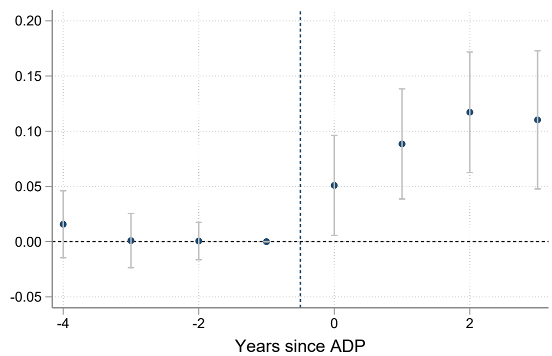
Figure 5: Principal Selection and Principal Effectiveness: Other Measures



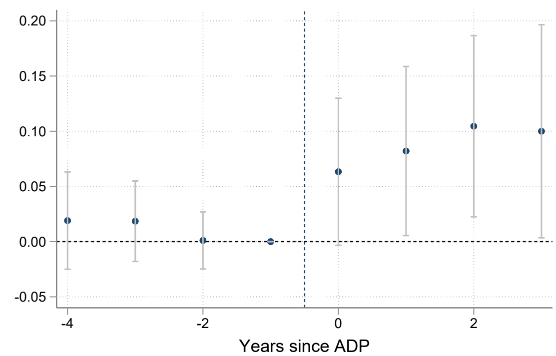
A. Attendance



B. Test score



C. GPA: Contemporaneous



D. GPA: No teacher FE

Notes: The figure shows the impact of appointments under the new selection system on different measures of the effectiveness of public schools' principals. Specifically, we plot the point estimates and 95% confidence intervals estimated from equation (3), including controls by school and municipality characteristics during the pre-reform period (measured in 2010) interacted with year dummies. Panels A and B present the results for a measure of principal fixed effect based on yearly students' attendance and 4th-grade standardized test scores, respectively. Panel C presents the result for our baseline model using contemporaneous GPA (t) instead of future GPA ($t + 1$), while Panel D presents the results obtained from our baseline value-added model but excluding teacher fixed effects.

Table 1: Summary Statistics

	Summary Statistics					By Type of School		
	Obs.	Mean	Std. dev.	10th pctile	90th pctile	Public	Private	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Student characteristics								
Math course grade	1,875,947	5.3	0.8	4.2	6.4	5.16	5.31	-0.15
						(0.80)	(0.85)	(0.00)
Spanish course grade	1,887,264	5.3	0.7	4.3	6.3	5.19	5.38	-0.19
						(0.74)	(0.73)	(0.00)
Math test scores	950,276	263.5	50.6	196.1	328.3	244.83	274.45	-29.61
						(47.21)	(49.26)	(0.00)
Spanish test scores	1,271,641	258.3	50.4	188.4	322.6	243.54	266.34	-22.80
						(49.72)	(48.97)	(0.00)
Ever Grade retention (%)	1,977,203	8.4	27.8	0.0	0.0	10.78	7.06	3.72
						(31.02)	(25.62)	(0.00)
Panel B: School characteristics								
Avg. College Admission Score	15,175	590.7	38.7	546.6	643.9	574.7	595.21	-20.47
						(33.80)	(38.86)	(0.00)
Enrollment	84,746	306	404.6	2	838	215.9	413.15	-197.25
						(303.61)	(477.15)	(0.00)
Annual subsidy per student (USD)	72,814	2,423.5	3,704.1	840.3	4,086.9	2977.05	1627.09	1349.96
						(4302.89)	(2397.34)	(0.00)
Share of disadvantaged students	80,282	46.1	36.0	0.0	92.0	57.30	31.70	25.60
						(32.88)	(34.64)	(0.00)
Teachers per hundred students	76,344	8.1	18.3	2.8	14.3	9.14	6.84	2.30
						(12.44)	(23.65)	(0.00)
Rural school	84,746	43.5	49.6	0.0	100.0	62.40	21.07	41.32
						(48.44)	(40.78)	(0.00)
School attendance	76,538	92.9	4.4	88.4	97.6	92.68	93.28	-0.60
						(4.69)	(3.90)	(0.00)
Panel C: Principal characteristics								
Wage (USD)	6,609	2846.5	2139.9	1609.8	4159.0	2648.59	3030.74	-382.15
						(2582.42)	(1601.50)	(0.00)
% Base salary	6,596	43.0	19.8	23.3	75.4	34.87	50.55	-15.68
						(13.68)	(21.47)	(0.00)
% Bonus	6,596	17.4	19.8	0.8	50.9	10.64	23.63	-12.99
						(15.45)	(21.30)	(0.00)
% Statutory	6,596	40.5	27.9	9.8	69.7	55.75	26.39	29.36
						(27.62)	(19.54)	(0.00)
Permanent contract	7,688	91.1	28.5	100.0	100.0	88.14	93.32	-5.18
						(32.33)	(24.96)	(0.00)
Age	7,688	54.6	10.2	40.0	66.0	54.79	54.40	0.39
						(8.58)	(11.26)	(0.09)
Female	7,688	54.6	49.8	0.0	100.0	47.07	60.41	-13.34
						(49.92)	(48.91)	(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: Principal Selection and Principal Effectiveness

	All Schools			Public Schools				Private Schools
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(a) Principal Turnover × ADP	0.060	0.062	0.062	0.045	0.035	0.040		
	(0.023)	(0.024)	(0.023)	(0.017)	(0.012)	(0.023)		
(b) Principal Turnover	-0.024	-0.027	-0.026				0.003	-0.030
	(0.018)	(0.018)	(0.018)				(0.088)	(0.018)
Observations	30,721	30,721	29,515	14,171	14,171	14,171	5,308	17,502
# of Schools	4,934	4,934	4732	2,389	2,389	2,389	1,668	2,802
R-squared	0.931	0.932	0.931	0.926	-	-	0.958	0.936
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School controls	No	Yes	No	No	No	No	No	No
Municipality controls	No	Yes	No	No	No	No	No	No
p-value a + b = 0	0.017	0.025	0.019					

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 using the new recruitment 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 estimate the regressions described by equation (2). Column 3 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 4 estimates the main regression only within public schools that selected a principal under the ADP system. Column 5 implements the model suggested by De Chaisemartin and d'Haultfoeuille (2020), while column 6 shows the result for the model suggested by Callaway and Sant'Anna (2020). Column 7 shows a placebo exercise where "Principal turnover" is a dummy that takes the value one after a principal turnover in a public school in the period 2010-2011 (pre-ADP reform). The number of schools that had a principal turnover in 2009 or 2010 is 292. Column 8 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 the school level in parenthesis.

Table 3: Principal Selection and Alternative Measures of Principal Effectiveness

	FE constructed based on:			
	Yearly Attendance	Test scores	GPA contemp.	GPA w/o teachers
	(1)	(2)	(3)	(4)
(a) Principal Turnover × ADP	0.084	0.146	0.078	0.067
	(0.029)	(0.048)	(0.023)	(0.034)
(b) Principal Turnover	-0.041	-0.074	-0.047	-0.031
	(0.021)	(0.037)	(0.017)	(0.026)
Observations	30,700	22,658	30,721	30,717
# of Schools	4,930	4,311	4,934	4,933
R-squared	0.883	0.892	0.933	0.819
Year FE	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes
p-value a + b = 0	0.043	0.015	0.059	0.126

Notes: This table presents the effects of the new selection system (ADP) on different measures of principal effectiveness. "ADP" is a dummy that takes the value one after the first time a school selects a principal using the new recruitment system. "Principal Turnover" is a dummy that takes the value one after the first time a school selects a new principal (after 2012). All columns estimate the regressions described by equation (2), adding controls by school and municipality characteristics during the pre-reform period (measured in 2010) interacted with year dummies. Column 1 presents the results for a measure of principal fixed effect based on yearly attendance. Column 2 presents the results for a measure of principal fixed effect based on 4th-grade standardized test scores. Finally, column 3 presents the result for a measure of effectiveness based on our baseline model using contemporaneous GPA (t) instead of future GPA ($t+1$), while column 4 presents the result for a measure of effectiveness based on our baseline model but excluding teacher fixed effects. Robust standard errors clustered at the school level in parenthesis.

Table 4: Principal Selection and Test Scores

	SIMCE Test score			College Entrance Exams score			
	Corr GPA & Test scores		Test scores	Math	Spanish	Average	Application score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(a) Principal Turnover × ADP	-0.005	-0.007	0.066	0.048	0.141	0.095	0.188
	(0.006)	(0.008)	(0.022)	(0.018)	(0.018)	(0.015)	(0.020)
(b) Principal Turnover	0.005	0.007	-0.074	-0.041	-0.061	-0.051	-0.079
	(0.004)	(0.006)	(0.016)	(0.010)	(0.011)	(0.008)	(0.011)
Observations	28,067	28,075	22,380	13,556	13,556	13,556	13,556
# of Schools	5,485	5,485	4,269	2,313	2,313	2,313	2,313
R-squared	0.398	0.311	0.825	0.865	0.772	0.869	0.747
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value a + b = 0	0.915	0.964	0.667	0.664	0.000	0.002	0.000

Notes: This table presents the effects of the new selection system (ADP) on SIMCE and college admission scores. “ADP” is a dummy that takes the value one after the first time a school selects a principal using the new recruitment 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 use as a dependent variable the within school-year level mapping between students’ standardized test scores and course grade in 4th grade, i.e., $\hat{\beta}$ obtained from a regression of test scores on course grades. In column 1, we weight the regression by the inverse of the variance of $\hat{\beta}$, while in column 2, we do not. Column 3 shows the results for the average SIMCE Test Scores at the school level. Columns 4 and 5 show the results for Math and Spanish college admission scores, while column 6 for the average of both. Finally, column 7 shows the estimates for the composite score actually 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 control by municipality characteristics during the pre-reform period (measured in 2010), interacted with year dummies. Robust standard errors clustered at the school level in parenthesis.

Table 5: Principal Selection and School Staff

	Principal's Support Team		Teaching Staff			
			All		Low Rating	
	Firing	Hiring	Firing	Hiring	Firing	Hiring
	(1)	(2)	(3)	(4)	(5)	(6)
(a) Principal Turnover × ADP	0.047	0.056	-0.002	0.006	0.116	-0.023
	(0.012)	(0.012)	(0.009)	(0.008)	(0.014)	(0.014)
(b) Principal Turnover	0.010	0.038	0.005	0.008	-0.022	0.010
	(0.009)	(0.008)	(0.007)	(0.006)	(0.009)	(0.008)
Observations	30,721	30,721	30,721	30,420	30,721	30,420
# of Schools	4,934	4,934	4,934	4,908	4,934	4,908
R-squared	0.279	0.272	0.346	0.333	0.273	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
Mean Dep Var	0.118	0.0939	0.911	0.931	0.177	0.192
p-value a + b = 0	0.000	0.000	0.751	0.022	0.000	0.293

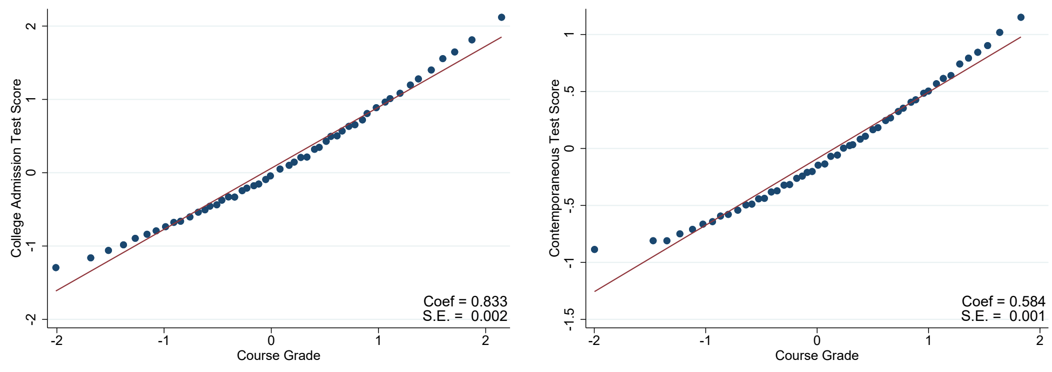
Notes: This table presents the effects of the new selection system (ADP) on the churn of the school staff. "ADP" is a dummy that takes the value one after the first time a school selects a principal using the new recruitment system. "Principal Turnover" is a dummy that takes the value one after the first 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 control by school and municipality characteristics during the pre-reform period (measured in 2010), interacted with year dummies. Robust standard errors clustered at the school level in parenthesis.

ONLINE APPENDIX

Managers' Productivity and Recruitment in the Public Sector
By Pablo Muñoz and Mounu Prem

A Additional Figures and Tables

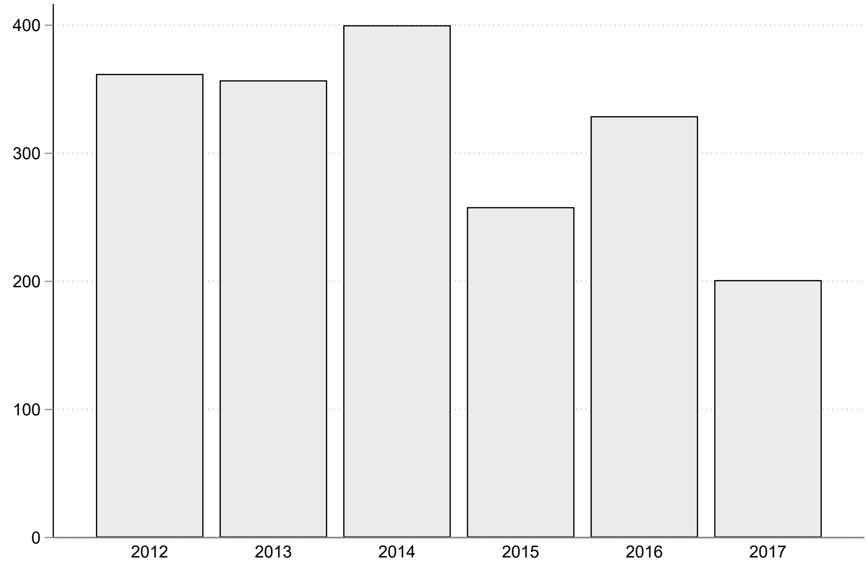
Figure A.1: Course Grades and Test Scores



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

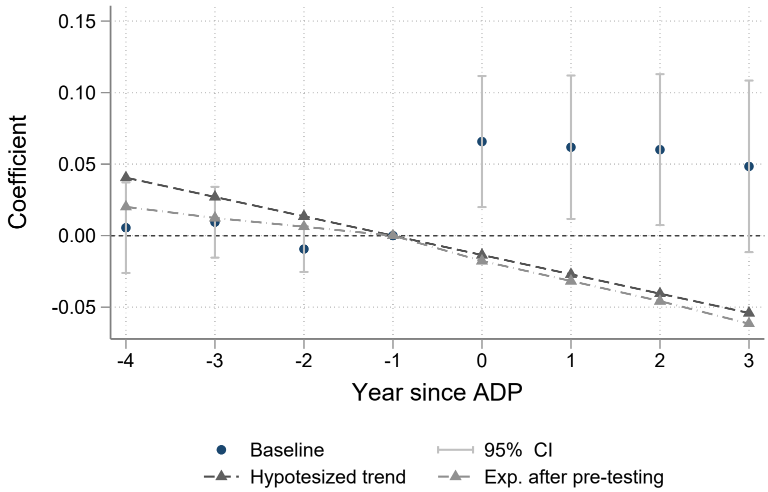
Notes: Panel A considers a sample of 132,585 students accepted into college and for whom we can compute college admission scores from the *Prueba de Selection Universitaria* (PSU), in the 2017 process. The college admission score is an institution-major-specific weighted average of applicants' high-school 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 scores on course grades.

Figure A.2: Number of Newly Elected Principals by ADP, per Year

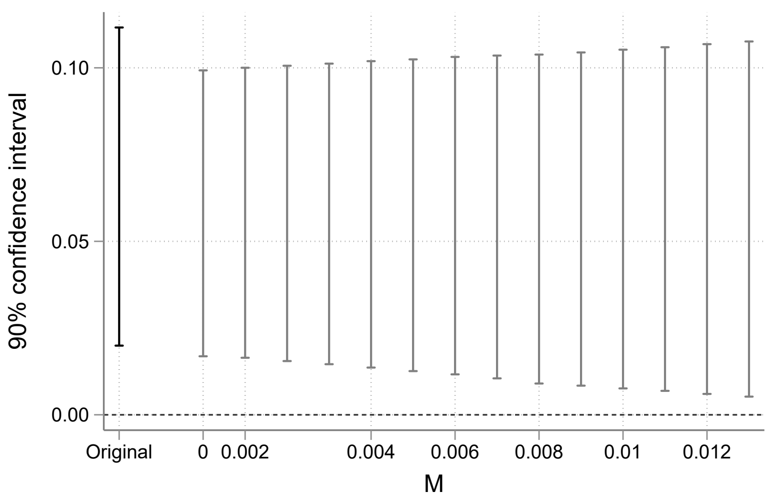


Notes: This figure shows the number of schools that elected a principal through the new ADP selection system for the first time, by year.

Figure A.3: Principal Selection and Principal Effectiveness: Parallel trends violations



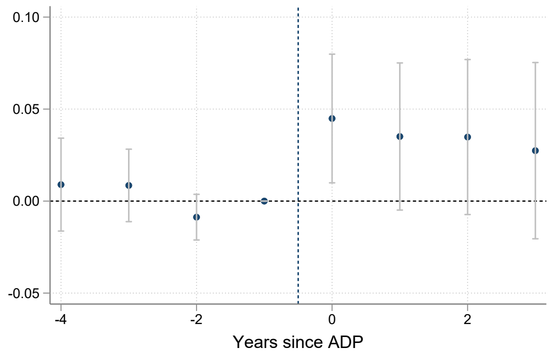
A. Pre-trend



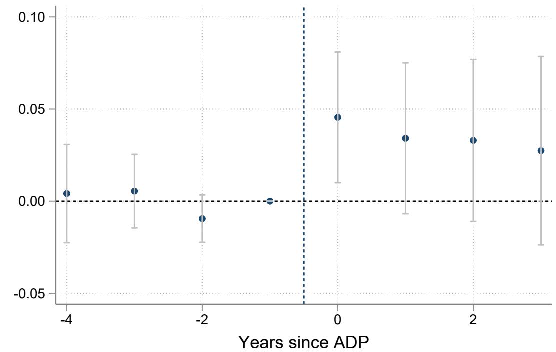
B. Violations from parallel trends

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, 2022). 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, 2022), which is 0.013. In Figure A.3 Panel B, we present the results.

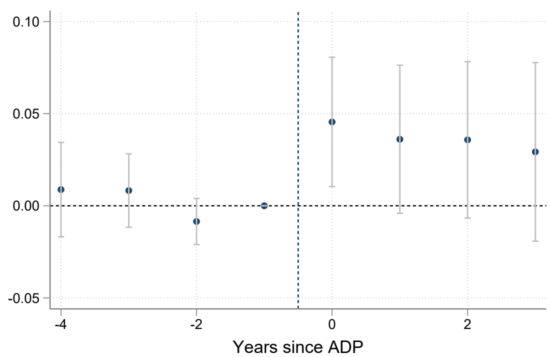
Figure A.4: Principal Selection and Principal Effectiveness: Weighted least squares



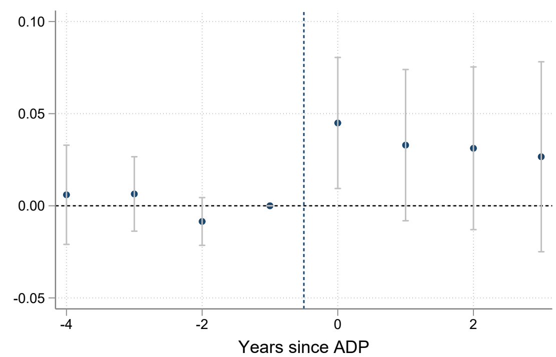
A. Baseline



B. Adds municipal controls



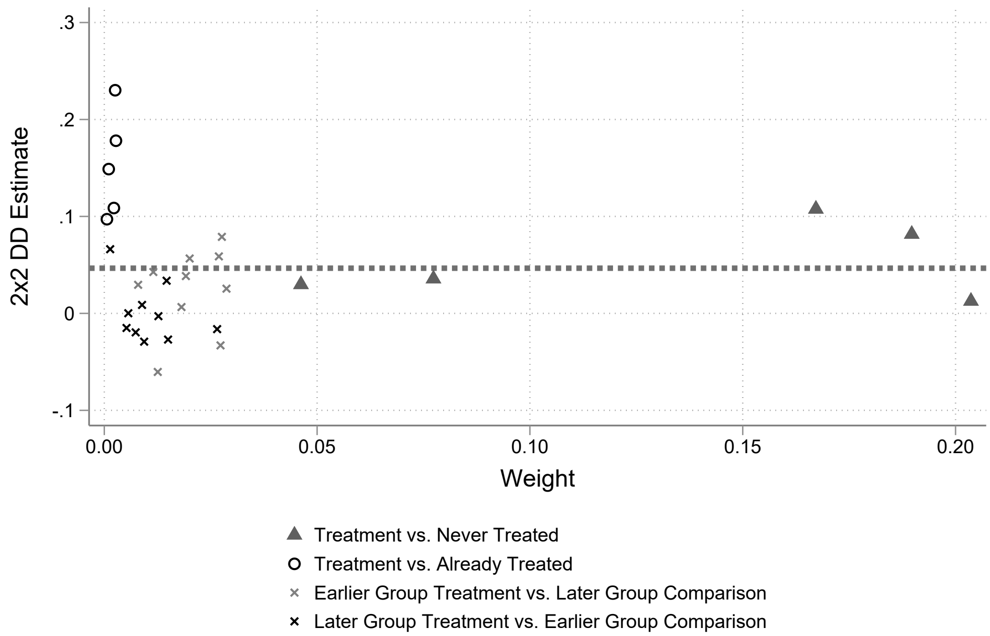
C. Adds school controls



D. Adds both set of controls

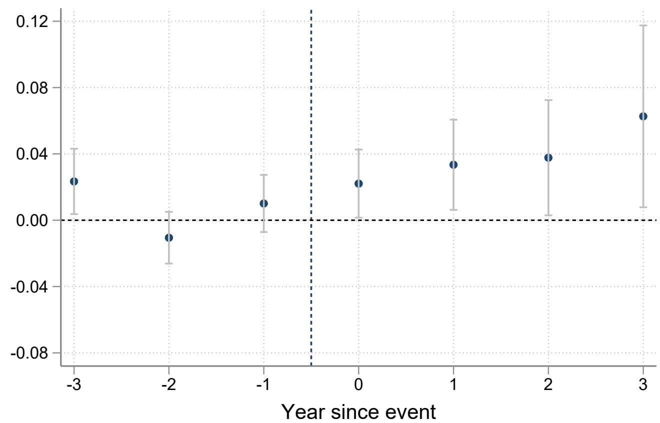
Notes: This figure presents the dynamic version of our *staggered difference-in-differences*. We use as dependent variable the un-adjusted principal fixed effects and we the regressions by the inverse of the standard deviation of the estimate. Panel A presents the dynamic version of the staggered difference-in-differences model suggested by De Chaisemartin and d'Haultfoeuille (2020). Panel B 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 panel B, 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 A.

Figure A.5: Goodman-Bacon (2021) Decomposition

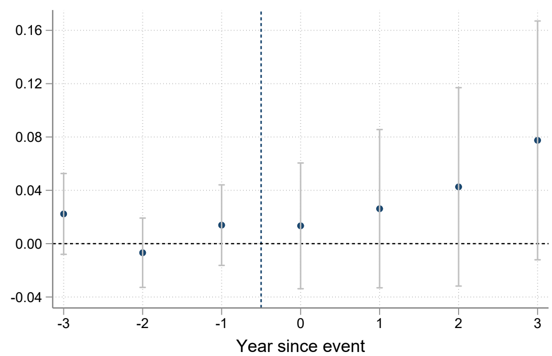


Notes: This figure presents the decomposition of the two-way fixed effect estimator suggested by Goodman-Bacon (2021).

Figure A.6: Principal Selection and Principal Effectiveness within Public Schools



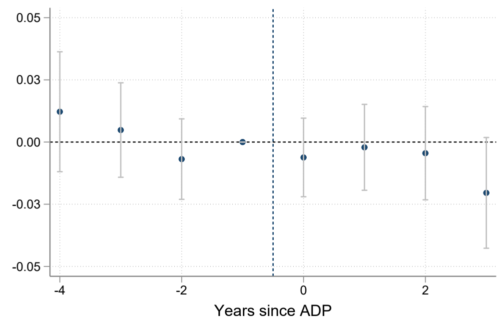
A. De Chaisemartin and d'Haultfoeuille (2020)



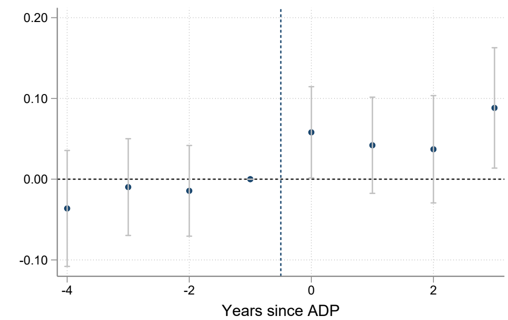
B. Callaway and Sant'Anna (2020)

Notes: This figure presents the dynamic version of our *staggered difference-in-differences* approach in the sample of public schools. Panel A presents the dynamic version of the staggered difference-in-differences model suggested by De Chaisemartin and d'Haultfoeuille (2020). Panel B 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 panel B, 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 A.

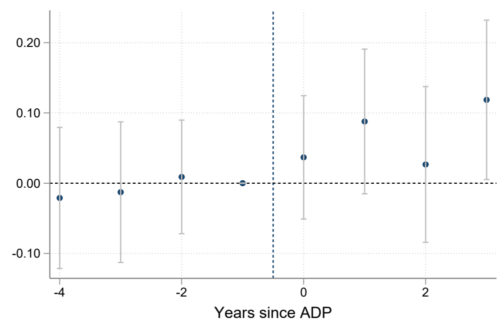
Figure A.7: Principal Selection and Test Scores



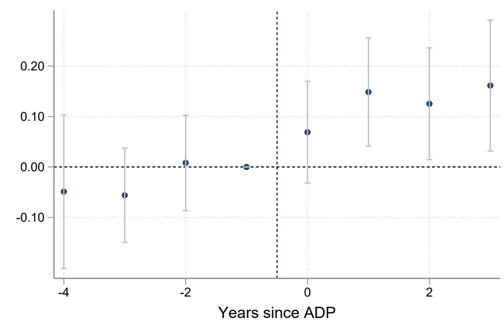
A. Corr GPA-Test scores (SIMCE)



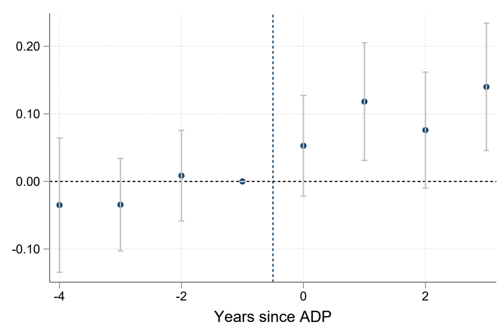
B. Test scores (SIMCE)



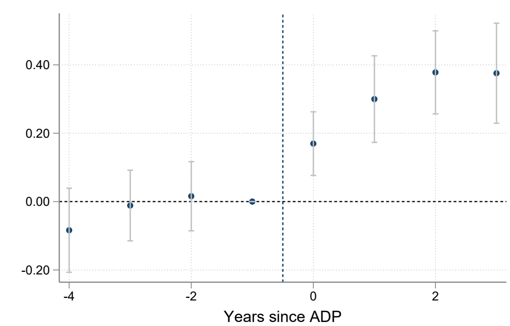
C. Math (PSU)



D. Spanish (PSU)



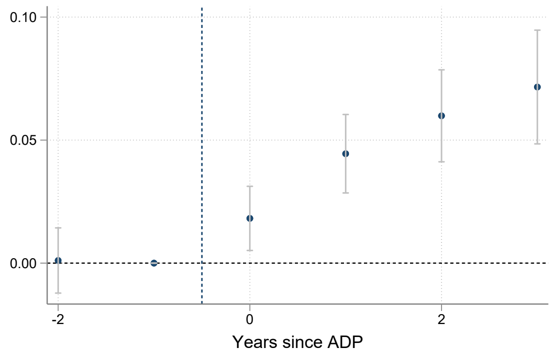
E. Average Math and Spanish (PSU)



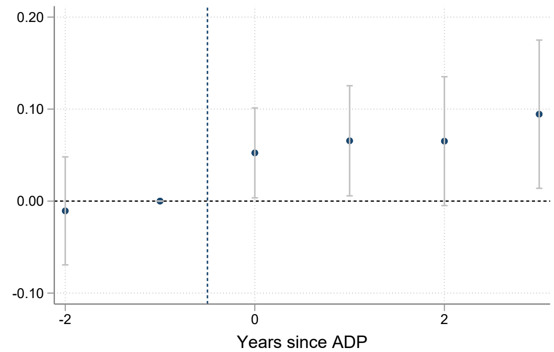
F. Application score (PSU)

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 (3). Panel A uses as a dependent variable the within school-year level correlation between standardized test scores in 4th grade and GPA at the student level. Panel B shows the results for the average SIMCE standardized test scores. Panels C and D show the impact on the mandatory exams of Math and Spanish, while Panel E shows the impact of the average between Math and Spanish. Panel F 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.

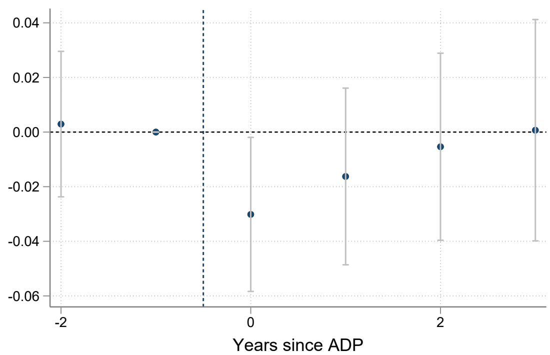
Figure A.8: Principal Selection, GPA, and School Attendance



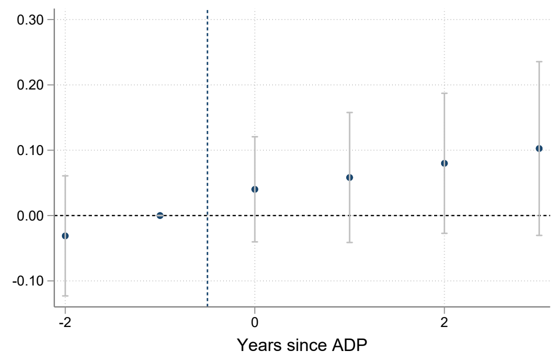
A. GPA: 1 to 8th grade



B. GPA: 9 to 12th grade



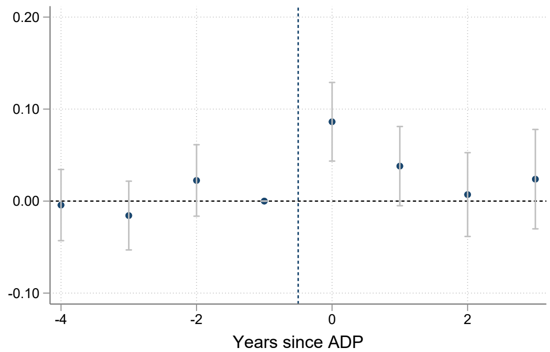
C. Attendance: 1 to 8th grade



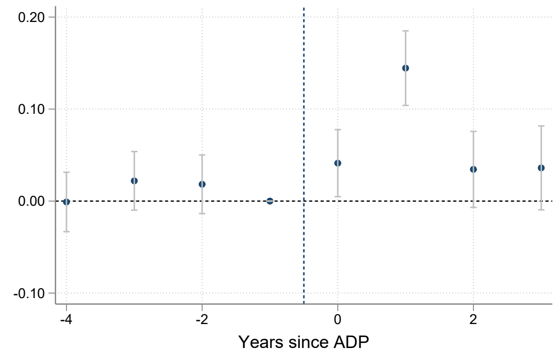
D. Attendance: 9 to 12th grade

Notes: This figure shows the impact of appointments under the new selection system on students' course grades and yearly attendance. The figure plots the point estimates and 95% confidence intervals estimated from equation (3). Panel A presents the results for standardized average GPA for grades 1 to 8, while Panel B shows the results for grades 9 to 12. Panels C and D present the results using students' yearly attendance as the dependent variable, for 1 to 8th and for 9th to 10th grade, respectively. All panels include school and year fixed effects.

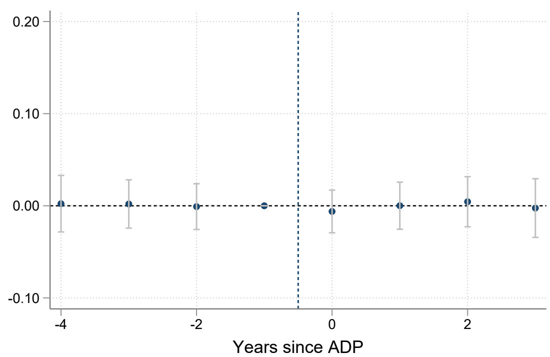
Figure A.9: Principal Selection and School Staff



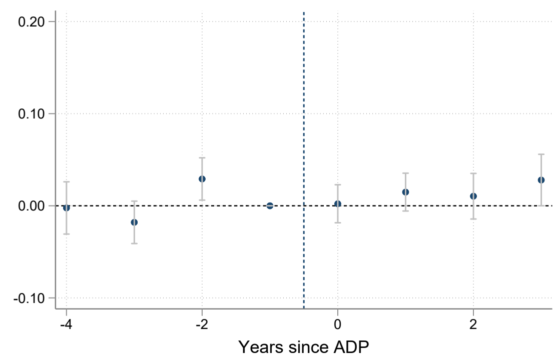
A. Firing: Principal's Team



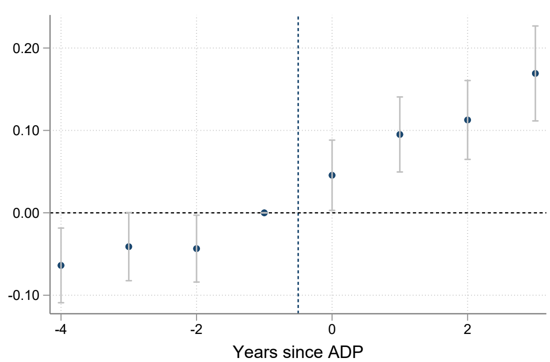
B. Hiring: Principal's Team



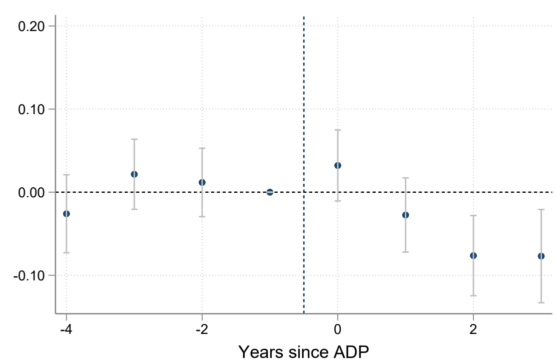
C. Firing: All Teachers



D. Hiring: All Teachers



E. Firing: Teachers with Bad Evaluation



F. Hiring: Teachers with Bad Evaluation

Notes: This figure plots the point estimates and 95% confidence intervals estimated from equation (3). The dependent variable is a dummy that takes the value one if there were any teachers fired (Panels A, C, and E) or any teachers hired (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), while 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.

Table A.1: Descriptive Statistics in Different Samples

	Full Sample		**Δ Teacher=1**		**LCS=1**	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
	(1)	(2)	(3)	(4)	(5)	(6)
Primary (2-8)	0.8	0.4	0.8	0.4	0.9	0.3
Secondary (9-11)	0.2	0.4	0.2	0.4	0.1	0.3
Subject = Math	0.5	0.5	0.5	0.5	0.5	0.5
Course Grade	5.7	0.6	5.7	0.6	5.7	0.6
% Attendance	91.7	3.3	91.7	3.3	91.8	3.1
% Rural School	0.1	0.3	0.1	0.3	0.1	0.3
% Public School	0.4	0.5	0.4	0.5	0.3	0.5
School Size	794.3	596.6	825.9	618.7	847.4	624.3
Sample Size	12,709,601		9,120,261		7,735,653	

Notes: This table presents descriptive statistics of students in three different samples. "Full Sample" includes all students in our dataset after excluding preschools, adults' schools, and special education schools. We also exclude classes that had more than one teacher per year and eliminate the bottom and top one percent of classroom size outliers. "ΔTeacher = 1" corresponds to the restricted sample of students for whom the teacher, in a given subject, changed between t and $t + 1$. Finally, "LCS" includes all students within the largest connected set 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	0.029	-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	0.050	0.049	0.065	0.042
	(0.009)	(0.010)	(0.015)	(0.013)
Perm. Contract	0.019	0.023	-0.000	0.042
	(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	-0.033	0.021	-0.097
	(0.015)	(0.016)	(0.023)	(0.022)
Ever Teacher		0.019	-0.005	0.027
		(0.012)	(0.017)	(0.016)
Ever Admin. Supp. Worker		-0.014	-0.040	-0.016
		(0.012)	(0.017)	(0.018)
Ever Admin Worker		-0.034	0.038	-0.027
		(0.033)	(0.055)	(0.040)
Observations	42,022	35,343	15,833	19,503

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 are presented in parentheses.

Table A.3: Teachers' Survey Responses

% Teachers highly agreeing that the principal:	$\hat{\beta}$ (1)	Standard 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	5,351	0.000	0.001
Can be trusted	0.016	(0.004)	0.521	5,349	0.000	0.001
Makes good decisions	0.024	(0.003)	0.459	6,386	0.000	0.001
Is effective	0.023	(0.004)	0.448	6,382	0.000	0.001
Is good at communicating	0.022	(0.005)	0.529	5,355	0.000	0.001
Engages teachers	0.028	(0.004)	0.444	6,367	0.000	0.001
Engages parents	0.028	(0.003)	0.464	6,386	0.000	0.001
Knows teacher needs	0.027	(0.004)	0.439	6,389	0.000	0.001
Knows student needs	0.029	(0.005)	0.502	5,351	0.000	0.001
Includes teachers	0.025	(0.004)	0.469	7,230	0.000	0.001
Promotes good work climate	0.022	(0.004)	0.525	5,273	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 "highly agrees" 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 A.4: Principals' Effectiveness and the Management of Schools

		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	-0.068	-0.024	-0.001	-0.002	-0.002	-0.019	-0.007	-0.002	-0.012	0.011
	(0.069)	(0.007)	(0.002)	(0.001)	(0.001)	(0.006)	(0.003)	(0.001)	(0.001)	(0.001)
	[0.415]	[0.009]	[0.584]	[0.415]	[0.415]	[0.009]	[0.039]	[0.297]	[0.009]	[0.009]
Observations	13,803	10,225	10,225	10,225	10,225	10,225	10,225	42,279	42,279	42,279
Mean Dep Var	5.019	-6.89e-09	0.0442	0.0272	0.0214	0.340	0.114	0.118	0.0571	0.0605
R-squared	0.094	0.099	0.098	0.060	0.067	0.102	0.071	0.106	0.079	0.075
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. The 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. In square brackets, we present p-values that control for the false discovery rate in related groups of outcomes following Romano and Wolf (2005).

Table A.5: Principals' Effectiveness and School Finance

		Expenditures					Income				
	Expenditures to Income Ratio	Log Total Expenditure	Personnel (%)	Learning (%)	Operations (%)	Other (%)	Log Total Income	Subsidies (%)	Self-Revenues (%)	Initial Budget (%)	Other (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Principal Effectiveness	-0.202	-0.012	-0.007	-0.030	-0.046	-0.027	0.003	-0.004	-0.066	-0.024	0.021
	(0.195)	(0.016)	(0.014)	(0.024)	(0.020)	(0.026)	(0.014)	(0.014)	(0.033)	(0.036)	(0.015)
Observations	4,122	4,114	4,080	4,006	4,052	4,054	4,115	4,127	2,839	3,342	4,121
R-squared	0.050	0.218	0.214	0.208	0.177	0.160	0.202	0.197	0.239	0.162	0.165
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep Var	0.60	19.97	19.73	16.68	17.23	17.18	20.13	19.83	17.80	17.61	17.14

Notes: This table shows the results from a set of regressions of different measures of school finance on the standardized measure of principal effectiveness discussed in section 3. The data on school finance is only for the year 2016. All the regressions include municipality fixed effects. Robust standard errors are presented in parentheses.

Table A.6: Principal Selection and Principal Effectiveness: Weighted least squares

	(1)	(2)	(3)	(4)
(a) Principal Turnover × ADP	0.046	0.049	0.048	0.047
	(0.024)	(0.025)	(0.024)	(0.025)
(b) Principal Turnover	-0.021	-0.023	-0.022	-0.023
	(0.018)	(0.019)	(0.018)	(0.019)
Observations	30,721	30,721	30,721	30,721
R-squared	0.925	0.925	0.925	0.925
Year FE	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes
School controls	No	Yes	No	Yes
Municipality controls	No	No	Yes	Yes
# of Schools	4,934	4,934	4,934	4,934
pvalue a + b = 0	0.112	0.098	0.111	0.130

Notes: This table presents the effects of the new selection system (ADP) on the un-adjusted standardized measure of principal effectiveness discussed in section 3. All regressions are weighted by the inverse of the standard deviation of each estimated fixed effect. "ADP" is a dummy that takes the value one after the first time a school selects a principal using the new recruitment 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 (2). Robust standard errors clustered at the school level in parenthesis.

Table A.7: Principal Selection and Principal Effectiveness by Rurality

	Rural	Urban	Rural	Urban
	(1)	(2)	(3)	(4)
Panel A: Based on school status				
(a) Principal Turnover × ADP	0.033	0.049	0.106	0.023
	(0.071)	(0.024)	(0.035)	(0.018)
(b) Principal Turnover	0.069	-0.033		
	(0.065)	(0.019)		
Observations	7,057	23,631	5,763	8,387
# of Schools	1,201	3,825	997	1,450
R-squared	0.933	0.931	0.926	0.927
Mean Dep Var	0.190	-0.051	0.266	0.015
p-value a + b = 0	0.003	0.316		
p-value rural-urban	0.798	0.798	0.057	0.057
Panel B: Based on municipality characteristic				
(a) Principal Turnover × ADP	0.123	0.049	0.058	0.039
	(0.064)	(0.025)	(0.036)	(0.019)
(b) Principal Turnover	-0.075	-0.017		
	(0.056)	(0.019)		
Observations	6,300	24,305	4,395	9,745
# of Schools	1,029	3,905	749	1,639
R-squared	0.934	0.929	0.927	0.924
Mean Dep Var	0.270	-0.0628	0.299	0.0364
p-value a + b = 0	0.179	0.049		
p-value rural-urban	0.331	0.331	0.642	0.642
Year FE	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes

Notes: This table presents the effects of the new selection system (ADP) on principal effectiveness by rurality of the school and municipality. "ADP" is a dummy that takes the value one after the first time a school selects a principal using the new recruitment system. "Principal Turnover" is a dummy that takes the value one after the first time a school selects a new principal (after 2012). All columns estimate the regressions described by equation (2). Columns 1 and 2 present the results for the comparison between public and private schools, while columns 3 and 4 show the results only within the public sector. Columns 1 and 3 (2 and 4) present the results for rural (urban) areas. Panel A uses the definition of rural/urban school from the ministry of education, while Panel B separates municipalities into rural/urban using the median of the empirical distribution of the share of rural population in municipality based on the 2017 Census. Robust standard errors clustered at the school level in parenthesis.

Table A.8: 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	466.429
	(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	-0.564
	(21.439)	(45.673)	(0.000)	(51.932)	(41.269)	(0.136)	(44.728)
Rural School	0.838	0.256	-0.582	0.224	0.276	0.051	.235
	(0.369)	(0.437)	(0.000)	(0.418)	(0.447)	(0.015)	(0.423)
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	-.064
	(15.703)	(1.173)	(0.080)	(1.739)	(0.590)	(0.209)	(6.139)
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)	(0.283)
Δ Share of disadvantaged students	-0.037	-0.014	0.023	-0.010	-0.016	-0.006	-0.004
	(0.175)	(0.114)	(0.000)	(0.121)	(0.108)	(0.249)	(0.110)
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)	(26.146)
Δ 4rd grade test scores (Spanish)	-1.590	-2.637	-1.047	-1.792	-3.140	-1.348	-4.876
	(33.854)	(22.127)	(0.308)	(21.959)	(22.223)	(0.251)	(21.871)
4rd grade test scores (Math)	238.975	245.459	6.484	245.793	245.262	-0.531	256.763
	(33.034)	(24.510)	(0.000)	(24.177)	(24.715)	(0.682)	(256.76)
Δ 4rd grade test scores (Math)	9.237	8.194	-1.043	8.549	7.982	-0.567	4.134
	(34.392)	(23.280)	(0.324)	(22.527)	(23.727)	(0.646)	(22.926)
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)	(76.776)
Δ Graduation test score (Spanish)	-4.119	-4.079	0.040	-3.011	-4.896	-1.885	-2.173
	(29.886)	(25.310)	(0.989)	(21.255)	(28.042)	(0.475)	(25.191)
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)	(80.753)
ΔGraduation test score (Math)	-7.798	-3.738	4.061	-3.929	-3.591	0.337	-3.281
	(31.307)	(25.322)	(0.155)	(21.378)	(28.017)	(0.898)	(24.495)
Panel B: Municipality characteristics							
Share of households in poverty	0.124	0.082	-0.042	0.082	0.082	-0.000	0.0933
	(0.075)	(0.056)	(0.000)	(0.057)	(0.055)	(0.952)	(0.076)
Income per capita	1.699	2.151	0.453	2.223	2.107	-0.115	2.386
	(0.489)	(1.115)	(0.000)	(1.400)	(0.892)	(0.033)	(1.652)
Unemployment rate	0.080	0.080	0.001	0.083	0.079	-0.004	0.083
	(0.047)	(0.047)	(0.626)	(0.050)	(0.045)	(0.064)	(0.042)
Average years of schooling	8.974	9.998	1.024	9.930	10.041	0.110	10.286
	(1.124)	(1.315)	(0.000)	(1.385)	(1.269)	(0.083)	(1.567)
Observations	3,029	1,820	4,849	695	1,125	1,820	3,782

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.9: Principal Selection and Principal Effectiveness: Placebos

	Public Schools (pre-reform)		Private Schools (post-reform)	
	(1)	(2)	(3)	(4)
Principal Turnover	0.006	0.021	-0.032	-0.026
	(.095)	(0.110)	(.019)	(0.036)
Observations	5,308	5,308	17,502	17,502
# of Schools	1,668	1,668	2,802	2,802
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 that 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 the school level in parentheses.

Table A.10: Principal Selection and Principal Effectiveness: Controlling for school markets

		Based on a distance of:			
Market:	Municipality	4kms	5kms	6kms	7kms
	(1)	(2)	(3)	(4)	(5)
(a) Principal Turnover × ADP	0.058	0.047	0.050	0.057	0.057
	(0.023)	(0.025)	(0.024)	(0.024)	(0.024)
(b) Principal Turnover	-0.027	-0.025	-0.023	-0.027	-0.024
	(0.018)	(0.019)	(0.019)	(0.018)	(0.018)
Observations	30,440	26,914	27,944	28,886	29,564
# of Schools	4,908	4,347	4,505	4,654	4,756
R-squared	0.937	0.937	0.936	0.935	0.933
Year FE	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes
School Market-Year	Yes	Yes	Yes	Yes	Yes
# of markets	295	276	220	178	127
p-value a + b = 0	0.053	0.184	0.096	0.058	0.034

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). All columns estimate the regressions described by equation (2), but adds a school market fixed effect interacted with year fixed effects. In column 1, we define a school market based on the municipality in which the school is located. In columns 2 to 4, we follow Cuesta et al. (2020) and Neilson (2019) and define a market based on the distance between schools. In Chile, the students' average distance to chosen schools is 2kms and the 90th percentile of such distribution is 5kms (Cuesta et al., 2020). Based on this, we define a market based on the schools that are close to each other. We use four different diameters (k) to define a market 4, 5, 6, and 7kms. We implement this by creating a symmetric adjacency matrix (A), where the element $A(i, j)$ takes the value one if i and j are less than k kms away from each other. Then, we construct the set of "connected components" of the matrix, where a component is defined as a set of schools where one can always find a "path" that connects two pairs of schools. Robust standard errors clustered at the school level in parenthesis.

Table A.11: Principal Selection, GPA, and School Attendance

	GPA		Attendance	
Grades:	1 to 8	9 to 12	1 to 8	9 to 12
	(1)	(2)	(3)	(4)
(a) Principal Turnover × ADP	0.046	0.072	-0.016	0.108
	(0.007)	(0.025)	(0.011)	(0.046)
(b) Principal Turnover	-0.018	-0.015	-0.001	-0.039
	(0.005)	(0.010)	(0.008)	(0.011)
Observations	25,178	9,459	25,178	9,458
R-squared	0.832	0.801	0.786	0.782
Year FE	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes
Mean dep var	0.079	-0.178	0.073	-0.026
p-value a + b = 0	0.000	0.017	0.054	0.117

Notes: This table presents the effects of the new selection system (ADP) on average GPA and average attendance rate. "ADP" is a dummy that takes the value one after the first time a school selects a principal using the new recruitment system. "Principal Turnover" is a dummy that takes the value one after the first time a school selects a new principal (after 2012). All columns estimate the regressions described by equation (2). In columns 1 and 2 we consider standardized course grades as the dependent variable and in columns 3 and 4 we consider standardized yearly attendance as the dependent variable. Robust standard errors clustered at the school level in parenthesis.

Table A.12: Principal Selection, Students' Sorting, and Teachers' Surveys

	Sorting Index	Does a good job	Can be trusted	Makes good decisions	Is effective	Is good at communicating	Is a good manager	Engages teachers	Engages parents	Knows teacher needs	Knows student needs	Includes teachers	Promotes good work climate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(a) Principal Turnover × ADP	-0.003	0.052	0.071	0.020	0.017	0.086	-0.004	0.048	0.011	0.009	0.088	-0.016	0.027
	(0.002)	(0.043)	(0.044)	(0.022)	(0.022)	(0.043)	(0.022)	(0.021)	(0.022)	(0.022)	(0.043)	(0.017)	(0.044)
(b)Principal Turnover	0.001	-0.045	-0.036	0.019	0.036	-0.061	0.021	-0.010	0.006	0.032	-0.052	0.032	-0.024
	(0.001)	(0.031)	(0.031)	(0.015)	(0.015)	(0.032)	(0.015)	(0.015)	(0.014)	(0.015)	(0.031)	(0.012)	(0.032)
Observations	12,547	11,658	11,604	16,808	15,969	11,674	16,522	16,364	16,177	16,542	11,551	21,142	10,520
R-squared	0.560	0.511	0.485	0.456	0.479	0.496	0.476	0.460	0.466	0.468	0.516	0.425	0.513
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep var	0.050	0.472	0.533	0.467	0.454	0.540	0.488	0.452	0.471	0.444	0.511	0.470	0.532
p-value a + b = 0	0.309	0.831	0.289	0.030	0.003	0.426	0.332	0.030	0.352	0.019	0.246	0.244	0.922

Notes: This table presents the effects of the new selection system (ADP) on answers from the teachers' surveys. "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). All columns estimate the regressions described by equation (2). The dependent variable in column 1 corresponds to the sorting index discussed in section 3. To construct the dependent variables in columns 2-13, 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. Robust standard errors clustered at the school level in parenthesis.

Table A.13: Principal Selection and Principals' Characteristics

	Ever teacher	Ever administrative	Ever principal	Ever in private sector	Age	Female	College degree
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Baseline							
(a) Principal Turnover × ADP	-0.009	-0.012	-0.029	0.022	-0.254	0.020	0.059
	(0.025)	(0.007)	(0.015)	(0.010)	(0.458)	(0.023)	(0.016)
(b) Principal Turnover	0.121	-0.007	0.011	0.001	-4.278	-0.003	0.023
	(0.019)	(0.007)	(0.014)	(0.009)	(0.384)	(0.017)	(0.011)
R-squared	0.778	0.476	0.550	0.919	0.728	0.752	0.742
p-value a + b = 0	0.000	0.000	0.0158	0.000	0.000	0.261	0.000
Panel B: Adds county-year fixed effects							
(a) Principal Turnover × ADP	-0.006	-0.010	-0.025	0.027	-0.098	0.011	0.065
	(0.025)	(0.007)	(0.015)	(0.011)	(0.472)	(0.024)	(0.016)
(b) Principal Turnover	0.124	-0.010	0.005	0.002	-4.643	0.005	0.024
	(0.019)	(0.007)	(0.014)	(0.009)	(0.392)	(0.017)	(0.011)
R-squared	0.800	0.520	0.589	0.925	0.750	0.772	0.763
p-value a + b = 0	0.000	0.000	0.0199	0.000	0.000	0.331	0.000
Panel C: Adds school market-year fixed effects							
(a) Principal Turnover × ADP	-0.009	-0.014	-0.033	0.025	-0.272	0.018	0.059
	(0.025)	(0.007)	(0.015)	(0.011)	(0.467)	(0.023)	(0.016)
(b) Principal Turnover	0.118	-0.008	0.011	0.000	-4.348	0.002	0.022
	(0.019)	(0.007)	(0.014)	(0.009)	(0.388)	(0.017)	(0.011)
R-squared	0.782	0.490	0.563	0.921	0.736	0.762	0.753
p-value a + b = 0	0.000	0.000	0.005	0.000	0.000	0.206	0.000
Observations	24,734	29,564	24,734	24,734	29,564	29,564	29,564
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep var	0.356	0.978	0.932	0.117	54.36	0.547	0.878

Notes: This table presents the effects of the new selection system (ADP) on principals' characteristics. "ADP" is a dummy that takes the value one after the first time a school selects a principal using the new recruitment system. "Principal Turnover" is a dummy that takes the value one after the first time a school selects a new principal (after 2012). All columns estimate the regressions described by equation (2). In Panel B, we define a school market based on the municipality in which the school is located and add an interaction between county and year fixed effects. In Panel C, we consider the market definition using the 7km distance. Robust standard errors clustered at the school level in parenthesis.

Table A.14: Principals' Characteristics by ADP Status

	Public Schools			
	Not ADP	ADP	Difference	Private Schools
	(1)	(2)	(3)	(4)
Panel A: Ever worked				
As teacher	0.541	0.441	-0.099	0.431
	(0.498)	(0.497)	(0.000)	(0.495)
As administrative before	0.950	0.964	0.014	0.937
	(0.217)	(0.185)	(0.034)	(0.244)
In private sector	0.009	0.035	0.025	0.229
	(0.096)	(0.183)	(0.000)	(0.420)
Panel B: Principal characteristics				
College degree	0.839	0.901	0.063	0.893
	(0.368)	(0.299)	(0.000)	(0.309)
Age	57.214	55.780	-1.434	54.294
	(8.762)	(8.940)	(0.000)	(11.979)
Female	0.490	0.489	-0.001	0.615
	(0.500)	(0.500)	(0.962)	(0.486)
Observations	2,057	1,770	3,827	4,434

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

Table A.15: Market Level Treatments and Principals' Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ever teacher	Ever administrative	Ever principal	Ever in private sector	Age	Female	College degree
Panel A: County							
Post ADP turnover in the market	-0.003	0.000	0.008	0.001	-0.401	-0.008	0.005
	(0.011)	(0.004)	(0.008)	(0.005)	(0.255)	(0.010)	(0.006)
R-squared	0.787	0.530	0.589	0.931	0.759	0.787	0.795
Panel B: School market							
Post ADP turnover in the market	-0.013	0.006	0.003	0.011	-0.684	0.000	0.016
	(0.019)	(0.008)	(0.014)	(0.008)	(0.450)	(0.021)	(0.008)
R-squared	0.787	0.530	0.589	0.931	0.759	0.787	0.795
Observations	14,298	16,945	14,298	14,298	16,945	16,945	16,945
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep var	0.320	0.972	0.916	0.191	53.50	0.607	0.886

Notes: This table presents the effects of the new selection system (ADP) on the relevant school market on principals' characteristics in private schools. "Post ADP turnover in the market" is a dummy that takes the value one after the first time a school selects a principal under the ADP system in a given market. The sample includes only private schools. In Panel A, the market is defined as the municipality in which the school is located, while in Panel B, the market is defined using the 7km distance. Robust standard errors clustered at the school level in parenthesis.

Table A.16: 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 (1,000 USD)	2.572	2.575	0.003
	(0.887)	(0.607)	(0.029)
Monthly school wage (1,000 USD)	1.323	1.350	0.027
	(0.270)	(0.243)	(0.009)
Share of disadvantaged students	34.922	62.684	27.762
	(23.404)	(16.754)	(0.717)
Average test scores	247.614	247.668	0.054
	(19.246)	(19.315)	(0.765)
Total enrollment	458.944	432.487	-26.457
	(351.421)	(321.265)	(11.863)
Income per student	8.048	10.696	2.648
	(3.942)	(3.179)	(0.126)
Rural school	0.247	0.221	-0.026
	(0.413)	(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	1,611	1,611	3,222

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 parentheses).

B Additional Specification Checks

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 our data to implement some of the validation exercises proposed in the literature.

First, in the spirit of Chetty et al. (2014) and the omnibus test in Angrist et al. (2017), we present quasi-experimental evidence from an analog to the ideal experiment of random principal assignment to schools. This design exploits principal turnover for identification, thus it rests on the identification assumption that principal turnover within a school is uncorrelated with student and school characteristics.[39] 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 B.1). 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 25% 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 quartile of the distribution of principal effectiveness, and we define exit events analogously. The series in Figure B.1 plots school-year means of standardized course grades in the two years before and after a low value-added principal exits the school. As in Chetty et al. (2014), we do not condition on any other covariates in this figure: each point simply shows average course grades for different years within a school. 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 principal effectiveness cannot be rejected. In all but the last panel (D), the change in course grade gains is significantly different from 0 with p-values < 0.01 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

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

principals enter or exit a school.

Second, 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).[40] 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 in our main specification (including lagged course grades), but excluding teachers' and principals' fixed effects. Figure B.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.

Third, to assuage concerns related to student sorting we follow Jackson et al. (2022) and show that conditional on our school level controls, predicted outcomes based on individual characteristics are unrelated to our measure of principal effectiveness. Specifically, we focus on students who took the SIMCE national exams at some point, for whom we have the following attributes: family income category (low, medium, high), parents' education (college graduate or not), and parents' ethnicity. Then, we predict course grades based on a linear regression of course grades on students' attributes, grade and year. Figure B.3, Panel A, shows a binned scatterplot of the actual course grades against the predicted course grades. The predicted outcomes tracks actual outcomes very well. We then examine if principal effectiveness is correlated with predicted course grades in a regression model with only the time-varying and across-time averaged school characteristics used in our correlated random

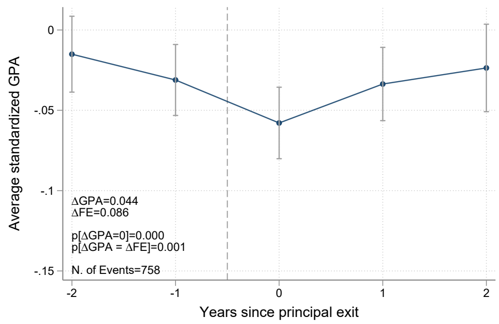
[40] This is because OLS estimates of worker and firm effects will be biased unless worker mobility is uncorrelated with the time-varying residual components of wages, a strong assumption on workers' mobility if one considered some specific models of wage determination (e.g., Gibbons et al., 2005).

effect approach. Figure B.3, Panel B, shows a binned scatterplot of the predicted course grades against our measure of principal effectiveness. We find that, after including our school-level controls, principal effectiveness is not significantly related to predicted course grades.

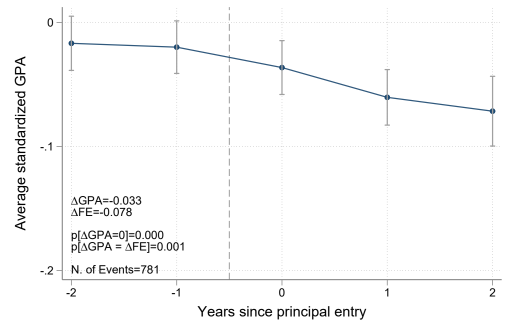
Finally, to complement the previous exercises, we also implement a falsification test similar to that in Rothstein (2010). We focus on a subset of students who switched schools at the end of primary and who were consequently exposed to more than one principal. The intuition of the test is simple: if the effectiveness of the principal in the school of destination s_{t+1} impacts GPA growth in the school of origin s_t, that would be evidence of model misspecification. We consider two sets of students. First, students who were "forced" to switch to another school because their school of origin *did not* offer secondary education. Second, students who switched from schools that *did* offer secondary education. For this exercise, we use "jackknife" estimates of principal effectiveness as the dependent variable, i.e., estimates of principal effectiveness in a sample that leaves out all observations of the students who switched schools. As shown by Table B.1, we fail to find evidence of a positive correlation between growth in course grades (the *pre-assignment* variable) and the effectiveness of their future principal (the *treatment* variable), in both cases.[41]

[41] 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. Consequently, we take this evidence only as suggestive.

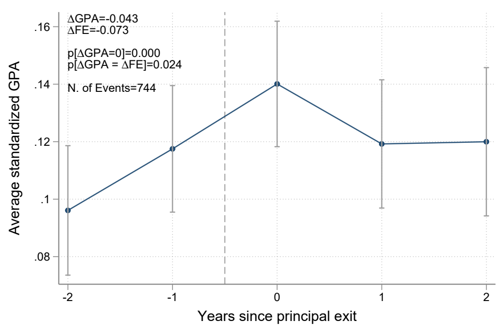
Figure B.1: Impacts of Principal Entry and Exit on Students' Performance



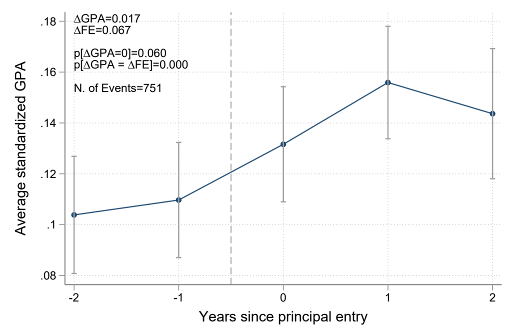
A. Low Value-Added Principal Exit



B. Low Value-Added Principal Entry



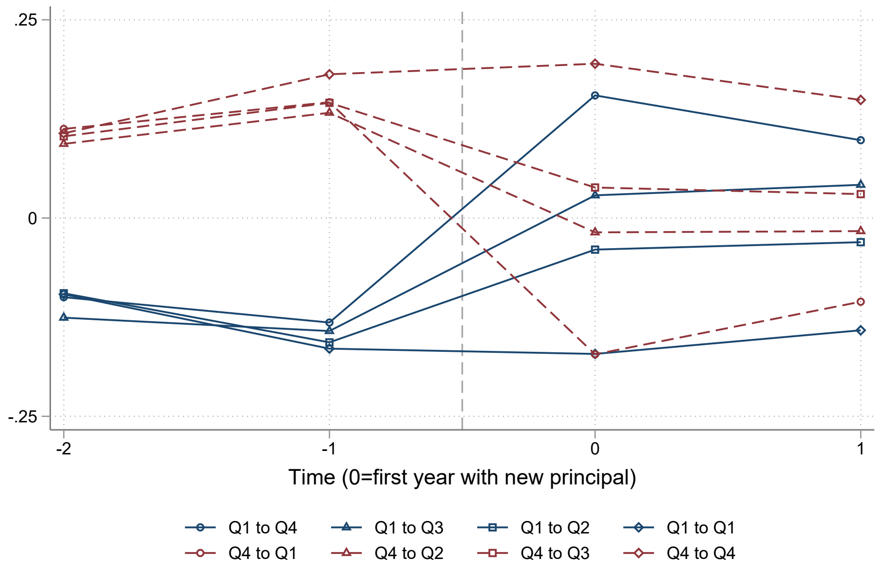
C. High Value-Added Principal Exit



D. High Value-Added Principal Entry

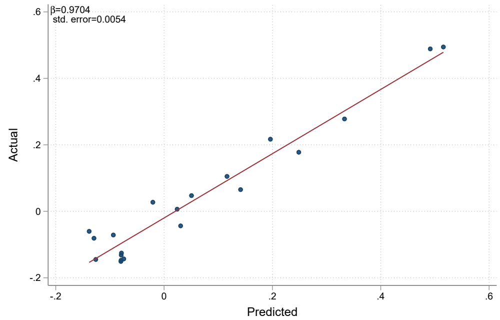
Notes: These figures plot event studies of standardized course grades as principals arrive at or leave a school at year t=0. Panels A and B plot the impact of the exit and entry of a low value-added principal (principals with VA in the bottom 25% of the distribution) on mean course grades. Likewise, Panels C and D plot the impact of the exit and entry of a high value-added principal (principals with VA in the top 25% 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. We only include observations where we observe students before and after the change in principal and plot the average course grade and their confidence interval at the 95% (based on the standard error of the mean) for each relative year across principal's turnover. 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 test 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 B.2: Mean Residualized GPA of Teachers who change Principal, classified by Quartile of Principals' Mean Residualized GPA at Origin and Destination

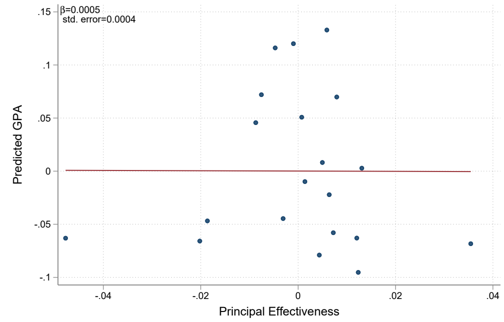


Notes: This figure plots the mean residualized course grades of teachers who changed principal between 2011 and 2016. We consider the first time a teacher switches to work under a new principal, but we do not condition on holding the old or new job relationship for a minimum number of years. Each principal 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 B.3: Predicted Course Grades and Principal Effectiveness



A. Actual GPA and Predicted GPA



B. Principal Effectiveness and Predicted GPA conditional on School CRE

Notes: Panel A shows a binned scatterplot of the actual course grades against the predicted course grades. Panel B shows a binned scatterplot of the predicted course grades (after removing school controls) against our measure of principal effectiveness (after removing school controls). Each figure reports the coefficient from a regression of the y-axis variable on the x-axis variable. In panel B we leverage the Frisch–Waugh–Lovell theorem to obtain the β coefficient. We predict course grades based on a linear regression of course grades on student attributes, which are available for students who took the SIMCE national exams at some point.

Table B.1: Falsification Test

	"Forced" Switches		"Non-Forced" Switches	
	$\theta_{p(-i)}$ at school s_{t+1}	$\theta_{p(-i)}$ at school s_t	$\theta_{p(-i)}$ at school s_{t+1}	$\theta_{p(-i)}$ at school s_t
	(1)	(2)	(3)	(4)
Growth GPA	0.001	0.013	0.002	0.015
	(0.001)	(0.002)	(0.002)	(0.003)
Observations	96,448	96,448	39,257	39,257

Notes: This table shows the results from the validation exercise discussed above. We consider a sample of students who switched schools at the end of primary. In columns 1 and 2, we consider students "forced" to switch because their school *did not* offer secondary education. In columns 3 and 4, we consider students who switched from schools that *did* offer secondary education. For this exercise, we use "jackknife" estimates of principal effectiveness as the dependent variable, i.e., estimates of principal effectiveness in a sample that leaves out all observations of the students who switched schools. Robust standard errors clustered at the county level are presented in parentheses.

C Principals' Labor Market

C.1 Descriptive Analysis of Wages at Public and Private Schools

Public sector compensation usually does not include pay for performance (Finan et al., 2017), and although there is a good rationale for this,[42] 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 C.1 presents some 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:

$$\ln(\text{wage}_{pt}) = \alpha + \beta_0 \text{Voucher}_{pt} + \beta_1 [X_{pt} - \bar{X}] + \beta_2 \text{Voucher}_{pt} \times [X_{pt} - \bar{X}] + \rho_{m(p,t)} + \gamma_t + \epsilon_{pt}, \quad (4)$$

where $\ln(\text{wage}_{pt})$ represents the logarithm of the average hourly wage paid to principal p at time t, Voucher_{pt} is an indicator that equals one if the principal works at a voucher-private school (and zero otherwise), γ_t are year fixed effects, and $\rho_{m(p,t)}$ is a fixed effect at the level of the municipality in which principal p works at time t. The parameter of interest is β_2, and it represents the factor price differential between sectors. Importantly, the vector X_{pt} includes principal characteristics such as our measure of her effectiveness $\hat{\theta}_p$, tenure, tenure squared, an indicator for whether the principal is female, and for whether she has a permanent contract. This specification also allows us to study how the different components of wages relate to principal effectiveness. For this, we decompose the dependent variable $\ln(\text{wage}_{pt})$ into two components: $\ln(\text{base}_{pt})$ and $\ln(\text{wage}_{pt}/\text{base}_{pt})$, where "base" corresponds to the sum of the minimum legal wage and the statutory payments described in section 2,

[42] 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).

and base_{pt} corresponds to the total wage minus the bonuses.

Table C.1 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 14% 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 7.5% 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 (2022) 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 the next subsection 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.

C.2 Two-sided Selection 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).[43] The timing of the model is the following:

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

The School's Decision

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

$$U_j(i) = m_j + \beta_j \theta_i + \epsilon_{1ij}, \tag{5}$$

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

$$U_j(\neg i) = s_j + \epsilon_{0ij}, \tag{6}$$

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

[43] 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 (5) is greater in value than expression (6), 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 (5) and (6). If ϵ_{1ij} and ϵ_{0ij} are *iid* type I extreme value, then the difference will follow a logistic distribution, and the probability that j will make an offer is given by:

$$Pr(o_{ij}) = \frac{\exp(\beta_{0j} + \beta_j \theta_i)}{1 + \exp(\beta_{0j} + \beta_j \theta_i)}, \tag{7}$$

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

The Worker's Decision

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

$$Pr(S_{ik}) = \prod_{m \in O_k} Pr(O_{im} = 1) \prod_{n \in \bar{O}_k} Pr(O_{in} = 0), \tag{8}$$

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:

$$V_{i(j)} = h_j + w_j \theta_i + v_{ij}, \tag{9}$$

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:

$$Pr(A_{ij} \mid O_k) = \begin{cases} \frac{\exp(h_j + w_j \theta_i)}{\sum_{h \in O_k} \exp(h_h + w_h \theta_i)} & , \quad j \in O_k \\ 0 & , \quad j \notin O_k. \end{cases} \tag{10}$$

Given our assumptions about the distribution of the random components in (5), (6), and (9), 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:

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

Numerical Simulation

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

$$\begin{aligned} Pr(A_{ip} \mid \theta_i) = & \Bigg(\frac{\exp(h_p + w_p\theta_i)}{\exp(h_p + w_p\theta_i) + \exp(h_v + w_v\theta_i)} \\ & \times \frac{\exp(\beta_{0p} + \beta_p\theta_i)}{1 + \exp(\beta_{0p} + \beta_p\theta_i)} \times \frac{\exp(\beta_{0v} + \beta_v\theta_i)}{1 + \exp(\beta_{0v} + \beta_v\theta_i)} \Bigg) \\ & + \left(1 \times \frac{\exp(\beta_{0p} + \beta_p\theta_i)}{1 + \exp(\beta_{0p} + \beta_p\theta_i)} \times \frac{1}{1 + \exp(\beta_{0v} + \beta_v\theta_i)} \right). \end{aligned} \tag{11}$$

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

$$\mathrm{E}[\theta_i \mid \text{Public}] = \int_\theta \theta_i f_{\theta|\text{Public}}(\theta_i \mid \text{Public}) d\theta \tag{12}$$

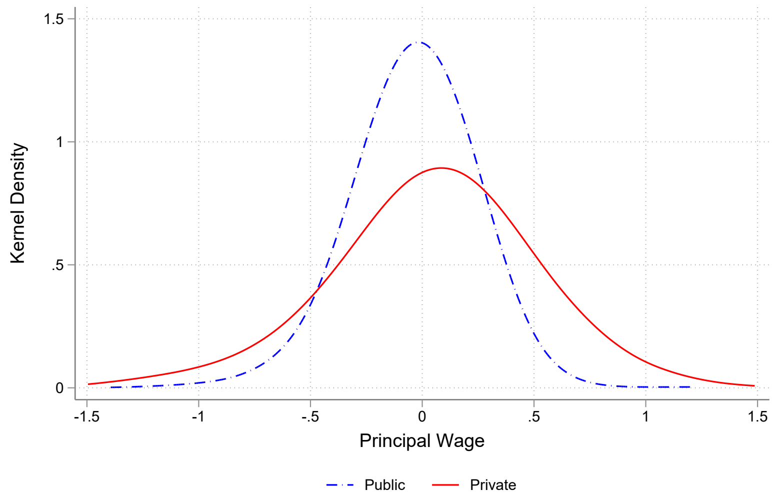
From Bayes' rule, we know that:

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

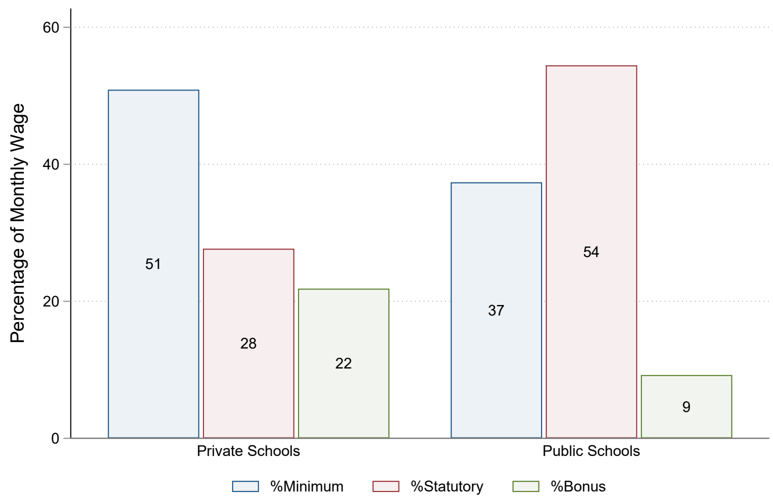
where $Pr(A_{ip} \mid \theta_i)$ is given by (11) and $Pr(\text{Public})$ is a scale factor equal to the fraction of public schools (0.5 in this case). Assuming that $f_\theta(\theta_i)$ is a standard normal, we can compute $\mathrm{E}[\theta_i \mid \text{Public}]$ using numerical integration. More importantly, we can study how this object depends on β_p and w_p, the two relevant parameters related to selection and payment policies in public schools, respectively.

Our simulation is presented in Figure C.2. 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 (12). To construct this figure, we created a grid for β_p and w_p from 1 to 10, and compute $\mathrm{E}[\theta_p \mid \text{school type: Public}]$ for each cell of this grid.

Figure C.1: Principals' Wages



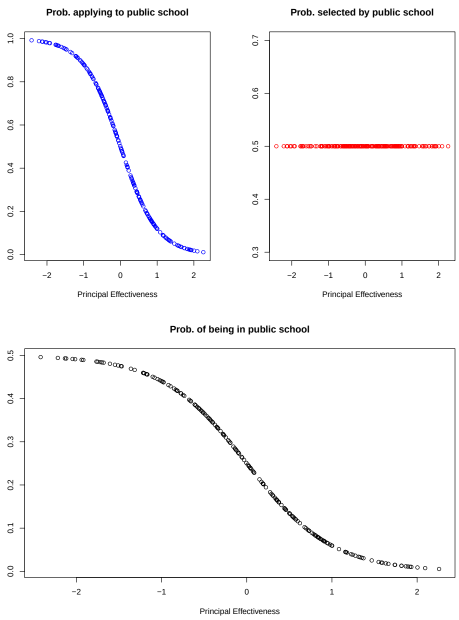
A. Residualized Log Wage



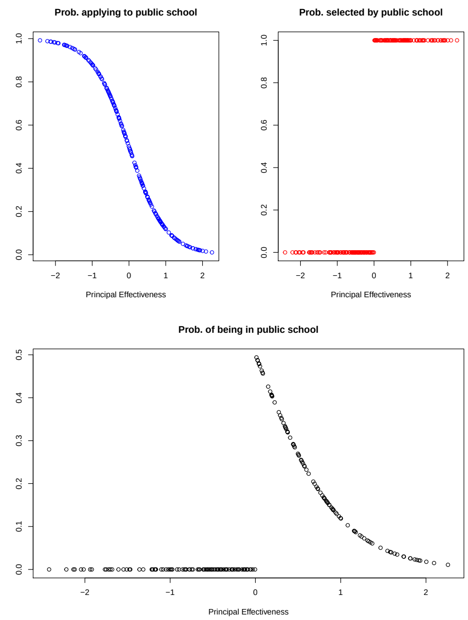
B. Wage Components

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

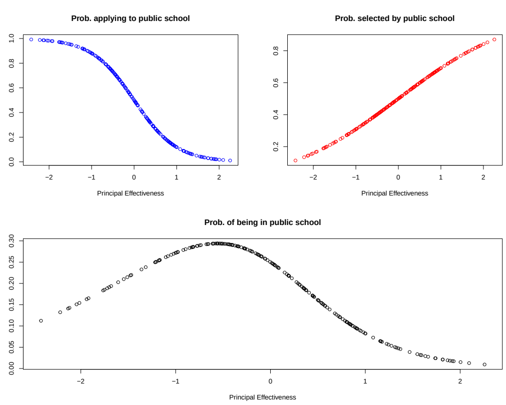
Figure C.2: Simulation of a two-sided matching model



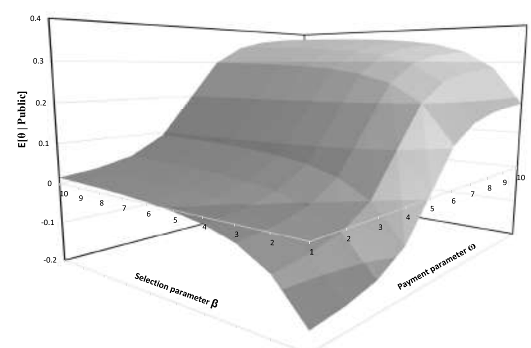
A. Selection at Random



B. Selection Above Cutoff



C. Selection on Quality



D. Allocation of θ_p if Selection on Quality

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

Table C.1: 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	0.139	-0.171	-0.129	0.276	0.268
	(0.008)	(0.009)	(0.010)	(0.012)	(0.008)	(0.013)
Principal Effectiveness	-0.016	-0.020	0.004	-0.003	-0.020	-0.018
	(0.015)	(0.015)	(0.021)	(0.019)	(0.015)	(0.015)
Principal Effectiveness × Private	0.087	0.075	-0.003	-0.008	0.091	0.083
	(0.033)	(0.030)	(0.039)	(0.036)	(0.030)	(0.029)
Female		-0.003		0.010		-0.014
		(0.007)		(0.011)		(0.007)
Female × Private		-0.114		-0.078		-0.036
		(0.016)		(0.020)		(0.015)
Age		0.032		0.043		-0.011
		(0.005)		(0.007)		(0.004)
Age × Private		0.005		-0.020		0.025
		(0.007)		(0.010)		(0.006)
Age^2		-0.000		-0.000		0.000
		(0.000)		(0.000)		(0.000)
Age^2 × Private		-0.000		0.000		-0.000
		(0.000)		(0.000)		(0.000)
Perm. Contract		0.067		0.078		-0.010
		(0.015)		(0.016)		(0.011)
Perm. Contract × Private		0.136		0.063		0.074
		(0.033)		(0.033)		(0.021)
Hours Contract		0.005		0.012		-0.008
		(0.005)		(0.005)		(0.008)
Hours Contract × Private		0.026		0.017		0.009
		(0.006)		(0.006)		(0.009)
College Degree		0.032		0.040		-0.008
		(0.016)		(0.022)		(0.014)
College Degree × Private		-0.019		-0.003		-0.016
		(0.026)		(0.034)		(0.024)
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 (4). We focus on a sample of principals for whom we have a 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 parentheses.

D 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 four 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' scores 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 choose 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 more 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 resources. 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.

School Finance: The Superintendency of Education provided access to school finance for the year 2016. These records correspond to reports that every school receiving vouchers must provide to the Superintendency in order to report the use of public resources.

Teacher surveys: The Ministry of Education provided access to the survey responses of teachers. Every time students take the nationwide standardized exam SIMCE, teachers must fill out 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 a 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 through 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.

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.

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

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 (2022): "Flexible wages, bargaining, and the gender gap," *The Quarterly Journal of Economics*, 137, 215–266.

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.

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

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

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

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.

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.

Cuesta, J. I., F. González, and C. L. Philippi (2020): "Distorted quality signals in school markets," *Journal of Development Economics*, 147, 102532.

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

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

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

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

Finan, F., B. A. Olken, and R. Pande (2017): "The personnel economics of the developing state," *Handbook of Economic Field Experiments*, 2, 467–514.

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.

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

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.

Jackson, C. K., S. C. Porter, J. Q. Easton, and S. Kiguel (2022): "Who Benefits From Attending Effective High Schools?" .

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

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

Neilson, C. (2019): "Targeted vouchers, competition among schools and the academic achievement of poor students," *Working Paper*.

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

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

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

Roth, J. (2022): "Pretest with caution: Event-study estimates after testing for parallel trends," *American Economic Review: Insights*, 4, 305–22.

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

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

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

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