Economics of Education Review 111 (2026) 102771

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Rule-based civil service: Evidence from a nationwide teacher reform in Mexico☆

Juan Bedoya [a], Rafael de Hoyos [b,*], Ricardo Estrada [c]

[a] VélezReyes+, Colombia
[b] Inter-American Development Bank, United States of America
[c] CAF–Development Bank of Latin America and the Caribbean, Mexico

ARTICLE INFO

JEL classification:
I21
J24
J45
M5

Keywords:
Civil service
Hiring methods
Teacher hiring
Teacher quality
Mexico

ABSTRACT

This paper studies the effect of a civil service reform on teacher hiring and student achievement in Mexico. The reform replaced discretionary hiring with a rule-based recruitment system, among other changes to the teacher civil service. We find that teachers hired after the reform had higher prior academic achievement. Two channels likely explain this improvement. First, the reform reduced the prevalence of discretionary hires, who were disproportionately drawn from the lower end of the achievement distribution. Second, it enhanced the screening efficiency of rule-based hiring, making prior academic achievement a more important determinant of hiring outcomes. Finally, we show that the reform led to an overall improvement in student achievement in mathematics.

1. Introduction

An efficient bureaucracy is a cornerstone of state capacity and economic development (Fukuyama, 2013). However, history shows that putting that cornerstone in place is often challenging (Grindle, 2012). For example, there is a consensus (going back to the Weberian definition of a professional bureaucracy) that the meritocratic recruitment of public officials is vital to a well-functioning bureaucracy. Yet, international indexes of bureaucratic quality indicate that many governments in low and middle income countries do not recruit public officials meritocratically (Besley et al., 2022). This pattern calls for a better understanding of the mechanisms used to recruit civil servants and the factors shaping their effectiveness.

Teachers are among the most important public sector employees. They are crucial inputs in the quality of educational services and, therefore, critical to productivity and long-term growth. At the individual level, the positive impact of being assigned to a highly effective teacher persists into adult life (Chetty et al., 2014; Rivkin et al., 2005). Furthermore, teachers account for a significant part of public sector employment and education spending. In Mexico, around 24 percent of public sector employees are teachers, according to our calculations using data from the national labor force survey, and teachers' salaries account for 79 percent of public education spending (OECD, 2022).

In this paper, we study the effect of a civil service reform on teacher hiring and student achievement. We do so in the context of a nationwide education reform adopted in Mexico in 2014, which revamped the teachers' civil service. (We refer to this as the "SPD reform" in the rest of the paper.) Among other changes, the SPD reform mandated centrally managed, competitive examinations to determine hire decisions, proscribing a discretionary system in which local officials and teachers' union representatives took such decisions at the state level. The SPD reform scaled up a previous reform that introduced competitive examinations for teacher hiring in response to criticisms about opacity, corruption, and absence of merit in the discretionary system (Estrada, 2019).

☆ This paper was written under an agreement with the Federal Education Ministry of Mexico and the National System of the Professional Teaching Service, allowing us to access data without compromising confidential information. We thank the Editor Daniel Jones, two anonymous referees, Felipe Barrera-Osorio, Gianmarco Leon-Cilliota, Maria Lombardi, Marcos Fernández-Gutiérrez, Mauricio Romero, and seminar and workshop participants at the Inter-American Development Bank, Universidad del Pais Vasco, AEDE Meetings, LACEA Impact Evaluation Network Meeting, LACEA Conference Bogota, RISE Conference at Oxford University, and Valle Nevado Workshop on Economics of Education. We are grateful to Florencia Buccari, Janina Cuevas, Mario Lechuga, Veronica Michel, and Camila Monzon for their excellent research assistance. Part of this research was conducted while de Hoyos was at the World Bank.

* Corresponding author.
E-mail addresses: Juanguillermo@velezreyesmas.com (J. Bedoya), rafaelde@iadb.org (R. de Hoyos), restrada@caf.com (R. Estrada).

https://doi.org/10.1016/j.econedurev.2026.102771
Received 8 October 2024; Received in revised form 7 January 2026; Accepted 23 January 2026

Our analysis of teacher hiring focuses on the prior academic achievement of newly hired teachers. Teacher quality is multidimensional, but several papers show that teachers with higher academic achievement are generally more effective—they add more value to student achievement (Gronqvist & Vlachos, 2016; Hanushek et al., 2019; Neilson et al., 2019; Rockoff et al., 2011). Furthermore, hiring higher-achieving teachers is particularly relevant in our context. Data from an international assessment of adult literacy and numeracy reveal that many teachers in Latin America, including Mexico, have very low levels of these competences (Estrada & Lombardi, 2023). For example, around half of teachers in Latin America struggle with paraphrasing and making low-level inferences from text. They also struggle with simple calculations involving common decimals and have difficulty interpreting basic data from text and tables. As teaching involves reading, writing, and making basic calculations on a regular basis, it would be surprising if teaching effectively did not require having at least basic levels of literacy and numeracy skills.[1]

For the empirical analysis, we use personnel data to construct a dataset of primary and lower-secondary public school teachers hired in Mexico between 2012 and 2017. We link these data with results from the competitive examinations used to select new teachers during the same period. In this way, we can identify teachers hired through the discretionary and rule-based processes before and after the SPD reform, plus unsuccessful applicants to the rule-based process. We measure teachers' prior academic achievement using their scores in "ENLACE Media Superior" (hereafter ENLACE), a census-based national standardized test applied at the end of upper secondary school (grade 12) between 2008 and 2013. Prior research shows that ENLACE test scores are strong indicators of cognitive skills (de Hoyos et al., 2021). We focus the analysis on recent university graduates.

Results show that teachers hired after the SPD reform had higher prior academic achievement than those hired before the reform. The mean ENLACE score of new teachers increased between 2.7 to 3.8 percentile points after the reform. To illustrate the magnitude, the average new teacher scored at the 65th percentile of the population distribution before the reform, and between the 68th and 69th percentiles afterwards. This improvement was driven primarily by gains at the bottom of the achievement distribution. The ENLACE score at the 0.1 quantile of new teachers increased by 4.6 to 6.2 percentile points following the reform, with no significant change observed at the 0.9 quantile.

Since the SPD reform was implemented at the national level, the above effects are estimated using time variation. However, we provide evidence supporting their causal interpretation. Importantly, our analysis rules out the main threat to identification: the potential influence of secular trends or shocks that could affect the pool of individuals from which teachers are selected. We can address this concern because we have direct data on the prior academic achievement of most applicants for teaching positions in Mexico.

We identify two main mechanisms that likely explain the improvement in hire quality. First, the SPD reform substantially increased the share of teachers hired through competitive examinations—from 63.6 percent in 2012 to 76.8 percent in 2014 and 86.4 percent in 2016. This shift mechanically raised the mean prior achievement of new hires, as rule-based hires had ENLACE scores 3.4 percentile points higher than discretionary hires even before the reform. Discretionary hires were disproportionately drawn from the lower end of the achievement distribution.

Second, the implementation of the SPD reform was followed by a shift toward hiring higher-achieving applicants: the mean ENLACE score of rule-based hires increased by about 5 percentile points in the first year of implementation. This improvement was largely driven by enhanced rule-based screening. Under the SPD reform, a one-percentile-point increase in an applicant's ENLACE score was associated with a 0.43 percentage-point higher likelihood of being hired, compared to just 0.13 percentage points before the reform.

We find no evidence of substantial changes in the ENLACE scores of rule-based applicants following the reform. This alleviates concerns that the observed improvement in hire quality was driven by changes in the applicant pool independent of the reform.[2]

Finally, we examine the overall effect of the SPD reform on student achievement using a difference-in-differences (DID) strategy. Specifically, we compare the evolution of student test scores in public and private schools before and after the reform, treating private schools as a control group since the SPD reform only applied to public schools. Our analysis draws on data from the Programme for International Student Assessment (PISA), which evaluates the reading and mathematics skills of 15-year-old students.

We find a positive effect of the SPD reform on mathematics scores, with point estimates ranging from 18.6% to 15.4% of a PISA standard deviation (SD). In contrast, we find no statistically significant effects on reading scores. Importantly, we do not observe differential pre-trends in student test scores by school type or changes in school composition following the reform, which supports the validity of our identification strategy.

This paper contributes to several literatures at the intersection of education, personnel, and public economics. Our main contribution is to the literature on teacher hiring. Our paper is closely related to a set of studies that evaluate the impact of rule-based hiring on student achievement in the context of recent teacher reforms in Latin America—Brutti and Torres (2021) and Busso et al. (2024) for Colombia, Araujo et al. (2020) for Ecuador, and Estrada (2019) for Mexico (studying the period previous to the SPD reform). Among these papers, Estrada (2019), Araujo et al. (2020), and Brutti and Torres (2021) find that rule-based hires are more likely to improve student achievement than discretionary hires. These reforms share a common feature: they were implemented, in part, due to criticisms of the abuse of discretion in the prevailing hiring regimes.

In the closest paper to ours, Busso et al. (2024) find – like we do – that the introduction of test-based screening led to the hiring of higher-achieving teachers. However, they also report a negative overall effect of the Colombian reform on student achievement. This difference appears to be context-specific. In Colombia, rule-based hiring was applied not only to vacancies created by retirements and other regular separations, but also to replace a large share of incumbent teachers who held temporary contracts. This led to significant teacher turnover and the departure of many experienced teachers—a pattern to which the authors attribute the reform's overall negative impact. This dynamic was not present in Mexico, where incumbent teachers generally held permanent contracts and were not massively replaced by rule-based hires.

Our main contribution to this literature is to document the selection patterns along the past achievement distribution shaping the profile of new hires and the mechanisms (compliance, self-selection, and screening) determining the effectiveness of rule-based hiring. Policy discussions often focus on the need to attract highly skilled individuals into teaching. However, the evidence presented here indicates that rule-based hiring can enhance the composition of the teaching force in a different way: by restricting the entry of individuals with very low skill levels into the profession.

[1] The abundance of teachers with low levels of skills seems to be common across developing countries. See a related review in Crawfurd and Pugatch (2020).

[2] The SPD reform limited the use of discretionary hiring to temporary positions, which in some cases were available only when no rule-based applicant was willing to accept them. These conditions may have negatively affected the quality of discretionary hires – which decreased after the reform – thus contributing to a lower overall effect on the quality of new hires.

We also contribute to the broader literature on personnel economics of the state and, more specifically, to a recent strand of papers that study the effect of rule-based and discretionary hiring on the profile of civil servants in Latin American countries (Brassiolo et al., 2020, 2021; Colonnelli et al., 2020; Dahis et al., 2023; Munoz & Prem, 2021). Our paper also connects to the literature on civil service reforms, which mainly focuses on the U.S. context (Aneja & Xu, 2024; Moreira & Pérez, 2024; Ornaghi, 2019; Rauch, 1995; Ujhelyi, 2014). We contribute to this literature by studying a large civil service reform in a different setting. Finally, we contribute to the literature on decentralization, particularly to the studies that highlight that the weak capacity of some local governments can compromise the success of decentralization reforms. (See reviews in Bardhan (2002), Mookherjee (2015).) Our findings point out one factor that could contribute to such weakness: a less meritocratic selection of bureaucrats.

2. Teacher hiring in the public education system

In Mexico, state governments are responsible for the operation of preschool, primary, and lower secondary schools, whereas the federal government defines the national curriculum, monitors the system's performance, among other regulations, and finances a large share of the public education budget. Before 2008, state education ministries and the teachers' union were responsible for the selection and promotion of teachers.[3] In practice, the union had close to full control in the hiring and promotion decisions in many states through, for example, the appointment of state officials who were also union leaders and the inclusion of participation in union activities as a requirement for promotions (Santibanez, 2008). Personnel decisions were characterized by a high use of discretion and opacity. The system was widely criticized for lack of merit in decisions, including the practice of entitlement (which allowed teachers who were retiring to pass on their jobs to relatives) and the selling of teaching positions.

A 2008 reform named the "Alianza por la Calidad de la Educación" or ACE introduced the use of national competitive examinations to hire teachers for public schools. (See details about the ACE reform in Estrada, 2019.) The ACE used test scores from national standardized examinations organized by the Federal authorities as the main criteria for hiring teachers. However, the use of competitive examinations in ACE was not mandatory. The decision of what share of vacancies would be filled by rule-based hiring was jointly taken by state authorities and the teachers' union, and the rest were assigned at the discretion of the teachers' union. The ACE reform did not introduce changes in other dimensions of the teaching career—it was limited to the selection of new teachers.

In 2013, the Mexican Congress enacted a major education reform that included revamping the teaching career. The newly created teachers' civil service (named "Servicio Profesional Docente" or SPD) mandated the use of competitive examinations to determine the hiring of teachers in preschool, primary, and secondary education.[4] The SPD used test scores from standardized examinations as the only criterion to screen applicants and make hiring offers. These assessments were designed specifically for this purpose by the newly created National Institute for the Evaluation of Education (INEE by its Spanish acronym).

Applications were submitted by teaching position type (e.g., general primary school teacher, mathematics lower secondary school teacher) and state. Applicants who scored above a predefined threshold were classified as qualified and ranked by exam score in lists organized by position type and state. During public sessions, authorities announced available school vacancies, and applicants selected positions in order of their ranking. If no qualified applicant chose a given vacancy, state authorities could fill the position through discretionary hiring under temporary contracts—these hires were still required to meet the educational qualifications mandated for the position. Discretionary hiring could also be used to temporarily fill positions open throughout the academic year due to, for example, retirements or other separations. There is anecdotal evidence that some state governments withheld teaching positions (e.g., by reporting vacancies only after the rule-based hiring process concluded) in order to use them for discretionary hiring. Table A.1 in the Online Appendix summarizes the main characteristics of rule-based hiring under both reforms.

The SPD reform sought to place merit at the core of the teaching career, introducing changes not only in hiring but also in tenure, promotion, compensation, and training policies. Before the SPD, teachers hired for permanent positions automatically gained tenure after a probation period of 6 months and 1 day, regardless of whether they were hired through a rule-based or discretionary process. Under the SPD, new teachers followed a tenure-track process, initially receiving a two-year appointment. To attain permanent status, they were required to pass an evaluation at the end of this period—around 1% failed to do so (de Hoyos & Fernández, 2019). The reform included the assignment of tutors – selected from among incumbent teachers – to support new teachers, although the implementation of the tutoring program was incomplete. Additionally, new teachers were required to take a diagnostic assessment at the end of their first year, with mandatory training assigned based on their results. Strong performance in the assessments qualified teachers for salary bonuses of up to 30% of their basic wage. No other changes to compensation policy were made.

The SPD reform also introduced mandatory performance evaluations for teachers with permanent contracts who were hired before the reform. Teachers who failed three consecutive evaluations were subject to reassignment to administrative positions. However, only around 0.3% of the evaluated teachers reached this stage (de Hoyos & Fernández, 2019). Teachers who performed well in the evaluations qualified for the salary bonuses mentioned above.

Like ACE before it, SPD faced fierce opposition from the teachers' union. Despite this and other implementation challenges, the federal Ministry of Education and the National Institute for Education Evaluation implemented in July 2014 the first SPD examination for teacher selection, a process that took place every year until 2018.

In 2019, Mexico abrogated the SPD reform due to a change in the political party in office, bringing this era of education reform to a close—although a modified version of rule-based hiring was kept in place. See Figure A.1 in the Online Appendix for a summary timeline. A detailed description of the reform and the political context is available in Islas, Calef and Aparicio (2021).

3. Data

3.1. Data sources

We obtain information on primary and lower secondary school teachers hired from 2012 to 2017 from the following three databases:

- *Registro Nacional de Maestros*, RENAME: quarterly administrative dataset on school personnel (3rd quarter of 2011–4th quarter of 2012). Identifies teachers hired in the 2012 school year (we refer to the 2012–2013 school year as 2012, and so on).
- *Censo de Escuelas, Maestros y Alumnos de Educación Básica*, CEMABE: census of schools, teachers and students carried out by the National Statistics Office, INEGI (September–December 2013). Identifies teachers hired in 2013.
- *Fondo de Aportaciones para la Nómina Educativa*, FONE: quarterly administrative dataset on school personnel (1st quarter of 2015–2nd quarter of 2018). Identifies teachers hired from 2014 to 2017.

[3] SNTE for its acronym in Spanish. See more about SNTE in Estrada (2019).

[4] The SPD reform did not cover either private schools or upper secondary schools operated by public universities.

The Federal Ministry of Education assembles the RENAME and FONE databases. They cover all public school teachers paid with federal funds, except those based in Mexico City schools. No official documents report the total number of teachers on the payrolls of state governments. However, according to our estimates, they accounted for around 13 percent to 15 percent of the teachers hired in 2013. Most teachers held positions funded by the federal government and are part of our analysis.

Information on the individuals who applied under the rule-based hiring process is based on micro-data from ACE for 2012–13 and SPD for 2014–17. There are no comparable records of "applicants" to the discretionary process. To the best of our knowledge, there is no documentation of the process or criteria used to select and promote teachers in the discretionary process.

The ENLACE exam is our measure of teachers' prior academic achievement. ENLACE was a census-based standardized test applied to students in their final year of upper secondary school (grade 12) between 2008 and 2013. Other versions of ENLACE – not included in our analysis – were administered at the primary and lower secondary levels. The test consisted of two components: mathematical and reading ability. The National Center for the Evaluation of Higher Education (CENEVAL by its Spanish acronym) was responsible for designing the exam, following a rigorous process aligned with the national curriculum and international assessment frameworks such as PISA.

Mathematical ability was defined as the aptitude to recognize and use mathematics in real-world contexts, reason mathematically, and engage with mathematical ideas in a reflective and informed way. The test covered four content area – quantity; space and form; change and relationships; and basic mathematics – and evaluated students' ability to solve problems using three cognitive processes: reproduction, connection, and reflection. Reading ability was defined as the capacity to understand, use, and reflect on written texts for personal, academic, and social purposes. The exam included four types of texts (argumentative, narrative, expository, and persuasive) and assessed a range of cognitive processes organized by complexity. These included information retrieval, comprehension, interpretation, and the reflection and evaluation of both content and form (CENEVAL, 2008).

ENLACE was a low-stakes exam. Its main purpose was to provide feedback to stakeholders in the education system, and its results did not have consequences for students' graduation or admission into the next schooling level. Participation in ENLACE was optional, but more than 90 percent of all students sat for the test each year.

de Hoyos, Estrada, and Vargas (2021) find that ENLACE scores are strong predictors of both educational attainment and labor market outcomes. Exploiting a large sample of twins to control for all between-family differences in school, household, and neighborhood inputs, the authors show that primary school test scores significantly predict secondary education trajectories. Using longitudinal data that link ENLACE results to later outcomes, they further demonstrate that secondary school test scores predict university enrollment and hourly wages. These findings provide robust evidence that ENLACE effectively captured the cognitive skills it was designed to measure, validating its use as a proxy for individual ability.

3.2. Merging and samples

Using the RENAME, CEMABE, and FONE databases, we construct a data set of new teachers that includes their name, population ID (CURP), taxpayer ID (RFC), gender, date of birth, type of teaching position(s), assigned school(s), and year of hire. We track individuals across these sources using their population and taxpayer identifiers. We define new teachers as those individuals who do not appear in any record from previous school years, as the datasets do not include explicit information on hiring dates.[5] Based on this approach, we identify a total of 181,590 individuals hired as primary and lower-secondary school teachers during the 2012–17 period.

We merge the results from the ACE and SPD entry examinations with the teachers' data and obtain applicants' results in the entry examination, and the type of teaching position they applied to.[6] The data does not include hiring offers or acceptances. Hence, we define rule-based hires as those who apply to the rule-based process, either the ACE or SPD examinations, and are part of the teacher payroll the following school year. Newly hired teachers who are on the payroll but did not participate in a rule-based process are considered discretionary hires.

We use the population ID to merge the teacher and applicant databases to the ENLACE data.[7] We successfully merge 59,345 (32.7%) of the 181,590 new teachers to their ENLACE scores. Because of their age, many of these individuals finished secondary school before 2008, the first year ENLACE was administered. For example, among the 2012 cohort of new teachers – the earliest in our analysis – we are only able to obtain ENLACE score for those hired four years after concluding secondary school.[8] Our analysis therefore focuses on the sample of 24,914 individuals who were hired four years after completing secondary school—Table A.2 in the Online Appendix. We refer to these individuals as "recent graduates". For robustness checks, we also consider teachers hired four or five years after completing secondary school.

Using a subsample of new teachers with a similar age profile to those in our main sample, Panel A in Table A.3 in the Online Appendix reports differences in means between observations that were matched with ENLACE scores and those that were not—74% of teachers in this age range were successfully matched. Differences in age, gender, and out-of-state birth are small. Rule-based hires are more likely to merge with ENLACE scores than discretionary hires. However, within each type of hire, the differences in means between matched and non-matched observations remain small (see Panels B and C in the same table). Among rule-based hires, merged individuals exhibit slightly higher entry-exam scores than non-merged ones (54% of the former scored in the top quartile, compared with 46.2% of the latter).[9]

The final dataset includes information on all new hires—both rule-based and discretionary—for primary and lower-secondary public schools between 2012 and 2017. For rule-based applicants, the data covers the 2013–2017 period, as were unable to merge ENLACE scores for unsuccessful rule-based applicants in 2012 due to data limitations.

Panel A of Table A.5 in the Online Appendix shows descriptive statistics. Both all new teachers and recent university graduates are majority female and were generally born in the state where they were hired. Recent graduates are, on average, seven years younger than new teachers and slightly more likely to be female, to have been born in

[5] Because CEMABE was designed to cover all public school teachers, it includes teachers paid with federal or state funds. To avoid counting state-funded teachers as federally funded teachers, we exclude all teachers in CEMABE who are based in schools that do not appear in either RENAME or FONE.

[6] Because test scores under SPD are only available in brackets corresponding to five performance categories, we construct a corresponding variable for the ACE test scores – reported as a continuous variable – that mimics the distribution of SPD applicants among these categories.

[7] Among the 4.9 million students who sat for the test between 2008 and 2012, 95% have a complete population ID. We transform the individual ENLACE scores into year-specific percentiles of the overall distribution of the mean score in mathematics and literacy.

[8] A teaching degree in Mexico typically requires four years of studies.

[9] Table A.4 in the Online Appendix reports merging rates to ENLACE by hiring year. The merging rate increases over time, especially for the initial cohorts. As we show in Section 5, this pattern does not threaten identification in our setting.

the state where they were hired, and to have entered through a rule-based hiring process. Their ENLACE scores are, on average, at the 67th percentile of the national distribution.

Panel B of Table A.5 in the Online Appendix shows some descriptive statistics for the 468,846 individuals who applied to the rule-based process during the period 2013–17. The majority are female, on average 29 years old, and 14 percent succeeded in being hired. The 55,660 recent university graduates have an even higher proportion of females and are about seven years younger on average. Recent graduates tended to perform well on the standardized test used in the competitive examinations. Thirty-eight percent (66 percent) scored in the top quartile (half) of their entry examination test, and 30 percent of them were hired. They have an average ENLACE score at the 61.5th percentile of the population distribution.

4. Empirical strategy

We are interested in the effect of the SPD reform on the hiring of higher-achieving teachers and in shedding light on its underlying mechanisms. Because the SPD reform had a national reach and was implemented simultaneously across the country, we rely on time variation for our first set of results:

$$Y_{it} = \alpha + \sum_{\pi=2013}^{2017} \beta_\pi \cdot \mathbb{1}[\pi = t] + \Gamma X_{st} + \theta_s + \epsilon_{it} \quad (1)$$

Where Y_{it} is the ENLACE score (in percentiles) of individual i hired at year t. $\mathbb{1}[\pi = t]$ is a vector of dummy variables that indicate the relative time (in years) with respect to 2012. We use 2012 as the baseline year (one year before the reform was announced and two years before it was implemented) to avoid any concern about anticipation effects. β_π are the coefficients of interest. X_{st} is a vector of labor market variables included as controls (the state's GDP per capita, unemployment rate, and the mean wages earned by tertiary-educated individuals working in non-teaching occupations) and Γ is the associated vector of parameters. θ_s is a vector of state fixed effects and ϵ_{it} is an error term. We present heteroskedasticity-robust standard errors.

To explore heterogeneous effects along the distribution of Y_{it}, we complement the main specification (Eq. (1)) with a quantile regression model:

$$Q_\tau(Y_{it}) = \sum_{\pi=2013}^{2017} \beta_\pi(\tau) \cdot \mathbb{1}[\pi = t] + \Gamma(\tau) X_{st} + \theta_s + u_{it} \quad (2)$$

Where $Q_\tau(Y_{it})$ is the ENLACE score of individual i hired at year t at the τ^{th} quantile and $\beta_\pi(\tau)$ and $\Gamma(\tau)$ are vectors of coefficients at the various quantiles.

The effect of the SPD reform in Eqs. (1) and (2) is identified using time variation. Hence, the main threat to identification arises from secular trends or shocks unrelated to the SPD reform affecting the pool of individuals from which new teachers are drawn. We do not have information about long-term trends in the skills profile of the teaching force in Mexico. However, a string of papers has documented a secular decline in the skills profile of individuals interested in pursuing a teaching career; mostly in developed countries (Corcoran et al., 2004; Fredriksson & Öckert, 2008; Nickell & Quintini, 2002), but also in Chile in Latin America (Neilson et al., 2019). In our specification such decline would lead to the *underestimation* of a positive effect of the SPD reform.

More importantly, we can directly observe the prior academic achievement of individuals who apply to the rule-based process, allowing us to study changes in the pool of applicants from which most hires are drawn. One limitation of our setting is that we cannot disentangle whether changes in this pool reflect contemporaneous shocks to applicant quality (which poses a threat to identification) or the potential effect of the SPD reform itself on attracting higher-achieving rule-based applicants (a potential mechanism). However, this limitation is unlikely to be problematic in practice. As Section 6 shows, the implementation of the SPD reform was followed by only a modest change in the prior achievement of rule-based applicants, which – regardless of its source – cannot account for the improvement in the prior achievement of new teachers documented in the next section.

Furthermore, we use a dynamic specification of treatment effects – instead of a simple before-and-after specification – which allows us to explore in a desegregated way the evolution of the new teachers' profiles. Similarly, by using the year-specific percentiles of ENLACE scores as the outcome variable, we concentrate on changes in the relative position of newly hired teachers in the national distribution of test scores, holding constant changes in the distribution across years.

5. The SPD reform and teacher hiring

We find a clear improvement in the hiring of higher-achieving teachers after the SPD reform. Fig. 1(a) reports the estimated changes relative to 2012 in the average ENLACE scores of the teachers hired each year—based on Eq. (1).[10] The point estimates for the post-reform years range from 2.7 to 3.8 percentiles points and are all statistically significant at the 95 percent level. This implies that, while the average teacher hired before the reform had an ENLACE score at the 65th percentile of the population distribution, the average teacher hired after the reform ranked between the 68th to 69th percentiles. We do not find clear evidence of an anticipation effect. The coefficient for 2013 has a small magnitude (−1.0 percentiles) and is not statistically significant.

The documented improvement is largely driven by gains at the lower end of the prior achievement distribution. Fig. 1(b) reports the estimated changes in test scores for teachers at the 0.1 and the 0.9 quantiles, relative to 2012. The point estimates – for the post-reform years – vary from 4.6 to 6.2 percentiles of the ENLACE score and are statistically significant at the 95 percent level. To contextualize, in 2012, new teachers at this quantile had ENLACE scores at the 24th percentile of the national test distribution. In contrast, the estimated changes at the 0.9 quantile are small (around 1 percentile) and not statistically significant. This likely reflects a ceiling effect, as incoming teachers in this quantile already had test scores in 2012 near the top of the ENLACE distribution (equivalent to the 94th percentile).[11]

Due to data limitations, we focus our analysis on teachers hired four years after finishing secondary school. Figure A.4 in the Online Appendix shows that our results are robust to the inclusion of individuals hired five years after finishing secondary school, using 2013 as the baseline year. Unfortunately, we cannot go beyond this extension.

As reported in Section 3, the merging rate to ENLACE increased over time, which improves coverage of the overall population and can enhance statistical power. One concern is that this rise could be mechanically drive our results if the "marginal" merged teachers have more favorable characteristics, including higher ENLACE test scores. To assess this, we exploit substantial variation in merging rates across states, over time, with some already exhibiting high rates in 2013. We split states into those with above- and below-median increases in merging rates between 2013 and 2017 and re-estimate our main specification separately for each group (Table A.8 in the Online Appendix). The first group had a merging rate of 78% in 2013 and experienced only a five percentage-point increase between 2013 and 2017, whereas the second group had a rate of 58% in 2013 and a 30 percentage-point increase over the same period. If the rise in merging rates were driving our results, the effects should be concentrated in states with the largest increases. Instead, we find significant effects in both groups, with larger estimates in the group with the smallest increase in merging rates. These results show that the improvement in merging rates does not threaten identification.

[10] Figure A.2a in the Online Appendix presents the raw yearly means, and Table A.6 in the Online Appendix reports the point estimates plotted in Fig. 1(a). It also presents results from estimations without the controls for the state-level labor market conditions and using yearly changes in these variables instead of levels. Results are consistent across specifications.

[11] A similar—though less stark—pattern is found from comparing the 0.25 and 0.75 quantiles—see figures A.2b and A.3 in the Online Appendix.

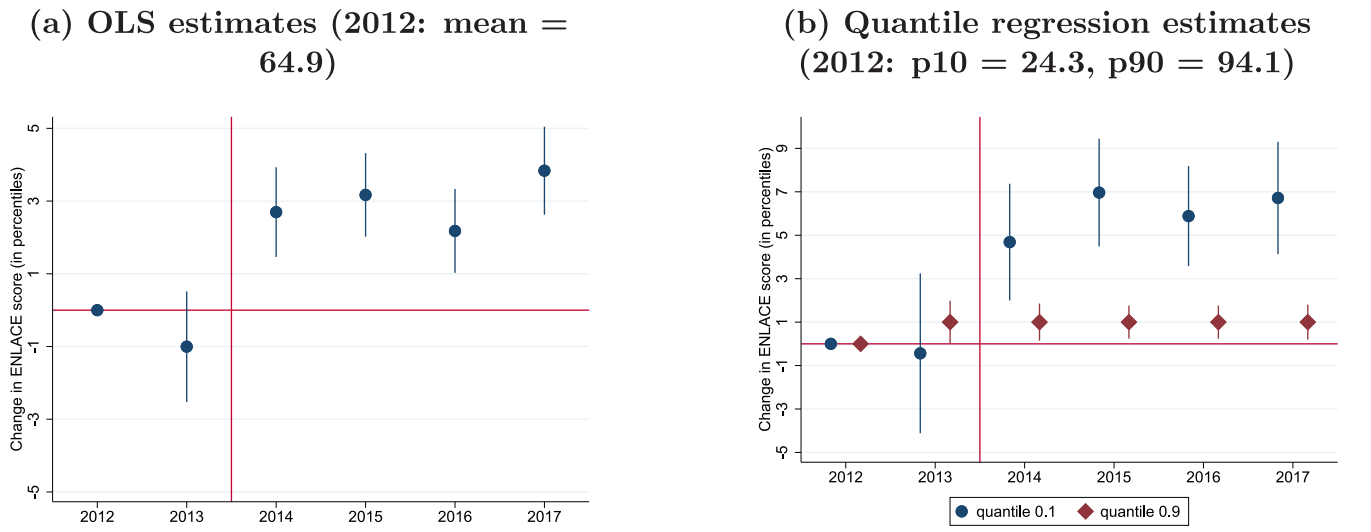


Fig. 1. New teachers: Change in ENLACE scores with respect to 2012.
Notes: Panel (a) shows the difference in percentiles of the ENLACE score of newly hired teachers with respect to 2012, corresponding to the β_π estimates of Eq. (1). Panel (b) shows the difference in percentiles of the ENLACE score of newly hired teachers for quantile 0.1 in blue and quantile 0.9 in red with respect to 2012, corresponding to the $\beta_\pi(0.1)$ and $\beta_\pi(0.9)$ estimates of Eq. (2). Regressions include state fixed effects, state job market controls, and robust standard errors. Confidence intervals at the 95 percent level are shown in bars.

6. Why did the hiring of higher-achieving teachers improve after the SPD reform?

Our results suggest that the rule-based personnel policies introduced under the SPD reform led to the hiring of higher-achieving teachers. We now turn to evidence supporting a causal interpretation and shedding light on the underlying mechanisms. To guide the analysis, consider the following decomposition equation:

$$\bar{Y}_t = \lambda_t \bar{R}_t + (1-\lambda_t)\bar{D}_t \tag{3}$$

Where $\bar{Y}_t$ denotes the mean prior academic achievement level of new hires, λ_t is the share of rule-based hires among new teachers, and $\bar{R}_t$ and $\bar{D}_t$ represent the mean prior academic achievement levels of rule-based and discretionary hires, respectively, at time t. Assuming these variables are differentiable functions of time, consider the following derivative of Eq. (3) with respect to t:

$$\frac{d\bar{Y}_t}{dt} = \frac{d\lambda_t}{dt}(\bar{R}_t - \bar{D}_t) + \lambda_t \frac{d\bar{R}_t}{dt} + (1-\lambda_t)\frac{d\bar{D}_t}{dt} \tag{4}$$

This expression decomposes the change in the average quality of new hires into three components: the first term captures changes in the composition of hires (i.e., shifts in the share of rule-based hires), while the second and third terms reflect changes in the quality of rule-based and discretionary hires, respectively. The sign of the first term depends on the achievement gap between the two groups of hires $(\bar{R}_t - \bar{D}_t)$. In terms of econometric identification, changes in $\bar{R}_t$ and $\bar{D}_t$ may causally result from the SPD reform – through improved screening or self-selection – or from other shocks unrelated to the reform, as previously discussed. We revisit this issue below after presenting some results.

6.1. The increasing prevalence of rule-based hiring

We study first the extent to which the SPD reform effectively increased the prevalence of rule-based hiring in the public education system. Fig. 2 reports the share of new teachers hired through the rule-based method. We make three observations. First, rule-based hiring was the most common way to select teachers even before the SPD reform, with 64 percent of the teachers hired through a competitive examination in 2012, as a result of the ACE reform. Second, the SPD reform significantly increased the prevalence of rule-based hiring—by

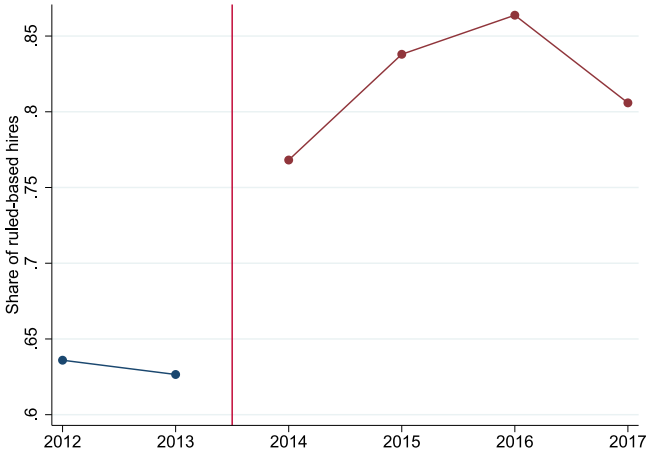


Fig. 2. New teachers: Share of rule-based hires.
Notes: The figure shows the share of rule-based newly hired teachers for each of the sample years. As explained in Section 3.1, we define rule-based hires as individuals who apply to the rule-based process (ACE or SPD) and are hired as teachers the following school year.

14 percentage points in its first year of implementation and by 23 percentage points two years later. (See point estimates in Figure A.5 in the Online Appendix.) Third, the use of discretionary hiring persisted after the SPD reform—albeit to a lesser extent.[12]

6.2. The achievement gap between rule-based and discretionary hires

Fig. 3(a) shows yearly means of ENLACE scores by hiring method. Three observations stand out. First, rule-based hires had higher ENLACE scores than discretionary hires starting from the baseline year,

[12] We might have overestimated the proportion of rule-based hires among total hires. Because we do not observe job offers, we define rule-based hires as individuals who apply to the rule-based process and are hired as new teachers. Our definition might encompass individuals who, despite having participated in a competitive examination, were hired through a discretionary process.

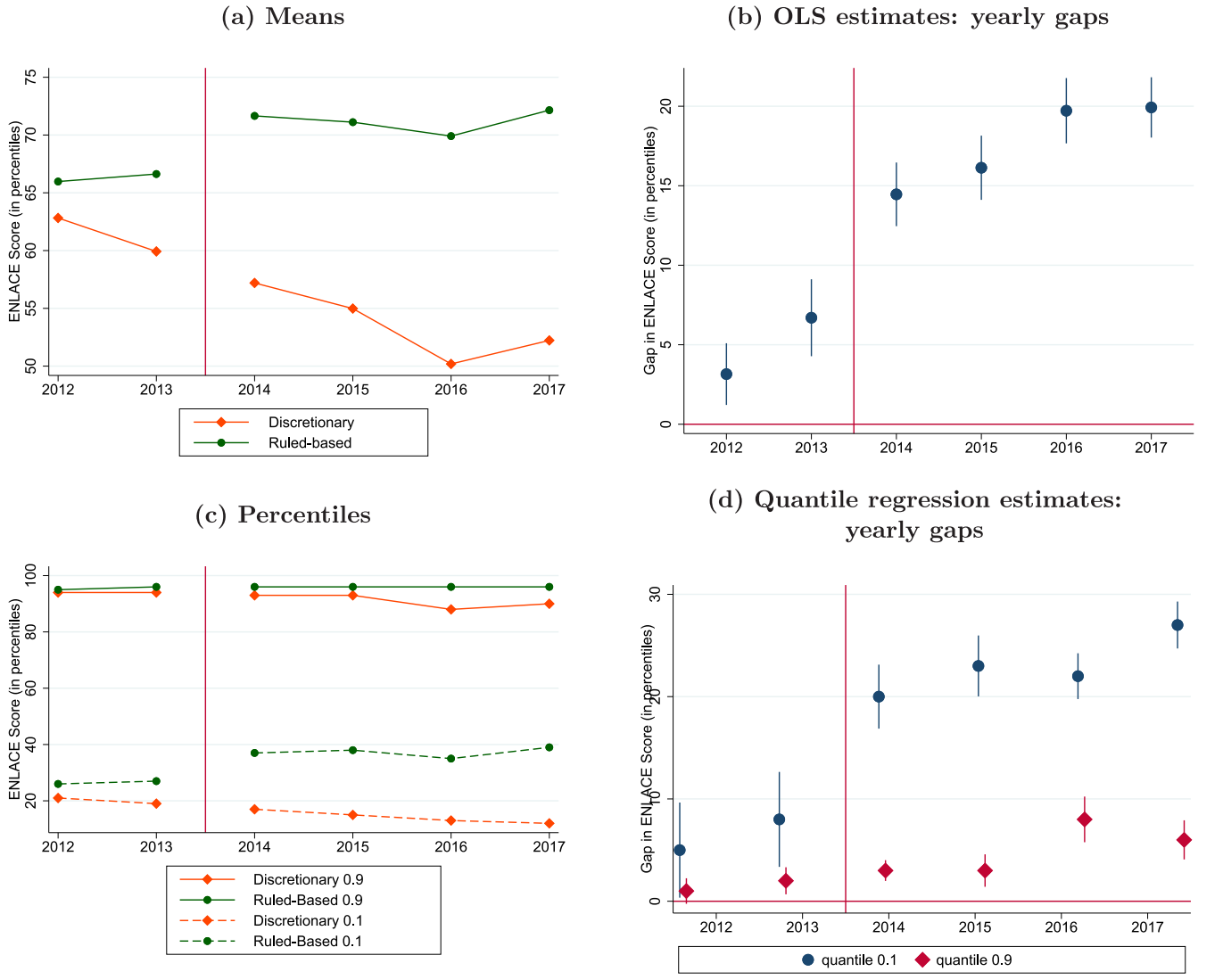


Fig. 3. New teachers: ENLACE scores of rule-based vs. discretionary hires.
Notes: Panel (a) shows the ENLACE score in percentile means for new rule-based and discretionary hires for each year in the sample. Panel (b) shows the estimates of the annual gaps in ENLACE scores in percentiles between new rule-based and discretionary hires; regressions include state fixed effects and robust standard errors. Panel (c) shows the ENLACE score in percentiles at quantiles 0.1 and 0.9 for new rule-based and discretionary hires for each year in the sample. Panel (d) shows estimates of the annual gaps in ENLACE score in percentiles at quantiles 0.1 and 0.9 between new rule-based and discretionary hires; regressions include state fixed effects and robust standard errors. For Panels (b) and (d), confidence intervals at the 95 percent level are shown in bars.

by 3.4 percentiles on average. Second, the scores of rule-based hires showed a marked improvement in the year the reform was implemented, increasing by around 5 percentiles. Third, the scores of discretionary hires followed a downward trend, with their ENLACE scores decreasing by 3.2 percentiles per year from 2012 to 2016. As a result, the gap in ENLACE scores between rule-based and discretionary teachers widened after the reform (Fig. 3(b)). (The point estimate for the increase in the gap between 2013 and 2014 is 4.4 percentiles, with a p-value of 0.01.)

The improvement in the hiring of higher-achieving rule-based applicants is striking. Figs. 3(c) and 3(d) present raw yearly percentiles and quantile regression estimates of the gap in ENLACE scores between rule-based and discretionary hires. The gap at the 0.1 quantile increased from 5 percentiles in 2012 to 20 percentiles in 2014. In contrast, the gap at the 0.9 quantile only changed from 2 percentiles in 2012 to 3 percentiles in 2014.[13]

6.3. *Self-selection and screening in rule-based hiring*

Two channels could explain the documented changes in the profile of new hires: changes in self-selection (in the pool of individuals who apply to a teaching position) and screening. We study the importance of both mechanisms in the rule-based process; we cannot do the same for the discretionary process because we only observe individuals who were hired, not the full pool from which they are selected.

Self-selection

The SPD reform had, at most, a modest effect on the likelihood that higher-achieving individuals applied to the rule-based process. As shown in Fig. 4(a), the average ENLACE scores of rule-based applicants declined slightly—by 1.4 percentile points (p-value of 0.00)—between 2013 and 2014. From 2015 onward, the estimates show no statistically significant changes relative to 2013.

[13] See Figures A.6a and A.6b in the Online Appendix for OLS and quantile regression estimates of the yearly change in the skills gap relative to 2012. See also Figures A.7a and A.7b in the Online Appendix for raw yearly percentiles and quantile regression estimates at the 0.25 and 0.75 quantiles.

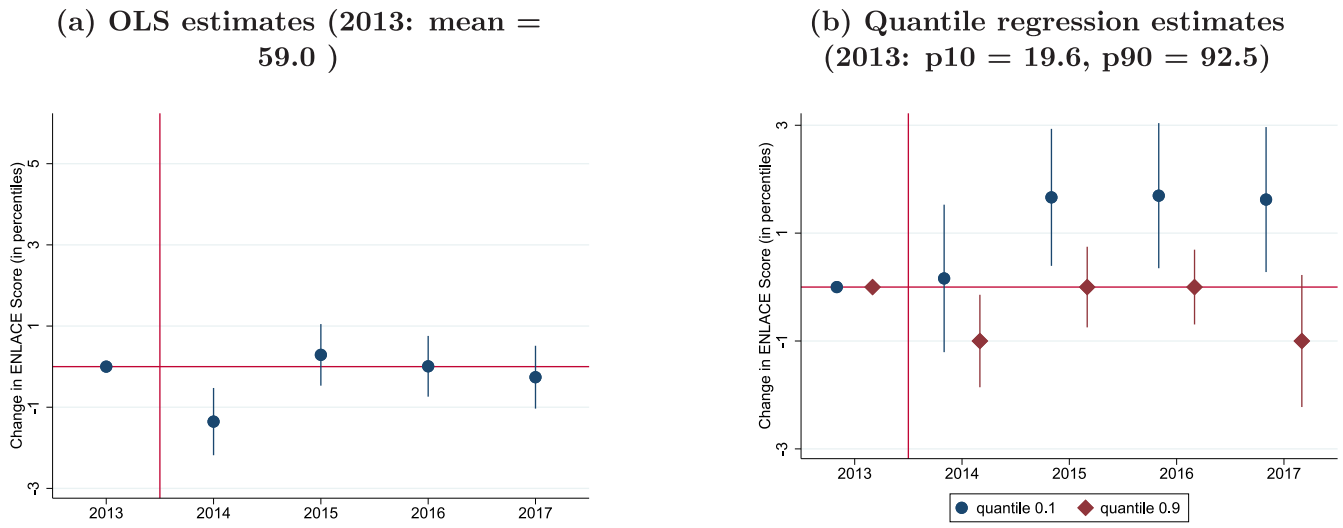


Fig. 4. Rule-based applicants: Change in ENLACE scores with respect to 2013.
Notes: Panel (a) shows the difference in percentiles of the ENLACE score of rule-based applicants with respect to 2013, corresponding to the β_π estimates of Eq. (1). Panel (b) shows the difference in percentiles of the ENLACE score of rule-based applicants for quantile 0.1 in blue and quantile 0.9 in red with respect to 2013 corresponding to the $\beta_\pi(0.1)$ and $\beta_\pi(0.9)$ estimates of Eq. (2). Regressions include state fixed effects and robust standard errors. Bars show confidence intervals at the 95 percent level.

At the lower end of the distribution, there was a modest improvement in the application of higher-achieving individuals. Fig. 4(b) presents the corresponding quantile regression estimates. At the 0.1 quantile, there is no detectable change between 2013 and 2014, followed by a significant increase of around 1.5 percentiles from 2015 onward. At the 0.9 quantile, by contrast, there are significant declines of around 1 percentile in 2014 and 2017 with respect to 2013, and no significant change in 2015 and 2016 with respect to 2013.[14] To interpret the magnitude of these point estimates, note that relative to 2013, the average ENLACE score of rule-based hires in the post-SPD years improved by around 3.4–4 percentiles and by 5–7 percentiles at the 0.1 quantile of the rule-based hires' distribution (Figure A.9 in the Online Appendix).

The modest change in the prior academic achievement of rule-based applicants—relative to the substantial improvement among hires—mitigates concerns that our main results reflect a secular trend or external shock—unrelated to the reform—to the pool of individuals self-selecting into teaching rather than the reform itself.

Enhanced screening

The SPD reform strengthened the screening process in rule-based hiring by making cognitive skills more important determinant of hiring outcomes. Fig. 5(a) plots local means of the probability of being hired among rule-based applicants by ENLACE score percentile, before and after the reform. As shown in the figure, ENLACE scores became a much stronger predictor of hiring under the SPD than under the ACE regime. (Figure A.10 in the Online Appendix reports similar results by year to show that this change is not the result of a secular trend.) While ACE was in effect, a one-percentile-point increase in the ENLACE score was associated with a 0.13 percentage-point increase in the probability of being hired, controlling for state fixed effects (see Table A.9 in the Online Appendix). Under the SPD regime, the corresponding increase was 0.43 percentage points. Although an applicant with an ENLACE score at the 10th percentile had the same probability of being hired under both regimes (10.9 percent), those with ENLACE scores at the 90th percentile had a 20.9 percent chance of being hired under ACE versus 43.5 percent under SPD.

The stronger relationship between applicants' prior academic achievement and hiring outcomes under the SPD regime was mediated by a tighter link between ENLACE scores and performance on the entry examination. Although both the ACE and SPD regimes relied on standardized tests to screen applicants, they differed in the institutions responsible for administering them, as well as in the exams' length and content. Fig. 5(b) plots local means of the probability of scoring in the top quartile or half of the entry-exam distribution by applicants' ENLACE score under both regimes. Table A.10 in the Online Appendix presents the corresponding regression estimates, controlling for state and type of teaching position fixed effects. As shown, the relationship between ENLACE scores and entry-exam performance was considerably stronger under SPD. For example, an ACE applicant with an ENLACE score at the 50th percentile had an unconditional probability of scoring at the top quartile (half) of the entry examination distribution of 9.8 percent (23.6 percent), compared to 21.3 percent (58.0 percent) under the SPD regime.

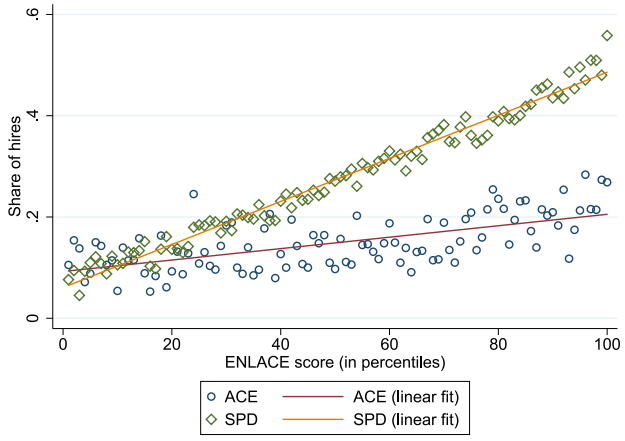
6.4. Discussion

The main threat to identifying the causal effect of the SPD reform on the hiring of higher-achieving teachers arises from positive shocks—unrelated to the reform—to the quality of prospective teachers, particularly those that could explain the observed improvement in the quality of rule-based hires—$(\bar{R}_t - \bar{D}_t)$ in Eq. (4). The evidence presented in this section mitigates this concern, as we only observe a modest change in the ENLACE scores of rule-based applicants. Instead, we document that the hiring of higher-achieving rule-based applicants is mainly driven by enhanced screening.

Returning to Eq. (4), our analysis points to two relevant mechanisms. First, a composition – or extensive-margin – effect: the expansion of rule-based hiring increased the recruitment of higher-achieving teachers because of the pre-existing achievement gap between rule-based and discretionary hires. In other words, the reform reduced the prevalence of discretionary hires, who were disproportionately drawn from the lower end of the test-score distribution. Second, a quality – or intensive-margin – effect: enhanced screening under the SPD reform raised the quality of rule-based hires.

[14] We obtain a similar picture if, instead of focusing only on individuals who finished secondary school four years before applying to a teaching position, we include in the sample those who finished four or five years prior, as shown in Figure A.8 in the Online Appendix.

(a) Probability of being hired by ENLACE score



(b) Probability of obtaining a high entry examination score by ENLACE score

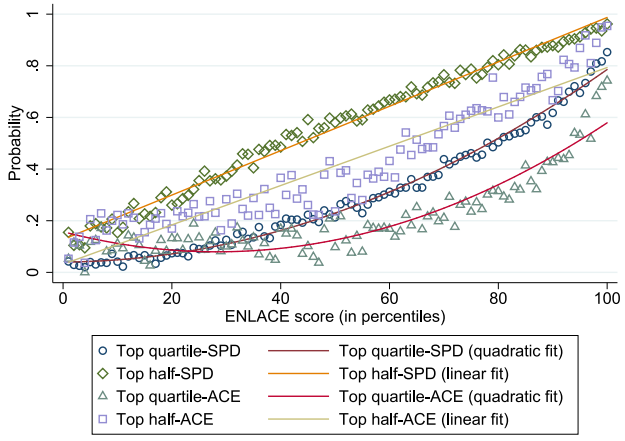


Fig. 5. Rule-based applicants.
Notes: Panel (a) plots local means of the probability of being hired by percentile of the ENLACE score for ACE and SPD rule-based applicants. The lines plot a linear fit by hiring system. Panel (b) plots local means of the probability of scoring in the top quartile and half of the ACE and SPD entry exams by percentile of the ENLACE score. The lines plot a linear (quadratic) fit for the probability of scoring in the top half (quartile) of the entry exam by hiring system.

Notably, the average quality of discretionary hires declined after the reform. This decline may reflect the increasing confinement of discretionary hires to temporary positions – possibly located in less desirable areas – or result from a broader deterioration in their selection or screening processes. Figure A.11 in the Online Appendix shows that, prior to the SPD reform, discretionary hires were indeed less likely to be assigned to schools located in municipalities with medium to very high levels of marginalization. This pattern reverses after the reform, peaking in 2014. The fact that the decline in discretionary hire quality persists beyond 2014 suggests that changes in school allocation are not the main mechanism behind this trend. However, we lack additional information to explore this further.

Using Eq. (4), we perform a back-of-the-envelope calculation to gauge the relative importance of the mechanisms outlined in the equation. For simplicity, we collapse the data into two periods – pre- and post-reform – and plug the resulting simple means into a two-period version of the equation. This exercise suggests that the increase in rule-based hiring accounts for approximately 25% of the overall effect of the SPD reform, the improvement in the achievement of rule-based hires explains about 115%, and the decline in the achievement of discretionary hires contributes roughly −40%.

7. The effect of the SPD reform on student achievement

The ultimate goal of the SPD reform was to improve student learning in the public education system. One key mechanism to achieve this was attracting higher-quality teachers—the central focus of this paper. Additionally, the reform may have improved student learning by enhancing teacher performance through the various incentives and additional resources it introduced. However, the SPD faced resistance from the teachers' union, and the prevailing dissatisfaction among teachers may have negatively impacted their performance, potentially harming student learning. In this section, we examine the overall impact of the SPD reform on student achievement.

For identification, we implement a DID strategy comparing the evolution of student test scores in public and private schools before and after the reform.[15] To measure student achievement, we use data from PISA, conducted by the Organization for Economic Co-operation and Development. PISA assesses the reading and mathematics skills of 15-year-old students in nationally representative samples every three years.[16] More precisely, we estimate the following model:

$$Y_i = \beta_0 + \beta_1 public_i + \beta_2 after_t + \beta_3 public_i * after_t + \Gamma Z_i + \tau_t + u_i \quad (5)$$

Where Y_i is the PISA test score of student i, $public_i$ is a dummy indicating that the student is enrolled in a public school (as opposed to a private school), $after_t$ is a dummy that equals one after the SPD reform was implemented. The term $public_i * after_t$ is an interaction between these two variables, with β_3 as the coefficient of interest. Z_i is a vector of exogenous covariates, τ_t represents time (cycle) effects, and u_i is an error term.

The identification assumption required for a causal interpretation of β_3 is that, in the absence of the SPD reform, student test scores in public schools would have followed a similar trajectory to those in private schools—this is known as the parallel trends assumption.

We estimate equation (5) using the repest command in Stata, which enables us to incorporate the full set of plausible values for PISA test scores and compute appropriate standard errors. Since PISA employs imputation methods to estimate student performance, it introduces imputation error that must be accounted for in the analysis (OECD, 2009).

We use data from the 2006 to 2018 PISA cycles (conducted every three years). Although Mexico has participated in PISA since its first cycle in 2000, we exclude data from 2000 due to the smaller sample size and from 2003 due to a higher number of missing values in the covariates, as well as differential trends between public and private schools from 2003 to 2006. The final sample includes six PISA cycles, with observations from 115,630 students across 1856 schools (we exclude from the analysis 1,842 observations with missing values in covariates). Table A.11 in the Online Appendix provides summary statistics for this sample.

Fig. 6 shows the evolution of average student test scores in public and private schools from 2006 to 2018 (Panel A for mathematics and Panel B for reading). Before the SPD reform, average test scores in public schools were lower than those in private schools. However, and crucially for the identification strategy, both groups exhibited similar trends—a pattern also observed when test scores are normalized to their 2006 mean, as shown in Figure A.12 in the Online Appendix. After the SPD reform was implemented, the test score gap between public and private schools narrowed.

[15] This strategy is similar to that of Busso et al. (2024), who use private schools as a comparison group to evaluate the impact of a reform introducing test-based teacher hiring in the Colombian public education system on student achievement.

[16] Unfortunately, Mexico does not have a national standardized assessment that provides comparable student test scores for the period before and after the SPD reform.

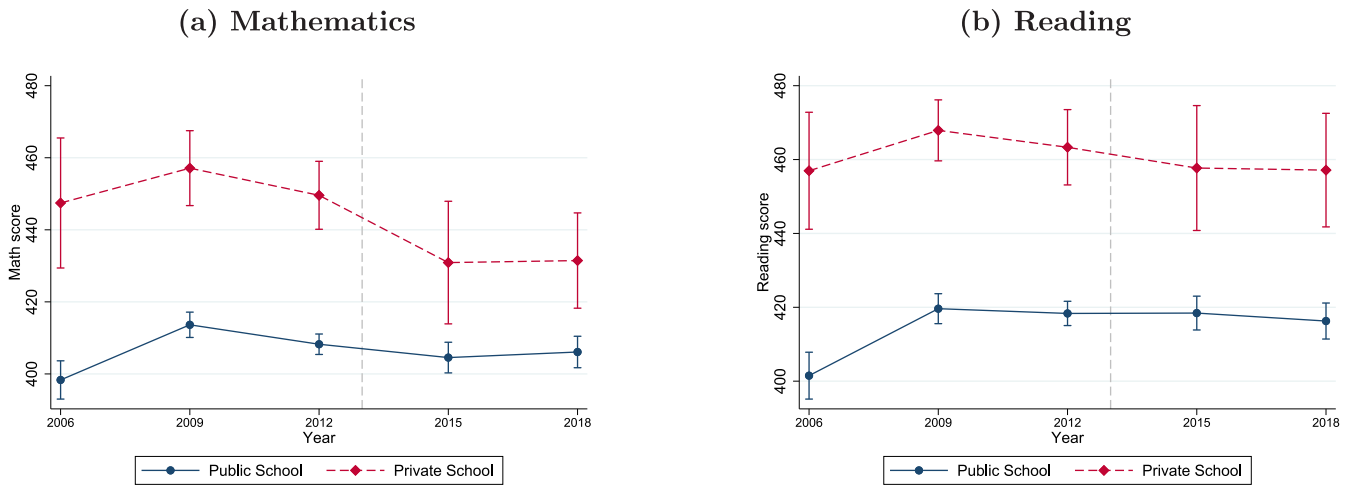


Fig. 6. Mean PISA scores in public and private schools.
Notes: This figure shows the evolution of mean PISA scores in Mathematics (Panel a) and Reading (Panel b) by school type in Mexico. The scores correspond to the first plausible value, and all estimates are weighted using sample weights.

Table 1
DID: Pre-SPD trends in PISA test scores.

	Math			Reading		
	(1)	(2)	(3)	(1)	(2)	(3)
Public school X 2009	5.36	1.91	4.23	9.51	5.37	4.89
	(11.02)	(11.07)	(8.99)	(9.71)	(9.74)	(7.80)
Public school X 2012	9.23	8.79	12.39	13.11	12.50	13.63*
	(11.42)	(10.78)	(8.92)	(10.95)	(10.03)	(8.19)
No. of obs.	101,721			101,721		
Mean public school score 2006	397.96			401.58		
Locality size	No	Yes	Yes	No	Yes	Yes
Student characteristics	No	No	Yes	No	No	Yes

Notes: This table reports regression coefficients from the estimation of a modified version of Eq. (5), which includes interaction terms between a public school indicator and cycle year indicators. The omitted cycle is 2006. Locality size is a vector of five locality size indicators. Student characteristics include a dummy variable equal to 1 if the student is female and a vector of quintile indicators for the PISA Economic, Social, and Cultural Status index. All specifications include cycle fixed effects. Estimates are computed using the repest command in Stata, which accounts for the full set of plausible values and produces appropriate standard errors. All estimates are weighted using sample weights. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2
DID: Changes in school inputs.

	ESCS index (1)	Father has tertiary education (2)	Mother has tertiary education (3)	Student–Teacher ratio (4)	School size (5)
Public	−1.06***	−0.27***	−0.23***	14.2***	1,033***
	(0.06)	(0.02)	(0.02)	(0.94)	(58.08)
After	−0.34***	0.12***	0.14***	6.54***	74.2
	(0.10)	(0.03)	(0.03)	(2.45)	(71.95)
Public X After	0.16	0.029	0.015	−0.25	23.3
	(0.10)	(0.03)	(0.03)	(2.66)	(83.80)
No. of obs.	115,630	107,755	107,755	98,380	114,304
Mean public school 2012	−1.3	0.23	0.21	32.06	920
Locality size	Yes	Yes	Yes	Yes	Yes

Notes: This table reports estimates from the estimation of Eq. (5) using different school input variables as dependent variable. ESCS index is the PISA Economic, Social, and Cultural Status index. Father (Mother) has tertiary education is a binary variable constructed from survey responses indicating whether the father(mother) completed an ISCED level 5B, 5A, or 6 program. Locality size is a vector of five locality size indicators. All specifications include cycle fixed effects. All estimates are weighted using sample weights. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We formally assess the presence of differential trends in test scores between public and private schools during the pre-SPD period by estimating a modified version of Eq. (5), in which we interact cycle dummies with the public school indicator using data from 2006 to 2012. Table 1 presents the results. As shown in the table, relative to 2006, there are no significant differences in the growth of test scores between public and private schools in either 2009 or 2012 for both mathematics and reading. This holds across all three specifications: (i) without exogenous covariates, (ii) with controls for locality characteristics (our preferred specification), and (iii) with controls for both locality and student characteristics. These results support the validity of the identification assumption.

Table 3
DID: Impact of the SPD reform on PISA test scores.

	Math			Reading		
	(1)	(2)	(3)	(1)	(2)	(3)
Public	−45.03***	−27.23***	−12.30***	−49.99***	−28.38***	−11.33***
	(3.57)	(3.62)	(3.27)	(3.27)	(3.39)	(3.11)
After	−12.15*	−12.09*	−11.69**	2.83	2.44	3.58
	(6.50)	(6.70)	(5.47)	(6.69)	(7.02)	(6.04)
Public X After	19.21***	16.89***	15.64***	10.99	8.42	6.41
	(6.17)	(6.29)	(5.28)	(6.77)	(6.82)	(5.93)
No. of obs.	115,630			115,630		
Mean public school score 2012	408.39			418.19		
Locality size	No	Yes	Yes	No	Yes	Yes
Student characteristics	No	No	Yes	No	No	Yes

Notes: This table reports regression coefficients from the estimation of Eq. (5). Locality size is a vector of five locality size indicators. Student characteristics include a dummy variable equal to 1 if the student is female and a vector of quintile indicators for the PISA ESCS index. All specifications include cycle fixed effects. Estimates are computed using the repest command in Stata, which accounts for the full set of plausible values and produces appropriate standard errors. All estimates are weighted using sample weights. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3 presents the results from estimating equation (5). We find a positive effect of the SPD reform on student achievement in mathematics, with point estimates ranging from 18.6 points (18.6% of a SD in the full international sample in 2000) to 15.2 points, depending on the specification (16.39 in our preferred specification). These magnitudes correspond to increases of 4.5% to 3.7% of the public school mean in 2012. In contrast, we do not find a significant effect on reading test scores. The point estimates are smaller, ranging from 9.3 to 4.9 points, and none are statistically significant at conventional levels. Given the size of the standard errors, we cannot rule out a small positive effect on reading. The relatively larger effect on mathematics is consistent with the findings in Estrada (2019), which studies the effect of rule-based hires (relative to discretionary hires) on student achievement during the ACE period.

One might be concerned that the above results are driven by differential changes in inputs between public and private schools that occurred concurrently with the SPD reform. For example, if the reform induced shifts from private to public schools that improved the composition or number of students or teachers, the parallel-trends assumption would be violated. We have no evidence to believe that such changes occurred (and the results in Column 3 already indicate this was not the case). Nevertheless, we evaluate this possibility using the PISA data.

We estimate equation (5) using the following input variables on the left-hand side: students' index of economic, social, and cultural status (ESCS), maternal and paternal education, school size, and the student–teacher ratio. Table 2 presents the results. We do not find any evidence that the reform was accompanied by a relative improvement in public school inputs. In all five cases, the coefficients of interest are small in magnitude and not statistically significant, which supports the causal interpretation of the main results.

In summary, our findings indicate that the SPD reform successfully improved student achievement in the public education system. This estimate captures potential effects of the reform on the composition and performance of both new and incumbent teachers. While we lack the data and variation needed to examine these specific mechanisms – and doing so is not the objective of this paper – the evidence presented aligns with the hypothesis that rule-based hiring, and, more broadly, a more meritocratic career system, can lead to both the recruitment of higher-achieving teachers and improved student learning.

8. Conclusions

The evidence presented in this paper indicates that the implementation of the SPD reform led to the hiring of higher-achieving new teachers. This improvement was mainly driven by gains at the lower end of the prior academic achievement distribution. Two main channels likely explain this result. First, the reform decreased the prevalence of discretionary hires, which were drawn disproportionately from the bottom of the distribution. Second, the reform improved the screening efficiency of rule-based hiring, making prior academic achievement a more important determinant of hiring outcomes.

The SPD reform reduced, but did not eliminate, the use of discretionary hiring. This persistence might reflect a lack of qualified rule-based applicants willing to accept positions in less desirable localities, a relative large number of vacancies arising within the school year and simple non-compliance. Due to data limitations, we are unable to assess the relative importance of these factors.

The SPD reform was based on the promise of making merit the key criterion of personnel decisions through the teaching career, in contrast to the accusations often made to the discretionary regime pre-SPD. One could expect that such a change in personnel policy made teaching a more attractive career for higher-achieving individuals, a possibility reinforced by the fact that the reform eliminated application restrictions among university graduates from fields different from the ones exclusively related to teaching ("escuelas normales" as they are known in Mexico). We only find a modest improvement in the average ENLACE test scores of rule-based applicants. However, we cannot discount the hypothesis that the self-selection effect would have grown in importance in the medium- to long-term (because our post-reform period of analysis only comprises four years).

Regarding external validity, our analysis focuses on a subset of new teachers – recent graduates – who represent approximately 14% of all hires in the years under study. To infer the reform's effect on the full pool of new teachers, consider the following intuition. Restricting the sample to recent graduates reduces heterogeneity in the quality of new hires and applicants (as measured by test scores). Figure A.13 in the Online Appendix shows that the mean of test scores declines with age while the coefficient of variation increases, using both ENLACE and ACE data. A more heterogeneous applicant pool makes rule-based hiring more consequential: as heterogeneity increases, the role of the hiring process in sorting candidates becomes more important. As a thought experiment, if we focused on applicants with identical test scores, the hiring method would be irrelevant. Importantly, we find that discretionary hires were disproportionately drawn from the lower end of the ENLACE score distribution and that ENLACE scores became a stronger predictor of rule-based hiring under the SPD reform. Because these patterns coincide with greater variance in test scores among the full pool of new teachers than among recent graduates, it is reasonable to expect that the effects we estimate constitute a lower bound of the reform's impact on all new hires.

We also study the overall effect of the SPD reform on student achievement and find evidence of a significant increase in mathematics – of around 19% to 15% of a PISA SD – but not in reading. This positive effect could be due to both the documented improvement in the achievement profile of new teachers and, potentially, changes

in the performance of incumbent teachers. We let for future work a more comprehensive analysis of the mechanisms through which the SPD reform could have affected student achievement.

Several countries in Latin America have implemented teacher reforms to make rule-based hiring mandatory, in some cases in a context of complaints about the abuse of discretion in hiring decisions in the status quo. The results presented in this paper highlight two important dimensions in evaluating the effect of such reforms on teacher quality. One is its overall efficiency vs. the alternative regime. The other is the technical efficiency of the assessment used to screen teachers.

The results presented in this paper show how a rule-based civil service system can improve the skills profile of incoming public officials. However, the ambitious reform under analysis faced a political backlash that led to its cancellation. This demonstrates that the political implementation and the generation of a broad coalition of support among public servants and other stakeholders is as important as the technical content of personnel policies for the long-run success of civil service reforms.

Declaration of competing interest

We the undersigned declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We understand that the Corresponding Author is the sole contact for the Editorial process. He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econedurev.2026.102771.

References

Aneja, Abhay, & Xu, Guo (2024). Strengthening state capacity: Civil service reform and public sector performance during the gilded age. *American Economic Review*, *114*(8), 2352–2387.

Araujo, Maria Daniela, Heineck, Guido, & Cruz-Aguayo, Yyannú (2020). Does Test-Based Teacher Recruitment Work in the Developing World? Experimental Evidence from Ecuador. IZA Discussion Papers 13830, Institute of Labor Economics (IZA).

Bardhan, Pranab (2002). Decentralization of governance and development. *Journal of Economic Perspectives*, *16*(4), 185–205.

Besley, Timothy, Burgess, Robin, Khan, Adnan, & Xu, Guo (2022). Bureaucracy and development. *"Annual Review of Economics"*, *14*(Volume 14, 2022), 397–424.

Brassiolo, Pablo, Estrada, Ricardo, & Fajardo, Gustavo (2020). My (running) mate, the mayor: Political ties and access to public sector jobs in Ecuador. *Journal of Public Economics*, *191*(C).

Brassiolo, Pablo, Estrada, Ricardo, Fajardo, Gustavo, & Martinez-Correa, Julian (2021). Family rules: Nepotism in the mexican judiciary. CAF Working Paper #2021/09.

Brutti, Zelda, & Torres, Fabio Sánchez (2021). Turning around teacher quality in latin america: Renewed confidence and lessons from Colombia. *Economic Analysis and Policy*.

Busso, Matias, Montaño, Sebastián, Muñoz-Morales, Juan, & Pope, Nolan G. (2024). The unintended consequences of merit-based teacher selection: Evidence from a large-scale reform in Colombia. *Journal of Public Economics*, *239*, Article 105238.

CENEVAL (2008). *Manual Técnico ENLACE Media Superior*. Secretaría de Educación Pública and Centro Nacional de Evaluación para la Educación Superior (CENEVAL).

Chetty, Raj, Friedman, John N., & Rockoff, Jonah E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679.

Colonnelli, Emanuele, Prem, Mounu, & Teso, Edoardo (2020). Patronage and selection in public sector organizations. *American Economic Review*, *110*(10), 3071–3099.

Corcoran, Sean P., Evans, William N., & Schwab, Robert M. (2004). Changing labor-market opportunities for women and the quality of teachers, 1957–2000. *American Economic Review*, *94*(2), 230–235.

Crawfurd, Lee, & Pugatch, Todd (2020). Teacher Labor Markets in Developing Countries. GLO Discussion Paper Series 473, Global Labor Organization (GLO).

Dahis, Ricardo, Schiavon, Laura, & Scot, Thiago (2023). Selecting Top Bureaucrats: Admission Exams and Performance in Brazil. *The Review of Economics and Statistics*, 1–47.

de Hoyos, Rafael, Estrada, Ricardo, & Vargas, María José (2021). What do test scores really capture? Evidence from a large-scale student assessment in Mexico. *World Development*, *146*, Article 105524.

de Hoyos, Rafael, & Fernández, Marco Antonio (2019). La mal llamada reforma punitiva. *Nexos. Distancia Por Tiempos*.

Estrada, Ricardo (2019). Rules versus discretion in public service: Teacher hiring in Mexico. *Journal of Labor Economics*, *37*(2), 545–579.

Estrada, Ricardo, & Lombardi, Maria (2023). Skills and selection into teaching: Evidence from latin america. *Economía LACEA Journal*, *22*(1), 218–239.

Fredriksson, Peter, & Öckert, Björn (2008). The supply of skills to the teacher profession. *Scandinavian Journal of Economics*, *110*(2), 277–296.

Fukuyama, Francis (2013). What is governance? *Governance*, *26*(3), 347–368.

Grindle, Merilee S. (2012). *Jobs for the Boys Patronage and the State in Comparative Perspective*. Cambridge, MA: Harvard University Press.

Gronqvist, Erik, & Vlachos, Jonas (2016). One size fits all? The effects of teachers' cognitive and social abilities on student achievement. *"Labour Economics"*, *42*, 138–150.

Hanushek, Eric A., Piopiunik, Marc, & Wiederhold, Simon (2019). The value of smarter teachers: International evidence on teacher cognitive skills and student performance. *Journal of Human Resources*, *54*(4), 857–899.

Mookherjee, Dilip (2015). Political decentralization. *Annual Review of Economics*, *7*(1), 231–249.

Moreira, Diana, & Pérez, Santiago (2024). Civil service exams and organizational performance: Evidence from the pendleton act. *American Economic Journal: Applied Economics*, *16*(3), 250–291.

Munoz, Pablo, & Prem, Mounu (2021). *Managers' Productivity and Recruitment in the Public Sector: The Case of School Principals: Technical Report*.

Neilson, Christopher, Gallegos, Sebastian, & Calle, Franco (2019). Screening and recruiting talent at teacher colleges using pre-college academic achievement. Working Paper.

Nickell, Stephen, & Quintini, Glenda (2002). The consequences of the decline in public sector pay in britain: a little bit of evidence. *The Economic Journal*, *112*(477), F107–F118.

OECD (2009). PISA Data Analysis Manual: SPSS, Second Edition. In *PISA*, Paris: OECD Publishing.

OECD (2022). Education at a glance 2022. (p. 462).

Ornaghi, Arianna (2019). Civil service reforms: Evidence from US police departments. Mimeo.

Rauch, James (1995). Bureaucracy, infrastructure, and economic growth: Evidence from U.S. cities during the progressive era. *The American Economic Review*, *85*(4), 968—-79.

Rivkin, Steven G., Hanushek, Eric A., & Kain, John F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

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

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

Ujhelyi, Gergely (2014). Civil service rules and policy choices: Evidence from US state governments. *American Economic Journal: Economic Policy*, *6*(2), 338–380.