

# THE EFFECTS OF SCHOOL ACCOUNTABILITY ON TEACHERS IN PUBLIC AND PRIVATE SCHOOLS: EVIDENCE FROM CHILE

## **Gregory Elacqua**

(corresponding author)  
Division of Education  
Social Sector  
Inter-American Development  
Bank  
Washington, DC 20577  
gregorye@iadb.org

## **Diana Hincapié**

Independent consultant  
Accra Ghana  
dianah@gwu.edu

## **Matías Martínez**

School of Education and Social  
Policy  
Northwestern University  
Evanston, IL 60208  
matias@u.northwestern.edu

## **Abstract**

This paper estimates the effects of a school accountability policy on year-to-year teacher mobility in publicly and privately managed low-performing schools in Chile. As school ranking depends on the institution's relative position according to a set of variables and their corresponding thresholds, we use a multivariate regression discontinuity design to evaluate the impact of the policy on teacher mobility. Our findings reveal that teachers are generally more likely to leave low-performing schools, although this result varies by type of management. Teachers in public schools tend to transfer to other schools within the system, and those who do so are more likely to be working in two or more schools. Meanwhile, teachers in private schools are comparatively more likely to exit the system altogether, with mobility concentrated among low-productivity teachers. Despite these differences, the introduction of accountability did not induce new hires at either type of school.

[https://doi.org/10.1162/edfp\\_a\\_00416](https://doi.org/10.1162/edfp_a_00416)

© 2023 Association for Education Finance and Policy

## 1. INTRODUCTION

The adoption of performance-based accountability mechanisms as a means of improving school quality has become widespread among policy makers around the globe (Spillane 2021). A key assumption for the success of this type of policy is that school communities will respond to incentives by increasing effort, designing high-quality teaching practices, and boosting management efficiency through the redefinition of work and redistribution of resources (Mbiti 2016; Espeland and Sauder 2007). Teachers are the most influential resource when it comes to student learning, leading many scholars to explore whether school accountability makes it more difficult for low-performing schools to retain and attract high-quality teachers (Clotfelter et al. 2004; Feng, Figlio, and Sass 2018). The primary concern is that accountability may lead to higher teacher turnover, thus disrupting social relations, weakening staff and community cohesion, and adversely effecting student learning (Johnson, Berg, and Donaldson 2005; Ronfeldt, Loeb, and Wyckoff 2013).

Conceptually, the effects of accountability on teachers are not clear. On one hand, these mechanisms can create incentives for effective teachers to remain in their school. If teachers receive more resources and support from the principal and other authorities, this may lead to cultural and institutional changes that make the low-performing school an attractive place to work (Boyd et al. 2011; Dizon-Ross 2020). On the other hand, this policy approach can discourage teachers from continuing to work in low-performing schools if accompanied by public labels suggesting incompetence, increases in administrative workload, reduced autonomy in teaching decisions, and/or shifts from effective pedagogical practices to student testing training (Rouse et al. 2013; Gjefsen and Gunnes 2020).

The most cited empirical evidence on this matter is from high-income countries and suggests that accountability causes detrimental teacher turnover. Feng, Figlio, and Sass (2018) show that higher value-added teachers in Florida are more likely to leave schools that have received a failing grade by the accountability system. Clotfelter et al. (2004) find that the introduction of accountability in North Carolina increased the share of teachers with less experience and fewer credentials. In Oslo, Gjefsen and Gunnes (2020) report that after school value-added information was made public, the likelihood of teachers leaving the system increased, especially those of high ability. In England, Sims (2016) observes that teacher turnover rose in schools whose performance category was downwardly reclassified by the inspection system from “requires improvement” to “inadequate.” Finally, Dizon-Ross (2020) reports that the accountability system in New York decreased teacher mobility at low-performing schools and increased the overall teacher quality measured by teacher value-added.

We contribute to this body of empirical literature by studying Chile’s school accountability policy and examining its heterogeneous effects on publicly and privately managed schools. The Chilean school system provides universal per-pupil vouchers such that parents can choose among public and private options, with around 55 percent of families opting for the latter (Carrasco and Gunter 2019). In 2012, a performance-based accountability policy was implemented, under which schools began to be ranked annually into different categories. The results of this school classification are publicly available, which means that those schools in the lowest category not only face the potential flight of the families, but also the threat of closure if they do not improve within

four years. We focus here on teachers working in low-performing schools during the first two years since the introduction of accountability, using a multivariate regression discontinuity design to evaluate whether they are more likely to leave their school or to leave the school system entirely. In addition, we estimate the potential effect for different teachers grouped according to proxies of their teaching ability, as well as track the moves made by teachers who remained in the school system. Our results indicate that the Chilean school accountability policy increased overall teacher turnover, though this took different forms according to school type. Teachers in public schools appear more likely to move to another school, and this mobility was concentrated among teachers working in two or more schools. Teachers in private schools were instead relatively more likely to leave the system altogether, with mobility concentrated among those teachers with less experience, lower scores on their college admission exam, and hired through temporary contracts.

The remainder of the paper is organized as follows. In the next section, we describe the school system in Chile, the origins of the accountability mechanism, and its main features. In section 3, we introduce the data employed in our analysis, the variables used to classify schools in the low-performing category, and our measure of teacher mobility. We then present our methodology and test its validity. In section 4, we report the main results, both for all teachers and different groups of teachers, as well as discuss other outcomes related to teacher mobility. Finally, in section 5, we present our conclusions and policy considerations.

## 2. SCHOOL ACCOUNTABILITY AND THE TEACHER LABOR MARKET

### School Accountability

The foundation of the present Chilean school system was laid in the early 1980s. One of the original design's main features was the provision of public funding based on a per-capita scheme consisting of a flat subsidy per student enrolled (and tied to attendance) and the possibility of choosing among three types of schools: (i) public schools, financed by government subsidies and administered by the local municipal government; (ii) private voucher schools, also funded by government subsidies, but managed by a private religious or secular (for-profit or nonprofit) organization; and (iii) private schools, both financed and administered privately. A central argument to support this arrangement was that parents would “vote with their feet” and prefer higher-quality schools, forcing poor-performing schools to either improve or go out of business (Friedman 1955).

Though it was hoped that the system's design would prevent the persistence of low-performing schools, studies conducted in the 2000s suggested that there had instead occurred an increase in the sorting of students by ability, resulting in schools with high concentrations of low academic achievement pupils (e.g., Hsieh and Urquiola 2006). This evidence, combined with pressure from secondary students demanding structural changes to the system, led to significant modifications in the public funding of schools (O'Malley and Nelson 2013; Bellei and Vanni 2015). In 2008, the *Subvención Escolar Preferencial* (SEP) [Preferential School Subsidy] law was enacted. This introduced, for the first time, differentiated funding favoring students from disadvantaged backgrounds (based on socioeconomic status [SES]). The SEP law assumes that low-SES students have higher educational needs that require extra funding, and transfers an additional per capita subsidy (50 percent higher than the base voucher) to schools serving students

classified as vulnerable. This extra funding was limited to public or private voucher schools that voluntarily agreed to participate in the program.

Along with the increase in funding, the SEP law introduced explicit school accountability measures. For schools to receive the additional subsidy, they were required to comply with several conditions, including signing an agreement in which they committed to developing and implementing an improvement plan over the course of the following four years. This agreement could be renewed if schools spent at least 70 percent of the SEP additional resources, and all the expenditures had been properly reported to the relevant authorities. They were also required to comply with minimum standards of quality, mainly defined by fourth-grade student performance on standardized tests in math, sciences, and language over the previous three years.

The quality standards were defined by a performance classification that distinguished between three categories of schools: (i) *autonomous* or high-performing schools; (ii) *emerging* or average-performing schools; and (iii) *in-recovery*, or low-performing schools, that did not meet the minimum national standards. Although these three categories were defined under the SEP law in 2008, it was not until 2012 that schools actually started to be classified as *in-recovery*. From 2008 to 2011, low-performing schools were instead ranked as *emerging*.

Accountability pressures to use resources efficiently were higher for *in-recovery* schools. Indeed, under the SEP law, if these schools failed to improve their performance and move up to the *emerging* category within three years, the Ministry of Education would inform the school community and encourage families to consider other schooling options, as well as facilitate transportation to another school. If the school remained in the *in-recovery* category for four years, the ministry could revoke the school's license to operate and cut its public funding. The SEP law furthermore established that information on school performance had to be made public, intended as a means of influencing parental preferences. Being classified as a low-performing school could hurt future enrollments, thus affecting the level of resources the school received through the per capita funding formula, and eventually, its ability to operate.

Qualitative evidence suggests that these pressures could affect teachers' decisions to continue working at *in-recovery* schools. An ethnographic study conducted by Assaél et al. (2014) consisting of observations, visits, and the analysis of interpersonal dynamics at two schools classified as *in-recovery* in 2012 revealed an already existent decreasing trend in student enrollment since 1995, which, combined with the threat of closure, led some school managers and teachers to see their school as "a dying patient at an intensive care unit." The study also reports tensions among some teachers related to the workload distribution and a perceived lack of opportunities to participate in school policy decisions.

It is important to note that this pressure exerted on *in-recovery* schools only lasted a few years. In addition to a sharp decline in the number of such establishments (from 196 schools in 2012 to 59 in 2014), the central authority eventually chose to implement a different accountability policy, using a more comprehensive conceptualization of school quality to rank institutions. This new policy became effective in 2016, accompanied by new timeframes for improvement, and a discounting of the sanctions previously established for *in-recovery* schools. In fact, no school classified for four consecutive years as *in-recovery* between 2012 and 2015 stopped receiving public funding. As the

accountability pressure was highest under the original policy, we focus our analysis on the 2012 and 2013 school years.

### The Teacher Labor Market in Chile

The teacher labor market in Chile can be conceptualized as the interaction of two components: (i) teacher supply and (ii) school demand. On the supply side, teacher candidates are free to apply to whichever school they are interested in, and do not face legal constraints should they wish to leave their current school. In general, teachers prefer working at schools that offer better working conditions, namely, the provision of pecuniary and nonpecuniary benefits (Boyd et al. 2005; Simon and Johnson 2015; Hanushek, Kain, and Rivkin 2004). However, informal barriers, such as applicant familiarity with school options and their social context, may limit the set of schools to which they ultimately decide to apply (Freedman and Appleman 2009; Lin and Dumin 1986). In particular, applicants' social network has been shown to be relevant in the Chilean context. Interview data from 2012 indicates that new teachers in Santiago heavily relied on their social networks to identify the schools they would apply to (Paredes et al. 2013; Cabezas et al. 2017).

Among more veteran teachers applying to schools, previous experience is informative for identifying a school that better fits their preferences (Grissom, Viano, and Selin 2016). Regarding teachers working at schools serving low-income students, studies show that teachers have both incentives to stay and to leave these schools. While they report great personal satisfaction from having the opportunity to teach the most disadvantaged children, they also express a high level of frustration relative to a perceived lack of support from the families of these pupils (Paredes et al. 2013).

On the demand side, school administrators seeking to hire teachers are governed by different legislation depending on the type of school. Public schools (55 percent of participant schools in SEP in 2012) are regulated by the *estatuto docente* [teacher statute], which defines the valid reasons for hiring new teachers (e.g., changes in student enrollment), the requirements for teacher applicants (e.g., holding a teaching degree), the deadlines for registering new hires (typically, 15 November) and completing public teacher competitive examinations (usually on 15 December, coinciding with the end of the school year), as well as the rules for terminating teaching contracts. In 2011, new legislation increased flexibility for terminating contracts, allowing public school administrators to dismiss up to 5 percent of teachers with permanent contracts, provided that they had a low score on the national teacher assessment.

Administrators at private voucher schools (45 percent of participant schools in SEP 2012) are not required to comply with the teacher statute definitions, and instead follow the same labor legislation that governs employer–employee relationships in most sectors of the economy. Thus, it is up to private voucher school administrations to define the teaching positions to be filled, their deadlines, and the hiring process. Contract termination is also stipulated in general legislation that provides employers with more discretion compared to the public sector, allowing them to dismiss teachers based on reasons related to company needs, business modernization, low productivity, or changes in market conditions (Flores, Ortúzar, and Milesi 2014; Ortúzar et al. 2016).

In practice, the process of applying and finding a teaching position can take months for teachers at both public and private voucher schools. For instance, based on 2012

data on teachers working at their first job, Cabezas et al. (2017) report that it took an average of 2.7 months to be hired (from the moment they sent their first resume until they received a formal offer). The authors also show that although teachers applied to a multitude of vacancies, they generally received a relatively low number of job offers. On average, teachers applied to ten schools and received one job offer for every seven applications.

### Differential Effects of Accountability by School Management Type

There are at least three reasons to believe that there could be differential mobility responses to accountability in the public and private school sectors. First, as mentioned above, the institutional rules are different for public and private voucher school teachers. Private schools have more freedom to hire and dismiss teachers than public schools. Moreover, there is a strong teachers' union that supports public school teachers in improving working conditions by negotiating better salaries, benefits, and work stability.

Second, in the public sector seniority is rewarded through a policy known as *bienios*, which is a salary increase every two years based solely on years of experience in public schools. This implies that the longer a teacher has taught at the municipality, the fewer incentives he or she has to change schools or leave the profession. Since the *bienios* incentive is not included in the private sector compensation scheme, public school teachers have stronger incentives to continue working in the public sector, even if it means transferring to another school managed by the same municipality.

Third, the government introduced early retirement packages for public school teachers in 2011 and 2012. Private voucher school teachers were not eligible to receive these voluntary bonuses. Law 20.501 (*Calidad y Equidad de la Educación*), enacted in 2011, included a voluntary retirement bonus for public school teachers in 2012 and 2013, and a bonus that paid teachers who retired in or before December 2010 a one-time lump sum between 1,000,000 and 2,000,000 pesos (US\$1,800 and US\$3,600) if they had worked for more than 10 years in a public school. Both incentives provided benefits for more experienced teachers in the public sector. Thus, many career public school teachers had strong incentives to remain in the same municipality to be eligible for these bonuses. Less experienced public school teachers may have also been more likely to continue to teach in the public sector if they had anticipated similar benefits would be granted in the future. For instance, Law 20.883 (*Bonificación por Retiro Voluntario*), enacted in 2015, reauthorized the bonus for voluntary teacher retirement in public schools, benefiting more cohorts of teachers.

## 3. DATA AND METHODS

### Data

To identify the effect of school accountability on teacher mobility, we analyze year-to-year changes in teacher employment. Our primary dependent variable is a binary variable indicating whether a teacher in year  $t$  leaves his or her school within one year. We are also interested in examining whether teachers leave their school to go work in another school or outright leave the school system. Our key independent variable is a continuous score that perfectly determines the SEP school category in a setting that allows the implementation of a regression discontinuity design. Specifically, when this

score is less than zero, the school is ranked as *in-recovery*, and when it is equal to or greater than zero, it is *emerging*, given that we exclude the *autonomous* schools from our analysis.

To perform our planned analyses, we combine five sets of administrative data: teacher censuses, SEP school classification databases, national teacher assessments, college admission exam results, and scores on the national education quality evaluations (administered under the Education Quality Measurement System [SIMCE]).<sup>1</sup> The primary database is at the school level and includes school performance rankings for the approximately 9,000 schools participating in SEP each year in 2012 and 2013. This database also includes the variables used for the definition of school performance.

Under the 2008 law, the SEP performance classification of schools follows three steps. First, considering the last three SIMCE national evaluations, schools that did not participate in at least two out of the last three evaluations or where fewer than twenty of their fourth-grade students participated are automatically assigned to the *emerging* category—though it is noted that they are not classifiable by their SIMCE performance. Second, the remaining schools fall into the *in-recovery* category if they simultaneously comply with two conditions for two out of the last three evaluations: (i) the school's SIMCE average is below 220 points for all the subjects assessed in fourth grade,<sup>2</sup> and (ii) fewer than 20 percent of its students scored higher than 250 points in the average of all subjects tested. Finally, independently of the criteria above, schools are also classified as *in-recovery* if they score below the tenth percentile in the distribution of a school quality index. This index is computed annually for the purpose of the SEP ranking and is based on SIMCE results (70 percent of the index) and other quality indicators such as student passing and retention rates, parental participation at school, pedagogical innovations, adequate working conditions, and teacher evaluation (30 percent of the index).

Our analysis focuses on the approximately 3,500 schools that each year were deemed as classifiable by their SIMCE performance. In 2012, 196 schools were ranked as *in-recovery*, 1,983 as *emerging*, and 1,384 as *autonomous*. The corresponding numbers for 2013 were 183, 1,931, and 1,370, respectively. As stated above, we exclude the *autonomous* schools from this analysis.

We merge the SEP school-level database with our primary database at the teacher level: teacher census data. The teacher census is reported annually to the central government by principals in public and private voucher schools. Each teacher has a numerical identifier that allows to trace their work trajectory. Here, we follow teachers' movements between  $t$  and  $t + 1$  between 2012 and 2014. Our teacher-level database also includes information about the number of schools at which a teacher is employed, the type of contract, and the number of hours a week he or she works at a given school. (See Appendix 1, available in a separate online appendix that can be accessed on *Education Finance and Policy's* website at [https://doi.org/10.1162/edfp\\_a\\_00416](https://doi.org/10.1162/edfp_a_00416), for details on how we build the database.)

1. Each year, the Education Quality Measurement System (*Sistema de Medición de la Calidad de la Educación*; SIMCE) tests all students in the second, fourth, sixth, eighth, and/or tenth grades in math, language, and sciences. The SIMCE also gathers detailed information on teachers, students, and parents.
2. The student scores for each subject tested in SIMCE follows a normal distribution with a mean of 250 and standard deviation of 50 for the first year the grade-subject is assessed. Math, language, and social sciences were first tested in 1999, and science was first tested in 2007.

**Table 1.** Teacher Characteristics by SEP Performance Classification and Year

	2012		2013		Total	
	In-recovery (1)	Emerging (2)	In-recovery (3)	Emerging (4)	In-recovery (5)	Emerging (6)
Teaching positions	4,270	50,977	3,965	52,658	8,235	103,635
Public schools	2,929	25,257	2,614	24,954	5,543	50,211
Private voucher schools	1,341	25,720	1,351	27,704	2,692	53,424
Teacher mobility (t and t + 1)						
Stays in their school	0.727	0.763	0.723	0.784	0.725	0.773
Moves to another school	0.166	0.140	0.183	0.136	0.174	0.138
Leaves the school system	0.107	0.098	0.094	0.080	0.101	0.089
No. of schools at which a teacher works						
One school	0.804	0.844	0.829	0.855	0.816	0.850
Two or more schools	0.196	0.156	0.171	0.145	0.184	0.150
Female	0.706	0.756	0.721	0.752	0.713	0.754
Working experience	14.591	13.934	13.245	12.841	13.943	13.379
Permanent contract	0.527	0.580	0.458	0.521	0.494	0.550
Contract time	31.931	32.588	32.536	32.928	32.222	32.761
College admission test - PSU	48.166	53.156	48.530	52.643	48.382	52.852
Teacher assessment						
High	0.047	0.079	0.046	0.076	0.046	0.078
Medium-high	0.564	0.627	0.571	0.632	0.567	0.629
Medium-low	0.379	0.288	0.374	0.287	0.377	0.287
Low	0.010	0.006	0.010	0.005	0.010	0.006

Notes: PSU stands for University Selection Test (*Prueba de Selección Universitaria*), a standardized college admission exam offered each year in Chile. The score corresponds to the average of the math and language tests. The percentiles for these two variables were computed relative to the cohort of students taking the exam the same year as the teacher in the dataset. All available data were included (scores from 2004 to 2011). If a teacher took the college entry exam more than once, the most recent score was considered. The variables related to the teacher assessment correspond to the official grade given by authorities ranking teachers in the public sector. Scores from 2008 to 2011 were included. If a teacher was assessed more than once, the most recent score was considered. SEP = *Subvención Escolar Preferencial*.

To describe further differences between *in-recovery* and *emerging* schools, we also complement our primary databases with additional databases at the school and teacher levels. Table 1 presents descriptive statistics at the teacher level and table 2 at the school level. On average, teachers at *in-recovery* schools are more likely to both move to another school and leave the school system entirely at the end of the school year, as well as to work in more than one school and be hired with temporary contracts. They also have been working slightly longer and performed less well on both the national teacher assessment and on college admission exams for teaching programs. (For the interested reader, in online Appendix 2 we present descriptive characteristics for teachers across sectors for pretreatment years.) School-level data meanwhile shows that *in-recovery* schools not only have lower results on SIMCE than do *emerging* schools but also tend to enroll fewer students, and serve pupils from less favorable socioeconomic backgrounds, as proxied by parental lower educational attainment and lower family income.

### Multivariate Regression Discontinuity Design

To estimate the causal effect of working at *in-recovery* schools on teacher mobility, we exploit the fact that the methodology used to classify schools in Chile is based on the



**Table 2.** School Characteristics by SEP Performance Classification and Year

	2012		2013		Total	
	In-recovery	Emerging	In-recovery	Emerging	In-recovery	Emerging
	(1)	(2)	(3)	(4)	(5)	(6)
Total schools	196	1,983	183	1,931	379	3,914
Public schools	119	1,000	109	940	228	1,940
Private voucher schools	77	983	74	991	151	1,974
Student Enrollment	349	505	343	509	346	507
School vulnerability	83	68	83	69	83	68
Family monthly income (US\$)	644	966	690	1,050	667	1,008
Parents' schooling	10.497	12.024	10.570	12.090	10.532	12.057
Student test scores (SD)						
Language	-0.590	-0.198	-0.666	-0.237	-0.627	-0.217
Math	-0.666	-0.244	-0.646	-0.199	-0.656	-0.222

Notes: Different sources of data were used for this table. Information on enrollment comes from the official registries of the Ministry of Education. School vulnerability data—an index ranging from 1 to 100 with higher values indicating higher social vulnerability (e.g., higher poverty)—was retrieved from the National Board of School Aid and Scholarships (*Junta Nacional de Auxilio Escolar y Becas*; JUNAEB). The remaining information was obtained from questionnaires and evaluations associated with the Chilean national assessment system of school quality. Test scores are expressed on standard deviation units (SD). SEP = *Subvención Escolar Preferencial*.

schools' position relative to a set of variables and their corresponding thresholds (see the previous section). We use a generalization of the traditional regression discontinuity design (RDD) in the case where multiple assignment variables and cutoffs are used for treatment assignment.

An RDD with multiple assignment variables (multivariate regression discontinuity design) poses challenges that are different from those in a traditional design. Indeed, the analytical procedures for estimating treatment effects are more complex than those for estimating a treatment effect at a single point along with a unique assignment variable. We base our method on a series of studies that have addressed these challenges (Papay, Willett, and Murnane 2011; Reardon and Robinson 2012; Wong et al. 2013; Porter et al. 2017).

Reardon and Robinson (2012) propose and discuss the merits of different estimation methods for multivariate regression discontinuity design. The authors present three main strategies, but do not provide evidence as to which is preferred. The first method (frontier) subsets the data and estimates separately the local treatment effect on each of the frontiers defined by each of the assignment variables. The second approach (fuzzy instrumental variables or fuzzy IV) uses all observations in a single estimation by centering on one of the frontiers and using instrumental variables to obtain estimates of the local complier average treatment effect. The third way (surface) of estimating the local treatment effects involves modeling a relationship between the assignment variables in a sole specification that uses all observations.

Porter et al. (2017) extend the work of Reardon and Robinson (2012) by comparing the methods in terms of their estimators' bias and precision. Based on simulations, the authors look at the performance of the three methods: frontier, fuzzy IV, and a specific variant of the surface method known as the binding score, which involves collapsing all assignment variables into one that perfectly defines treatment assignment. They show

that although all three strategies can achieve unbiased estimators, the fuzzy IV is more prone to bias than the other two methods. With regard to precision, the binding score is more robust than the frontier method, especially when the correlation between the assignment variables is high.

While the binding score is the method that maximized precision, Porter et al. (2017) suggest using this method when the local treatment effects are expected to be homogeneous across frontiers. In the case of significant differences in the estimated treatment effect across frontiers, the binding score provides estimators that are harder to interpret as it is based on a weighted average that might depend on the distribution of the assignment variables.

Based on theory and simulated data, empirical studies working with a large number of observations should implement both the frontier method and the binding score. In our case, however, the number of observations is not large enough to allow us to estimate separate effects for each frontier defined by each of the assignment variables. We thus opted to use, primarily for two reasons, the binding score strategy. First, most schools are ranked by their value on only one assignment variable. Overall, more than 75 percent of schools are closest to the frontier defined by the tenth percentile of the school quality index, which makes other frontiers less relevant for assignment to treatment. Second, there is a high correlation between the seven assignment variables, with average pairwise correlations of over 0.8 each year. Conceptually, a high correlation among the assignment variables indicates that the binding score method is more appropriate than the frontier approach and that their estimators are close to the traditional RDD (Porter et al. 2017).

The binding score strategy allows us to collapse scores from multiple assignment rules into a single assignment variable, achieving a sharp RDD that uses all the observations simultaneously in a single estimation of the local treatment effects. This approach simplifies the analyses, allows for an estimation based on a sharp RDD, and provides efficient estimators. That said, a disadvantage of this method is that the pooling of units from different frontiers may increase the heterogeneity of the outcome at the pooled cutoff, requiring a larger bandwidth for nonparametric estimates (Wong, Steiner, and Cook 2013). Despite this potential limitation, other studies have similarly chosen the binding score strategy to obtain unbiased estimations (e.g., Robinson 2011; Reardon et al. 2010; Gill et al. 2009).

### Binding Score Regression Discontinuity

This method is based on the construction of a new assignment variable  $Z$  (*binding score*) that combines all the assignment variables to perfectly determine the treatment status. For example, suppose that assignment to treatment for unit  $i$  depends on two variables ( $R_i$  and  $M_i$ ). Units are assigned to a single treatment condition  $T$  if they score below both cutoffs ( $r_c$  and  $m_c$ ) and to the control condition  $C$  if they score above either cutoff ( $R_i \geq r_c$  or  $M_i \geq m_c$ ). Neither of these variables individually defines treatment allocation, but we can construct a new variable  $Z_i$ , defined as the maximum between both assignment variables centered at its respective cutoff:

$$Z_i = \max(R_i^c, M_i^c), \quad (1)$$

where  $R_i^c = R_i - r_c$  and  $M_i^c = M_i - m_c$ , and by construction,  $T_i = 1$  if  $Z_i < 0$  and  $T_i = 0$  if  $Z_i \geq 0$ . Thus, the problem becomes a sharp RDD, and standard analytical methods can be used by defining  $Z$  as the assignment variable and zero as the cutoff. Although the original assignment variables are transformed, Wong et al. (2013) and Porter et al. (2017) show that this method estimates the same causal effect as alternative methods, and estimators are more efficient in the case where there is a high correlation between assignment variables.

For the Chilean setting, we use the SEP ranking database to construct a binding score for both 2012 and 2013. First, we consider only those schools classified by their performance indicators, thus excluding establishments that either did not participate in at least two out of the last three national assessments, or with an average of less than twenty students taking each test. As described above, the schools that do not meet these two criteria are by default—and not by actual education outcomes—classified as *emerging*. Second, we combine the different assignment variables according to the criteria established in the classification methodology. Specifically, we standardize all the assignment variables so that their cutoffs are zero and then combined them according to the defined joint conditions.

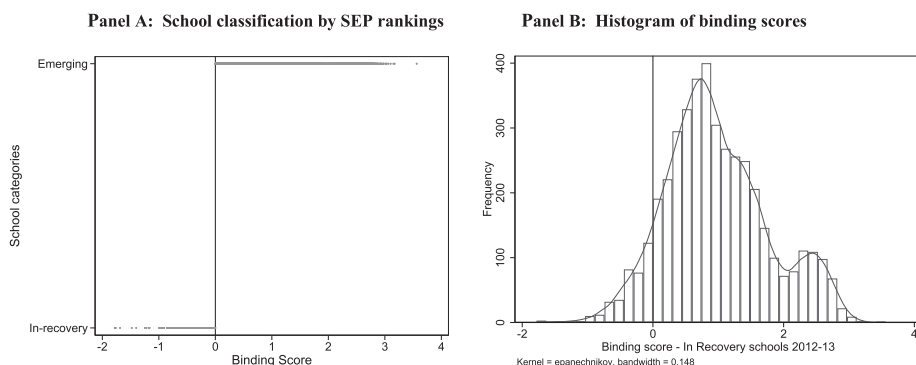
The process of computing the binding score for each year starts by considering that for a school to be classified as *in-recovery*, it must have less than 20 percent of its fourth-grade students scoring higher than 250 points in the average of all subjects tested by SIMCE, and an average lower than 220 points on that same variable. These two conditions translate into the computation of a variable equivalent to the maximum between these two standardized assignment variables. We calculated this variable separately for each of the last three years with available data. Since this condition must be met for two out of the previous three years, the second maximum score between these three variables represents a preliminary binding score that includes all six of the conditions (two conditions each year).

To obtain the final binding score for each year that schools are SEP-classified, the preliminary binding score above is combined with one additional condition related to a school quality index: All schools below the tenth percentile in the distribution of this index are classified as *in-recovery*. Thus, the final binding score corresponds to the minimum between both the preliminary binding score and the school quality index centered at zero. The density of schools near this final binding score each year is presented in figure 1. Most schools are located within  $\pm 1$  unit of the binding score from the cutoff separating *in-recovery* and *emerging* schools.

Using the final binding score, we can use a sharp RDD to estimate the effect of the program on teacher mobility. The credibility of this design depends critically on the inability of schools to manipulate the assignment variables, and thus to influence their classification.

### Regression Discontinuity Diagnostics

As indicated, a fundamental assumption of the regression discontinuity analysis is that schools cannot manipulate the assignment variables; thus, falling on either side of a threshold can be considered random. Several features of the SEP classification and data collection processes make it unlikely that schools manipulated these variables.



Notes: Panel A shows the accountability category received by a school as a function of the underlying binding score. Panel B displays a frequency histogram and density plot to inspect its smoothness and mass around the threshold. SEP = *Subvención Escolar Preferencial*.

Figure 1. Density of Schools Near the Binding Score Cutoff for 2012 and 2013 Pooled Data

While the original 2008 SEP law did spell out the cutoffs dictating classification as *in-recovery*, this was explicated only for the SIMCE variables and not for the school quality index. Furthermore, schools were neither informed relative to the methodology used to combine the assignment variables, nor might they have expected to be classified as *in-recovery*, given that no school received that ranking between 2008 and 2011, regardless of their performance. Finally, the lagged timing of the classification would also likely jeopardize a school's ability to manipulate its ranking. The school's classification is based on the past three assessments with a two-year lag. For instance, the 2012 classification uses SIMCE scores from 2008, 2009, and 2010. Thus, for a school to be able to manipulate the classification would require planning from two to four years in advance, when the *in-recovery* category was not even operational, which seems highly unlikely.

Additionally, the Ministry of Education closely monitors the SIMCE data collection process both during and after testing. It hires external staff to prevent teachers and principals from accessing the tests or the classroom where students are tested. Furthermore, once available, the test data are analyzed to ensure that the results reliably represent student ability. If analysts at the Ministry of Education determine that is not the case, then the SIMCE results are disregarded and not reported to the public.

One data-driven approach to determine whether there is evidence of manipulation consists of using a formal test that attempts to identify discontinuities in the density of the assignment variables around the threshold that defines the treatment status. McCrary (2008) was the first to develop such a test, consisting of a two-step procedure. In the first step, the data are binned to construct a histogram of the assignment variable, a procedure that introduces tuning parameters for the bins. In the second stage, this histogram is “smoothed” by estimating a local linear regression separately on both sides of the threshold. The test is implemented as a Wald test whose null hypothesis is that the discontinuity is zero. Under the null hypothesis of continuity, the test distribution is very close to a normal distribution.

More recently, Cattaneo et al. (2020) propose an updated version of McCrary's strategy, which is entirely data-driven and differs in avoiding transforming the data or choosing multiple tuning parameters. Specifically, the authors construct a local-polynomial

**Table 3.** Density Test Implemented for All Assignment Variables and Binding Scores ( $p$ -values)

	2012	2013
Binding score	0.313	0.404
Assignment variables		
Fourth gr. SIMCE < 220, $t - 4$	0.580	0.188
Fourth gr. SIMCE < 220, $t - 3$	0.567	0.901
Fourth gr. SIMCE < 220, $t - 2$	0.655	0.546
Prop. 4th gr. SIMCE < 0.2, $t - 4$	0.329	0.375
Prop. 4th gr. SIMCE < 0.2, $t - 3$	0.642	0.996
Prop. 4th gr. SIMCE < 0.2, $t - 2$	0.402	0.371
School Quality Index < p10	0.322	0.473

Notes: Own calculations based on SEP (*Subvención Escolar Preferencial*) classification data.

density estimator based on kernel functions allowing to test for differences at each side of the cutoff. Based on this method, table 3 presents the  $p$ -values associated with all assignment variables and binding scores for 2012 and 2013. The results suggest that there is no evidence that schools manipulated any of the assignment variables.

### Empirical Strategy

We are interested in estimating the effect that working at an *in-recovery* school has on the probability of teachers leaving their school. To this end, we define a set of binary variables as outcomes that are equal to one when teachers leave the school, either to move to another school or to exit the school system entirely. We then pool the 2012 and 2013 data and follow a nonparametric approach that fits a weighted linear regression for teacher observations in an interval around the cutoff defined by a bandwidth  $h$ . The local specification within  $h$  has the following general form assuming a polynomial of order one to model the relationship between the running variable and the dependent variable:

$$Y_{i,s,t+1} = \alpha + \tau \cdot IR_{s,t} + \beta_1 \cdot (Z_{s,t} - c) + \beta_2 \cdot IR_{s,t} \cdot (Z_{s,t} - c) + X_{s,t-1} + \varepsilon_{i,s,t}; \quad (2)$$

where  $c - h \leq Z_{s,t} \leq c + h$ .

Teacher mobility for teacher  $i$  in school  $s$  is represented by  $Y_{i,s,t+1}$ , which examines whether the teacher stays or leaves the school in the year immediately following the school classification.  $IR_{s,t}$  is a dummy variable that takes the value of one when the schools are classified as *in-recovery* in year  $t$  and zero otherwise. The expression  $(Z_{s,t} - c)$  represents the distance to the threshold of the assignment variable computed using the binding score method with  $c$  as the cutoff. Note that the treatment condition is defined by  $IR_{s,t} = 0$  if  $Z_{s,t} \geq c$  and  $IR_{s,t} = 1$  if  $Z_{s,t} < c$ . Our binding scores are computed so that the cutoff that defines treatment status is zero (i.e.,  $c = 0$  for all binding scores defines the cutoff that divides *in-recovery* and *emerging* schools). Finally,  $X_{s,t-1}$  is a vector with a set of school-level covariates measured one year before the implementation of the accountability policy, including averages for the SIMCE scores for math and reading, the maximum value between the years of schooling of mothers and fathers, a school vulnerability index based on family socioeconomic status and risk of student drop out, the proportion of teachers leaving the school at baseline, and student

enrollment. The impact that working at a school classified as *in-recovery* has on teacher mobility is captured by the coefficient  $\tau$ .

We follow this strategy to estimate the impact that working at an *in-recovery* school has on the three binary variables that measure teacher mobility between  $t$  and  $t + 1$ : (i) equal to one when the teacher moves to another school or leaves the system, and zero otherwise; (ii) equal to one when the teacher moves to another school and zero otherwise; and (iii) equal to one when the teacher leaves the school system and zero otherwise. Our estimations and inference follow the work of Calonico et al. (2017) and Calonico, Cattaneo, and Titiunik (2014), who propose an estimation method that computes an optimal bandwidth based on the minimization of the mean square error (MSE), a procedure that optimizes the bias-variance trade-off associated with the choice of bandwidth. In terms of inference, we implement an approach known as robust bias correction, which removes potential biases specific to hypothesis testing that can occur when using MSE-optimal bandwidth. This is achieved by recentering the point estimate and rescaling the standard errors in comparison to a conventional ordinary least squares-based  $t$ -statistic. All estimates are based on a triangular kernel for weighting the observations and standard errors clustered at the school level.

A key decision under the nonparametric approach above is the bandwidth size, which defines the weight assigned to each observation. As the bandwidth gets smaller, the observations close to the cutoff receive more weight in the estimation. We present as our main specification the results associated with the optimal bandwidth proposed by Calonico et al. (2017), based on the approach known as MSE-optimal bandwidth for the regression discontinuity treatment effect estimator. To test the robustness of our results, we estimate specifications with various bandwidths, using three factors that expand the optimal bandwidth to both sides (1.75, 1.50, 1.25), and three factors that contract it at both sides (0.75, 0.50, 0.25).

To test the validity of the results, we also perform a placebo test replicating the estimation strategy for teacher mobility measured one year before the implementation of the accountability policy. Specifically, we estimate the effect of the binding scores for 2012 and 2013, respectively, on teacher mobility in 2010 and 2011. As for the main sample, we locally estimate the placebo effects using a triangular kernel, clustering the standard errors at the school level, and including the same set of covariates at baseline (i.e., 2009 and 2010).

In addition, to further assess the validity of our identification strategy, we test for discontinuities in the baseline covariates and in the probability of school closing by using the same above-described nonparametric approach. Twenty-six school-year observations correspond to schools that closed in the following year within a bandwidth of  $\pm 0.5$  binding-score units (seven *emerging* and five *in-recovery* schools). We exclude these from the analysis, but only after testing for discontinuities associated with the closing decision. The results are presented in online Appendix 3 and show that our approach is valid given that we do not find discontinuities in pretreatment covariates or in the probability of school closing.

#### 4. RESULTS

In what follows, we present three sets of results. First, the outcomes of our main analysis on the overall effects of accountability and the extent to which teacher mobility

**Table 4.** Regression Discontinuity Estimates of the Effect of Working at an *In-recovery* School on Teacher Mobility

	All Schools			Public Schools			Private Voucher Schools		
	Departees	Schools Movers	System Leavers	Departees	Schools Movers	System Leavers	Departees	Schools Movers	System Leavers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<b>Panel A: Main Results, Years 2012 and 2013</b>									
ADP	0.252	0.156	0.097	0.256	0.142	0.105	0.241	0.161	0.085
LATE	0.044** (2.029)	0.048*** (2.931)	0.003 (0.355)	0.037 (1.259)	0.068*** (2.703)	-0.026 (1.548)	0.080* (1.829)	0.028 (0.938)	0.055* (1.661)
Bandwidth	0.308	0.279	0.435	0.191	0.153	0.230	0.270	0.297	0.276
Obs. (left-right from the threshold)									
Schools	177–341	169–312	213–505	89–144	78–116	101–178	50–94	53–103	53–96
Teachers	4,566– 9,915	4,267– 8,783	6,050– 16,042	2,397– 4,062	2,109– 3,290	2,771– 5,153	1,071– 2,279	1,116– 2,529	1,102– 2,316
<b>Panel B: Falsification Test, Years 2010 and 2011</b>									
ADP	0.212	0.121	0.090	0.208	0.117	0.089	0.232	0.142	0.089
LATE	0.026 (1.328)	0.015 (0.993)	0.013 (0.895)	0.015 (0.305)	0.011 (0.544)	0.004 (0.012)	0.047 (0.925)	0.033 (0.767)	0.013 (0.600)
Bandwidth	0.440	0.292	0.442	0.249	0.241	0.234	0.269	0.238	0.400
Obs. (left-right from the threshold)									
Schools	245–870	227–639	243–804	132–293	135–302	128–282	49–86	47–76	68–131
Teachers	6,000– 15,198	4,291– 8,747	6,024– 15,198	2,812– 5,392	2,753– 5,232	2,722– 5,004	1,057– 1,970	951– 1,715	1,534– 3,244

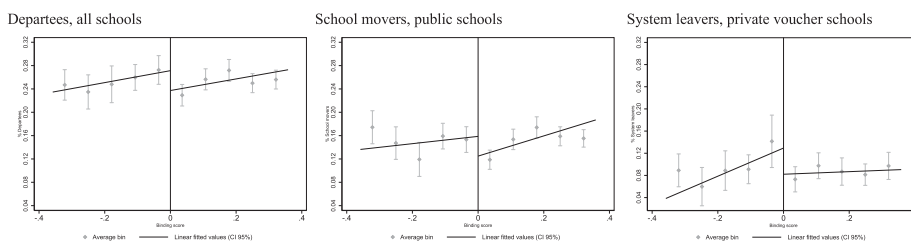
Notes: Each cell shows results from separate regressions. All dependent variables are dichotomous and equal to zero for teachers staying in the same school in  $t + 1$ . All estimates are based on mean square error (MSE)-optimal bandwidth choice, triangular kernel, and covariates at the school level measured in  $t - 1$ : 4th-grade test scores on math and reading, maximum schooling years between both parents, school vulnerability index, proportion of teachers leaving the school, and student enrollment. Inference involves recentering of point estimator to correct for potential bias associated with MSE-based bandwidth selection, and rescaling standard errors to be consistent with point estimator recentering that are clustered at the school level. Corresponding z-scores in absolute value are shown in parentheses. ADP = Average of dependent variable for emerging schools to the right of the threshold within the optimal bandwidth. LATE = Local Average Treatment Effect. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

induced by accountability varies by public and private schools and different groups of teachers. Second, we show differences between teachers who moved and those who stayed one year after schools were ranked as *in-recovery*. Finally, we present the effects of school accountability on new teacher hires.

### Teacher Mobility Results

#### Main Effects by Type of School and Groups of Teachers

Table 4 presents the effect that working at an *in-recovery* school has on teacher mobility as measured by the three binary variables described in section 3. For ease of interpretation, the coefficients were multiplied by  $-1$ , thus a positive value corresponds to an increase in the likelihood of leaving the school. Panel A shows the effects of the accountability policy on affected teachers, and panel B the results from a falsification test using teacher mobility for 2010 and 2011 with the corresponding covariates measured in 2009 and 2010. All estimations were performed separately for all schools, public schools, and private voucher schools. Columns 1, 4, and 7 display the effect on the likelihood of leaving the school either by moving to a different school or exiting the school system entirely for each of these three groups of schools, respectively. The results including all schools indicate that the accountability policy did in fact induce teachers



Notes: Each dot represents the average of the binding-score binned in five equally distant categories within the optimal bandwidth. The x-axes show the schools' average binding-score, and a horizontal line is at the cutoff (equal zero). The y-axes show the average teacher mobility for each of the three teacher mobility variables analyzed: (i) departees in all schools; (ii) school movers in public schools, and (iii) system leavers in private voucher schools. The straight lines correspond to the local estimation based on a polynomial equal one within the optimal bandwidth when analyzing departees in all schools.

Figure 2. Main Results

working at *in-recovery* schools to either move to another school or leave the system. The magnitude of this effect is 4.4 percentage points.

The results by type of school show that the mobility induced by accountability seems to differ depending on school management. Whereas in public schools, being ranked as *in-recovery* led teachers to move to other schools instead of leaving the system (columns 5 and 6), teachers in private voucher schools were relatively more likely to exit the system altogether, when compared with their peers at public schools (columns 6 and 9).<sup>3</sup>

To test the robustness of these results, we independently replicated the estimations using different bandwidths and polynomials of orders 1 and 2 (see online Appendix 5). The overall effect and within-system mobility at public schools are robust to all bandwidths and functional forms. Meanwhile, the finding that teachers at private voucher schools leave the system is slightly less robust, though it is more relevant when including in the analysis schools that are relatively further away from the cutoff.

To gain a deeper understanding of how these three main results (i.e., teacher departure from *in-recovery* schools, public school teacher moves, and private voucher school teacher exits—see figure 2) are distributed between different teacher profiles, we follow the same identification strategy and separately estimated the effect of accountability on teacher mobility for twelve distinct groupings of teachers. The teacher groups differ relative to their type of employment contract, number of schools at which they work, performance on college admission exam, national assessment ranking (conducted annually for teachers in public schools), teacher experience, and whether they teach in a school grade including subjects tested in the accountability system (first to fourth grades) or not (other grades). Table 5 displays the main effects. When all schools are included in the analysis, the effects on teacher mobility are concentrated among teachers

3. We also constructed a statistical test to study differences between coefficients across sectors, which assume independence between the effects in public and private voucher schools. Next, based on the bias corrected coefficients (which differ from the conventional coefficients in tables 4 and 5) and robust standard errors, we computed a z-score of the difference of effects by sector (see Appendix 4). The results are presented in table A3. In general, they suggest a higher likelihood to leave the school system among teachers working in private voucher schools. The differences between system leavers at private v. public schools concentrate among more experienced teachers working in one school, hired through temporary contracts, and teaching in non-accountability grades. We thank Matias Cattaneo for his guidance and help implementing this test. More details can be found in Calonico et al. (2020).



**Table 5.** Regression Discontinuity Estimates of the Effect of Working at an *In-recovery* School on Teacher Mobility for Different Groups of Teachers

	All Schools						Public Schools						Private Voucher Schools					
	Departees		School Movers		System Leavers		Departees		School Movers		System Leavers		Departees		School Movers		System Leavers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)									
Permanent contract (nt = 7,147; ns = 763; av = .58; LM = .13)	-0.002 (0.032)	0.016 (1.014)	-0.013 (0.692)	0.0180 (0.694)	0.074*** (3.126)	-0.026 (0.996)	-0.010 (0.179)	-0.007 (0.248)	0.003 (0.024)									
Temporary contract (nt = 10,157; ns = 1,314; av = .43; LM = .32)	0.075*** (2.911)	0.064*** (2.982)	0.017 (0.967)	0.072** (2.207)	0.079** (2.480)	-0.017 (0.812)	0.176*** (2.973)	0.088** (2.001)	0.107** (2.146)									
Works in one school (nt = 12,508; ns = 832; av = .83; LM = .19)	0.015 (0.733)	0.025* (1.660)	0.008 (0.728)	-0.004 (0.053)	0.029 (1.360)	-0.038* (1.928)	0.091* (1.744)	0.017 (0.482)	0.076** (1.976)									
Works in two or more schools (nt = 2,344; ns = 844; av = .17; LM = .29)	0.159*** (3.526)	0.141*** (3.167)	0.009 (0.658)	0.197*** (3.794)	0.183*** (3.100)	0.021 (0.823)	0.050 (0.738)	0.040 (0.637)	0.005 (0.053)									
Lower score on CAE (nt = 1,397; ns = 774; av = .41; LM = .31)	0.138*** (1.991)	0.157*** (2.766)	-0.018 (0.385)	0.097 (1.132)	0.111 (1.175)	-0.025 (0.25)	0.382*** (2.915)	0.294** (2.388)	0.076 (0.974)									
Higher score on CAE (nt = 1,436; ns = 779; av = .59; LM = .32)	0.052 (0.860)	-0.031 (0.245)	0.063 (0.886)	0.051 (0.366)	0.022 (0.492)	-0.030 (0.588)	0.021 (0.049)	0.092 (0.738)	-0.028 (0.328)									
Lower score on NTA (nt = 1,875; ns = 432; av = .29; LM = .21)	-	-	-	0.030 (0.789)	0.033 (1.02)	-0.020 (0.534)	-	-	-									
Higher score on NTA (nt = 3,436; ns = 443; av = .71; LM = .15)	-	-	-	0.021 (0.875)	0.060** (2.118)	-0.022 (1.176)	-	-	-									
Less experienced (< 3 years) (nt = 4,048; ns = 1,065; av = .24; LM = .34)	0.091** (2.203)	0.114*** (3.161)	-0.007 (0.409)	0.087 (1.545)	0.122*** (2.756)	-0.044 (1.197)	0.219*** (2.604)	0.178*** (2.918)	0.028 (0.534)									
More experienced (> 3 years) (nt = 11,215; ns = 818; av = .76; LM = .17)	0.030 (1.269)	0.024 (1.498)	0.005 (0.323)	0.025 (0.89)	0.068** (2.352)	-0.015 (0.848)	0.069 (1.441)	0.009 (0.210)	0.063** (2.08)									
Non-accountability grades (nt = 9,023; ns = 850; av = .77; LM = .21)	0.029 (1.151)	0.039* (1.767)	0.000 (0.096)	0.016 (0.621)	0.057** (2.005)	-0.031 (1.462)	0.072 (1.441)	0.009 (0.222)	0.066* (1.761)									
Accountability grades (1-4) (nt = 3,959; ns = 1,013; av = .23; LM = .22)	0.060* (1.857)	0.075*** (2.788)	-0.019 (0.749)	0.058 (1.298)	0.098** (2.176)	-0.036 (1.226)	0.085 (1.293)	0.089** (2.011)	0.013 (0.458)									

Notes: Each cell shows results from separate regressions. All dependent variables are dichotomous and equal to zero for teachers staying in the same school in  $t + 1$ . All estimates are based on mean square error (MSE)-optimal bandwidth choice, triangular kernel, and covariates at the school level measured in  $t - 1$ . 4th-grade test scores on math and reading, maximum schooling years between both parents, school vulnerability index, proportion of teachers leaving the school, and student enrollment. Inference involves recentering of point estimator to correct for potential bias associated with MSE-based bandwidth selection, and rescaling standard errors to be consistent with point estimator recentering that are clustered at the school level. Corresponding z-scores in absolute value are shown in parentheses. nt = number of teachers; ns = number of schools; av = average of variable of interest; LM = percent of departees in  $t + 1$ ; CAE = college admission exam; NTA = national teacher assessment. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ .

with temporary contracts, teachers who work in two or more schools, teachers who had lower scores on the college admission exam, and those with less work experience. These averages may, however, hide heterogeneous effects by management type. In public schools, teacher mobility seems to be exclusively concentrated in within-system moves, with teachers working in two or more schools being the group with the highest likelihood of moving to a different school. Accountability-induced mobility is also higher among those public school teachers with less work experience, hired through both permanent and temporary contracts, those with a higher score on the teacher assessment, and, to a lesser degree, teachers with more work experience.

In private voucher schools, the pressure exerted from accountability is, for some groups, linked to increased mobility to a different school, and for others, to leaving the system altogether. With regard to the type of teachers moving out of *in-recovery* schools, the effect is concentrated among teacher characteristics commonly correlated with lower value added such as teachers with temporary contracts, with a lower score on the college admission exam, with less working experience, and, to a lesser degree, who are working in one school and have more years of experience, with the latter effect only being statistically significant when analyzing the probability of leaving the system.

In addition to proxies of teacher value added, we followed a similar strategy as Feigenberg et al. (2019) to test the heterogeneity of the accountability effect by school grade. Specifically, we separately analyzed teacher mobility for those teaching in any grade between first and fourth grade (grades reviewing and teaching curricular content that is more likely to be evaluated on the fourth-grade national assessments used to rank schools in the accountability system) and teachers teaching in other grades. The results suggest that teachers in accountability grades at public and private schools are more likely to move to other schools in response to accountability. Among those teachers in grades not used for calculating the accountability ranks, it seems that they are more likely to move to other schools in the public sector and to leave the system in private schools. In summary, teachers in public schools appear to be more likely to move to other schools regardless of the grade they are teaching, while teachers at private schools in non-accountability grades seem more inclined to exit the system, and to move to other schools if they were teaching in accountability grades.

In interpreting these results, it should be noted that we have data on college admission exam performance for teachers that entered teaching programs only after the year 2004. While this metric has been shown to be predictive of teacher effectiveness, our results should be taken with caution since our sample of teachers with college admission data is more limited, and the predictive power of this score on teacher effectiveness appears to be weaker than that associated with the teacher assessment scores (Gallegos, Neilson, and Calle 2019; San Martín et al. 2013).

Another issue worth noting is that the teacher mobility induced by the accountability policy appears to be unrelated to changes in student enrollment. One might assume that a potential mechanism pushing teachers to depart in the first years of the policy from *in-recovery* schools was decreasing enrollment. In other words, under the assumption that fewer students would require fewer teachers, an increased likelihood of teacher mobility may have been a response to families opting out of *in-recovery* schools. This does not, however, seem to be the case in Chile. Specifically, following the same methodological approach defined in section 3, we do not find that school

accountability affected either total student enrollment, the enrollment of the incoming cohort (i.e., typically at prekindergarten or kindergarten), or enrollment two years after schools were assigned this ranking. In other words, while declining enrollments is observed at both sides of the threshold for *in-recovery* and *emerging* schools, no systematic difference is attributable to the introduction of accountability. We present these results in further detail in online Appendix 6.

In summary, our main results suggest that the pressure of accountability increased the likelihood of teacher departure from *in-recovery* schools in Chile. It remains, however, unclear the extent to which this was a decision on the part of teachers or school administrators. In private voucher schools, the concentration of mobility among teachers with temporary contracts, those who work in a single school, and with teaching scores associated with lower productivity suggests that school administrations may have played an important role in inducing staff changes. With regard to public schools, the story is less clear. Mobility is characterized by within system moves, with a higher concentration among teachers who work in two or more schools, and a minor difference in terms of the type of contract. Furthermore, teachers with higher scores on the national assessment also appear to be more likely to move from an *in-recovery* school to a different establishment. To better understand these moves within the school system, in the next subsection we look at the differences between teachers who stayed in *in-recovery* schools and those who moved to another establishment one year after their school began to experience pressure associated with the threat of closure.<sup>4</sup>

#### *Destination School Differences for Movers and Stayers*

To compare the working conditions between *in-recovery* school stayers and movers, we begin by defining ten dependent variables measuring school and position characteristics one year after the school received their ranking. These include whether teachers move to schools of a different network or management type, contract hours, the number of schools in which teachers worked, and the 2012–13 averages for the binding score, SEP ranking, SIMCE performance, and socioeconomic status. For those teachers working in two or more schools, we used the average values for all of their schools. We then followed a linear regression approach to study position differences between stayers and movers. The variable of interest is a dummy equal to 1 for movers and 0 for stayers, and the model includes school of origin fixed effects and clustered standard errors at the school level. Specifically, we estimate the following ordinary least squares regression models:

$$Y_{i,s,t} = \alpha + \pi \cdot IRM_{i,s,t-1} + \theta_{s,t-1} + \varepsilon_{i,s,t}, \quad (3)$$

where  $Y_{i,s,t}$  represents characteristics of the school or position of destination in the year after schools were ranked as *in-recovery* for teacher  $i$  in school  $s$ .  $IRM_{i,s,t-1}$  is a binary variable equal to one if the teacher  $i$  who was working at an *in-recovery* school at  $t - 1$  moves to another school in the year immediately following the school classification

4. Another exercise to better understand the dynamics associated with accountability is to examine whether longer exposure to greater accountability pressure has a stronger effect. We do not have enough statistical power to explore this question by type of school management; however, we study the differential effects between schools ranked as *in-recovery* once vs. those ranked as *in-recovery* twice. The details are presented in online Appendix 7.

and zero otherwise.  $\theta_{s,t-1}$  represents school of origin by year fixed effects, and  $\epsilon_{i,s,t}$  is an error term. The differences in the characteristics of the destination school or position between movers and stayers at *in-recovery* is captured by the coefficient  $\pi$ . The standard errors are clustered at the school level.

Given that mean reversion would predict an improvement in school characteristics for those moving from *in-recovery* schools independent of the potential effects of accountability, we also use the same models to estimate differences in the destination schools for teachers working in 2010 and 2011 in the same schools ranked as *in-recovery* in 2012 and 2013. That is, we examine differences between movers and stayers in 2010 at schools ranked as *in-recovery* in 2012, and between movers and stayers in 2011 at schools ranked as *in-recovery* in 2013. The sample for this analysis is composed of teachers that have worked in any of the 259 different schools that were ranked as *in-recovery* in 2012 or 2013 and that were also operating in 2010 and 2011. To evaluate differences between the cohorts of teachers unexposed (i.e., 7,263 teachers working at these schools in 2010–11) and exposed to the accountability ranks (i.e., 7,207 teachers working at these schools in 2012–13), we used a statistical test of mean differences that assumes independence between their corresponding coefficients.<sup>5</sup>

Table 6 presents a summary of the overall results as well as separate results for teachers in public schools and those in private voucher schools. The left panel shows results for the 2012–13 teacher cohorts, and the right panel the results for the 2010–11 teacher cohorts. The general pattern is the same for both cohorts, and suggests that teachers who leave their *in-recovery* school to go to another school, compared with their peers who stay in the same schools, are more likely to work in schools with higher binding scores and a higher SEP ranking, with better SIMCE outcomes, more educated parents, and lower social vulnerability. On average, they also tend to work for about 1.2 fewer hours a week and in fewer schools. There are no statistically significant differences across teacher cohorts when we examine teachers in public and private voucher schools, based on mean difference tests and a 95 percent confidence level.

When we analyze the results by school sector, the patterns across teacher cohorts also do not differ significantly. For both teacher cohorts, the main difference between movers from public schools and those from private voucher schools is the likelihood of continuing to work in the same school network. For movers in the public sector, the probability of changing employer is slightly below 0.50, while the corresponding value for those moving from private voucher schools is 0.96. In other words, whereas almost all teachers moving from low-performing private sector schools change employers, about half of teachers who move from publicly managed schools continue to work for the same municipal government.

There are no sharp differences by type of school management for the remaining variables. In general, teachers moving from private voucher schools improve their

5. Specifically, we computed a z-score based on the distribution of the estimate of  $\pi$  in equation 3 for 2012–13 teacher cohorts ( $\hat{\pi}_1 \sim N(\mu_1, \sigma_1^2)$ ) and the distribution of the estimate of  $\pi$  for 2010–11 teacher cohorts ( $\hat{\pi}_0 \sim N(\mu_0, \sigma_0^2)$ ), as  $zscore = (\mu_1 - \mu_0) / (\sigma_1^2 + \sigma_0^2)$ , with  $\mu_1$  as the parameter capturing differences between movers and stayers at *in-recovery* schools in 2012–2013, and  $\sigma_1^2$  as its standard deviation. Similarly,  $\mu_0$  and  $\sigma_0^2$  are the corresponding values for differences between movers and stayers at *in-recovery* schools in 2010–11. We tested for mean differences for (1) all teachers, (2) teachers at public schools, (3) teachers at private voucher schools, and (4) differences between teachers at private voucher schools and public schools. When we detected differences at a 95 percent confidence level, we reported them in table 6 and online appendix table A7 in online Appendix 8.

**Table 6.** Destination School Characteristics for Movers from *In-recovery* Schools Compared to Stayers in the Same Schools

	Movers (IR) v Stayers (IR) 2012–13				Movers (IR) v Stayers (IR) 2010–11			
	All schools (1)	Public schools (2)	PV schools (3)	Columns (3)–(2) (4)	All schools (5)	Public schools (6)	PV schools (7)	Columns (7)–(6) (8)
Changes school network (nt = 7,207; ns = 259; ay = .127; ax = .179)	0.651*** (0.024)	0.474*** (0.022)	0.960*** (0.021)	0.486*** (0.031)	0.641*** (0.026)	0.449*** (0.026)	0.981*** (0.010)	0.532*** (0.029)
Changes type of management (nt = 7,207; ns = 259; ay = .051; ax = .179)	0.290*** (0.015)	0.289*** (0.019)	0.292*** (0.022)	0.003 (0.029)	0.295*** (0.018)	0.305*** (0.021)	0.277*** (0.031)	-0.028 (0.037)
Hours of contract (nt = 7,207; ns = 259; ay = 39.423; ax = .179)	-1.233*** (0.368)	-1.231*** (0.454)	-1.236*** (0.633)	-0.005 (0.775)	-0.831*** (0.414)	-0.978** (0.492)	-0.571 (0.754)	0.407 (0.895)
Number of schools (nt = 7,207; ns = 259; ay = 1.180; ax = .179)	-0.035** (0.016)	-0.034 (0.021)	-0.037 (0.024)	-0.003 (0.032)	-0.074*** (0.019)	-0.099** (0.024)	-0.031 (0.031)	0.068* (0.039)
Binding score (nt = 6,700; ns = 259; ay = -.086; ax = .117)	0.994*** (0.042)	0.822*** (0.047)	1.345*** (0.070)	0.523*** (0.083)	0.973*** (0.041)	0.889*** (0.045)	1.161*** (0.078)	0.272*** (0.088)
Destination school is IR (nt = 7,207; ns = 259; ay = -.755; ax = .179)	-0.768*** (0.023)	-0.721*** (0.031)	-0.850*** (0.026)	-0.129*** (0.041)	-0.796*** (0.022)	-0.782*** (0.028)	-0.820*** (0.037)	-0.038 (0.046)
SIMCE math (nt = 6,966; ns = 259; ay = -.578; ax = .156)	0.329*** (0.027)	0.284*** (0.034)	0.406*** (0.044)	0.122** (0.055)	0.371*** (0.028)	0.333*** (0.035)	0.436*** (0.047)	0.103* (0.058)
SIMCE language (nt = 6,968; ns = 259; ay = -.545; ax = .156)	0.312*** (0.026)	0.269*** (0.031)	0.386*** (0.044)	0.117* (0.054)	0.342*** (0.028)	0.296*** (0.033)	0.423*** (0.048)	0.127** (0.057)
Parents' schooling (nt = 6,980; ns = 259; ay = 10.782; ax = .157)	1.220*** (0.080)	1.109*** (0.098)	1.410*** (0.137)	0.301* (0.168)	1.292*** (0.088)	1.206*** (0.099)	1.441*** (0.169)	0.235 (0.194)
Vulnerability Index (nt = 7,029; ns = 259; ay = 80.419; ax = .158)	-10.156*** (0.635)	-9.378*** (0.724)	-11.527*** (1.194)	-2.149 (1.388)	-10.589*** (0.688)	-10.596*** (0.799)	-10.576*** (1.297)	0.020 (1.513)

Notes: All regressions are based on 259 schools ranked as *in-recovery* in 2012 or 2013. Descriptive statistics below each dependent variable correspond to 2012–13 teacher cohorts. Each cell shows results from separate regressions. The left panel shows results for teacher cohorts 2012–2013. The right panel shows results for teacher cohorts 2010–2011. Differences between schools reported in columns (4) and (8) are captured by an interaction term between the variable of interest and a binary variable differentiating private voucher schools (=1) from public schools (=0). All specifications include school fixed effects. Standard errors are clustered at the school level and shown in parentheses. IR: *In-recovery*; PV: Private voucher; nt: number of teachers; ns: number of schools; ay: average dependent variable; ax: average variable of interest (descriptive statistics below the name of each dependent variable corresponds to all teachers in 2012–2013 – sample for estimates reported in column [1]). \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ . Differences in means tests ( $p > 0.05$ ) between coefficients in columns (1)–(5); (2)–(6); (3)–(7); and (4)–(8) are identified by †. SIMCE = Education Quality Measurement System.

working conditions slightly more relative to their counterparts moving from public schools. For instance, while teachers moving from public schools end up working in establishments with students scoring 0.285 standard deviation higher in SIMCE math compared with those of their peers who did not move, the corresponding difference for teachers moving from private voucher schools is 0.406 standard deviation. A similar pattern emerges for SIMCE reading scores, family socioeconomic status, likelihood of continuing to work in an *in-recovery* school, and average binding score.

Teachers who moved thus benefited, though not in every dimension. While the schools they move to seem to offer better working environments, these teachers also work fewer hours and in fewer schools.<sup>6</sup> In addition, these patterns do not differ significantly compared to movers before the introduction of the *in-recovery* rank in 2012, which suggest that while the accountability system induced higher teacher mobility (see table 4), the characteristics of the destination schools for these additional movers was similar to the characteristics of the destination schools for other movers from these schools that were not responding to the accountability system. This combination of results may be the outcome of a negotiation process between employers and teachers in the context of high pressure to improve.

#### *Did In-recovery Schools Respond by Hiring New Teachers?*

The last part of our analysis investigates whether *in-recovery* schools responded to accountability pressure and teaching staff mobility by increasing new teacher hires. We follow the same methodological approach described in section 3, and estimate the effects of accountability on the likelihood of new hires one and two years after schools received their SEP ranking.

The overall results, as well separate results for public and private voucher schools, are reported in columns 1–3 of table 7. The results show no discernable impact of accountability on the hiring of new teachers for any of the groups examined. In other words, the school accountability policy in Chile seems not to have incentivized the recruitment of new teachers at *in-recovery* schools, notwithstanding the departure of some of their teachers.

To gain a better understanding of why schools do not respond to the increase in teacher departees by hiring more teachers, we studied the impact of accountability on the year-to-year changes in the student–teacher ratio (YSTR).<sup>7</sup> Columns 4–6 of table 7 show that accountability increased the YSTR, mainly in private voucher schools. This occurred in a context of a decreasing YSTR trend in low performing and high poverty schools. This trend is probably due to a combination of demographic changes and parental preferences that have led to a reduction in the size of student cohorts over time at low-performing schools. Since the average YSTR is similar in magnitude to the effect of accountability on the YSTR, class sizes are probably decreasing overall but

6. Another comparison group we used to assess the destination school characteristics for movers from *in-recovery* schools is all other teachers at schools ranked by SEP as *in-recovery* or *emerging*. The results suggest that movers from *in-recovery* schools move to schools below average on academic outcomes such as SIMCE; however, this pattern is similar to the one observed for teacher cohorts before the introduction of the SEP ranks. The details are presented in online Appendix 8.

7. In addition, we also studied the effects of accountability on student–teacher ratio, total number of teachers, and total number of teaching hours. The results are presented in online Appendix 9.

**Table 7.** Regression Discontinuity Estimates of the Effect of Being Classified as *In-recovery* on New Hires

	New Hires in $t + 1$			Student: Teacher Ratio $\Delta(t + 1, t)$		
	All (1)	Public Schools (2)	Private Voucher Schools (3)	All (4)	Public Schools (5)	Private Voucher Schools (6)
ADP	0.209	0.206	0.212	-1.091	-0.962	-1.524
LATE	-0.010 (0.383)	0.016 (0.725)	-0.018 (0.646)	0.952** (2.328)	0.516 (1.163)	1.587** (2.119)
Estimation bandwidth	0.455	0.196	0.369	0.226	0.241	0.217
Obs. (left-right from threshold)						
Number of schools	245–879	126–268	88–227	145–249	101–190	45–72
Number of teachers	5,960–16,276	2,305–4,057	1,444–2,994	160–292	111–219	51–85

Notes: Each cell shows results from separate regressions. All dependent variables are dichotomous and equal to one if the teacher was not present in the same school in  $t - 1$  (or  $t - 2$ , respectively), and zero for teachers that were present in the same school in  $t - 1$  (or  $t - 2$ , respectively). All estimates are based on mean square error (MSE)-optimal bandwidth choice, triangular kernel, and covariates at the school level measured in 2011: 4th grade test scores on math and reading, maximum of schooling years between both parents, school vulnerability index, proportion of teachers leaving the school  $t - 1$ , and student's enrollment. Inference involves recentering of point estimator to correct for potential bias associated to MSE-based bandwidth selection, and rescale standard errors to be consistent with point estimator recentering that are clustered at the school level. Corresponding z-scores in absolute value are shown in parentheses. LATE = Local Average Treatment Effect; ADP = Average of dependent variable for emerging schools to the right of the threshold within the optimal bandwidth. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ .

maintaining a similar level at *in-recovery* schools. In other words, it is plausible that *in-recovery* schools that are losing teachers do not hire new teachers to replace them, because they are also facing a decline in enrollments.

### 5. DISCUSSION AND CONCLUSIONS

Conflicting findings on the impact of accountability on teacher mobility persist in the literature. Critics express concern that the negative stigma of labeling schools as low performing discourages effective teachers from continue to work at disadvantaged establishments. Advocates maintain that accountability creates incentives for schools to retain their highest performing teachers so as to avoid negative sanctions, and/or to dismiss their least effective teachers. This argument is predicated on the assumption that low-performing schools will be able to replace their lowest performing teachers with better candidates.

Our novel analysis explores this question in the context of a developing country. Specifically, we exploit a set of arbitrary rules used to classify low-performing schools in Chile, allowing us to evaluate the impact of school accountability on teacher mobility, as well as the heterogeneous effects by type of school management (public or private). The overall effect we observe aligns with that of most empirical studies; namely, that classifying a school as low-performing increases teacher departure the following year. In addition, we find that public school teachers typically move to other schools, while private school teachers are relatively more likely to leave the system altogether. In both cases, the teacher departure induced by accountability is not offset by the hiring of new teachers.

The increased teacher mobility caused by the accountability policy was larger for certain groups of teachers and differed by school management type. Specifically, teachers with less work experience and those scoring lower on their college admission exam have a higher likelihood of leaving privately managed *in-recovery* (i.e., lowest-ranking)

schools. Given that we have data on college admission exam outcomes only for younger teachers, our findings suggest that, among less experienced teachers, average teacher effectiveness may have improved at these *in-recovery* schools. Yet, our results also indicate that departing teachers are not being replaced by new recruits, which may mean an increased workload for those who stay, and/or make teaching more challenging with a reduced staff.

At public schools, teacher mobility largely consists of moves within the system, where most continue to work for the same employer. It is unclear whether this redistribution of teachers in the public sector leads to higher levels of efficiency. Given that teachers are moving to other schools and not leaving the system altogether, it is possible that overall teacher effectiveness does not change in response to the accountability system. However, if the teacher redistribution is such that fewer teachers are working in multiple schools, the overall effectiveness in the sector is likely to increase as workload decreases (Elacqua and Marotta 2020).

Taken together, our findings indicate that the higher observed teacher mobility is related both to teachers preferring to avoid *in-recovery* schools and to school principals deciding not to retain less experienced teachers, who are less likely to be effective (Clotfelter, Ladd, and Vigdor 2006; Kane et al. 2008). Other studies of the Chilean accountability system support this view. Elacqua et al. (2016), for example, present suggestive evidence that *in-recovery* schools in the city of Santiago initially reacted to the accountability policy by implementing a series of practices aimed at improving student achievement in the short run, such as reallocating the most experienced teachers to the fourth grade (whose national evaluation results form, in part, the basis of the school's classification). Yet this was also accompanied by less time and resources invested in teacher training and evaluation. Similarly, Falabella (2016) provides qualitative evidence suggesting that government pressure tends to overwhelm low-performing schools in Santiago, leading them to adopt teaching strategies focused on content and skills measured on the standardized tests, as opposed to developing a more comprehensive approach to student learning. On the one hand, such practices might deteriorate teacher working conditions, especially for those who are less experienced and have a greater need to learn during the early years of their career. On the other hand, schools that must urgently need to improve their academic outcomes might benefit from having fewer teachers who need training.

The overall effectiveness of accountability is unclear. More research is needed on how school administration and management can improve working conditions and academic outcomes at *in-recovery* schools. It is crucial to determine whether such schools can attract effective teachers, how they can reallocate school resources more efficiently, and how they can promote a culture favoring student engagement in the learning process. Similarly, a better understanding of teacher preferences and behaviors would beneficially inform school policies. How teachers perceive working at a school labeled as low performing might differ for more and less effective teachers. Studying teachers' rationales—those that consider, for example, working at an *in-recovery* school as a positive career experience, or alternatively, as something that may hurt their future job prospects—can also contribute to improving the design of accountability mechanisms.

While further work is needed to determine whether Chile's accountability policy benefits or hurts low-performing schools, the failure of *in-recovery* schools to attract



effective teachers to offset the turnover generated by the reform suggests that it may not be the optimal solution to boosting student learning. Complementary policies, such as increasing monetary incentives for teachers who work in low-performing schools or channeling more resources and technical and pedagogical support into *in-recovery* schools, could strengthen the potential effectiveness of this type of accountability policy.

#### ACKNOWLEDGMENTS

We thank the staff of the Chilean Ministry of Education, Agency for Quality in Education, and DEMRE for their invaluable work and cooperation, as well as Ofer Malamud, Jonathan Guryan, Amilcar Velez, Federico Crippa, Matias Cattaneo, Rocio Titinuik, and participants at Northwestern Causal Inference Workshop, LACEA, and Education Division of Inter-American Development Bank meetings. This study was funded by the Inter-American Development Bank (IDB) and hence reviewed by the organization before publication. Nonetheless, the IDB was not involved in the design of the study, data collection processes, interpretation of data, or writing of the article. No other organizations provided financial support. None of the authors have anything further to disclose. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the IDB, its Board of Directors, or the countries they represent. The authors have no conflicts of interests or financial and material interests in the results. All errors remain our own. The data used in this article were obtained from the Chilean Ministry of Education (<https://www.ayudameduc.cl/ley-transparencia>; <http://datos.mineduc.cl/dashboards/20031/descarga-bases-de-datos-de-cargos-docentes/>), the Agency for Quality in Education ([https://formulario.agenciaeducacion.cl/solicitud\\_cargar](https://formulario.agenciaeducacion.cl/solicitud_cargar)), and the Department of Evaluation, Measurement and Educational Registry (DEMRE – <https://ayuda.demre.cl/forminvestigador.dmr>).

#### REFERENCES

- Assaél, J., F. Acuña, P. Contreras, and F. Corbalán. 2014. Transformaciones en la cultura escolar en el marco de la implementación de políticas de accountability en Chile: Un estudio etnográfico en dos escuelas clasificadas en recuperación. *Estudios Pedagógicos (Valdivia)* 40(2): 7–26. 10.4067/S0718-07052014000300001
- Bellei, C., and X. Vanni. 2015. Chile: The evolution of educational policy, 1980–2014. In *Education in South America*, edited by S. Schwartzman, pp. 179–200. London, UK: Bloomsbury Academic.
- Boyd, D., P. Grossman, M. Ing, H. Lankford, S. Loeb, and J. Wyckoff. 2011. The influence of school administrators on teacher retention decisions. *American Educational Research Journal* 48(2): 303–333. 10.3102/0002831210380788
- Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff. 2005. The draw of home: How teachers' preferences for proximity disadvantage urban schools. *Journal of Policy Analysis and Management* 24(1): 113–132. 10.1002/pam.20072
- Cabezas, V., R. Paredes, F. Bogolasky, R. Rivero, and M. Zarhi. 2017. First job and the unequal distribution of primary school teachers: Evidence for the case of Chile. *Teaching and Teacher Education* 64:66–78. 10.1016/j.tate.2017.01.017
- Calonico, Sebastian, Matias Cattaneo, Max Farrell, and Rocío Titiunik. 2017. rdrobust: Software for regression discontinuity designs. *Stata Journal* 17(2): 372–404. 10.1177/1536867X1701700208
- Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82(6): 2295–2326. 10.3982/ECTA11757
- Carrasco, A., and H. M. Gunter. 2019. The “private” in the privatisation of schools: The case of Chile. *Educational Review* 71(1): 67–80. 10.1080/00131911.2019.1522035

- Cattaneo, M. D., M. Jansson, and X. Ma. 2020. Simple local polynomial density estimators. *Journal of the American Statistical Association* 115(531): 1449–1455. 10.1080/01621459.2019.1635480
- Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor. 2006. Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources* 41(4): 778–820. 10.3368/jhr.XLI.4.778
- Clotfelter, C. T., H. F. Ladd, J. L. Vigdor, and R. A. Diaz. 2004. Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management* 23(2): 251–271. 10.1002/pam.20003
- Dizon-Ross, Rebecca. 2020. How does school accountability affect teachers? Evidence from New York City. *Journal of Human Resources* 55(1): 76–118. 10.3368/jhr.55.1.1015.7438R1
- Elacqua, G., and L. Marotta. 2020. Is working one job better than many? Assessing the impact of multiple school jobs on teacher performance in Rio de Janeiro. *Economics of Education Review* 78(102015): 1–9.
- Elacqua, G., M. Martínez, H. Santos, and D. Urbina. 2016. Short-run effects of accountability pressures on teacher policies and practices in the voucher system in Santiago, Chile. *School Effectiveness and School Improvement* 27(3): 385–405. 10.1080/09243453.2015.1086383
- Espeland, W. N., and M. Sauder. 2007. Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology* 113(1): 1–40. 10.1086/517897
- Falabella, A. 2016. ¿Qué aseguran las políticas de aseguramiento de la calidad?: un estudio de casos en distintos contextos escolares. *Estudios Pedagógicos (Valdivia)* 42(1): 107–126.
- Feigenberg, Benjamin, Rui Yan and Steven Rivkin. 2019. Illusory Gains from Chile’s targeted school voucher experiment. *Economic Journal* 129:2805–2832. 10.1093/ej/uez023
- Feng, Li, David Figlio, and Tim Sass. 2018. School accountability and teacher mobility. *Journal of Urban Economics* 103:1–17. 10.1016/j.jue.2017.11.001
- Flores, C., S. A. P. Ortúzar, and C. Milesi. 2014. *Caracterización del proceso de contratación de docentes de Chile. Oportunidades para profesores recién egresados. Proyecto FONIDE NF711278*. Santiago: Mineduc.
- Freedman, S. W., and D. Appleman. 2009. “In it for the long haul”—: How teacher education can contribute to teacher retention in high-poverty, urban schools. *Journal of Teacher Education* 60(3): 323–337. 10.1177/0022487109336181
- Friedman, Milton. 1955. The Role of Government in Education. In *Economics and the public interest*, edited by Robert Solo, 123–144. New Brunswick, NJ: Rutgers University Press.
- Gallegos, S., C. Neilson, and F. Calle. 2019. Screening and recruiting talent at teacher colleges using pre-college academic achievement. IRS Working Papers No. 636. Princeton University. <http://arks.princeton.edu/ark:/88435/dsp013197xp935>.
- Gill, B., J. R. Lockwood, F. Martorell, C. Messan, Booker, G. Vernez, B. Birman, and M. Garet. 2009. *An Exploratory analysis of adequate yearly progress, identification for improvement, and student achievement in two states and three cities*. Washington, DC: U.S. Department of Education.
- Gjefsen, H. M., and Trude Gunnes. 2020. The effects of school accountability on teacher mobility and teacher sorting. MPRA Working Paper No. 69664. Ludwig Maximilian University of Munich. 10.1080/09645292.2020.1788514

- Grissom, J. A., S. L. Viano, and J. L. Selin. 2016. Understanding employee turnover in the public sector: Insights from research on teacher mobility. *Public Administration Review* 76(2): 241–251. 10.1111/puar.12435
- Hanushek, E. A., J. F. Kain, and S. G. Rivkin. 2004. Why public schools lose teachers. *Journal of Human Resources* 39(2): 326–354. 10.2307/3559017
- 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(8–9): 1477–1503. 10.1016/j.jpubeco.2005.11.002
- Johnson, S. M., J. H. Berg, and M. L. Donaldson. 2005. Who stays in teaching and why: A review of the literature on teacher retention. Project on the Next Generation of Teachers, Harvard Graduate School of Education.
- Kane, T. J., J. E. Rockoff, and D. O. Staiger. 2008. What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review* 27(6): 615–631. 10.1016/j.econedurev.2007.05.005
- Ley de Subvención Escolar Preferencial N 20.248. 2008. Biblioteca del Congreso Nacional de Chile. <https://www.leychile.cl/Navegar?idNorma=269001>.
- Lin, N., and M. Dumin. 1986. Access to occupations through social ties. *Social Networks* 8(4): 365–385. 10.1016/0378-8733(86)90003-1
- Mbiti, I. M. 2016. The need for accountability in education in developing countries. *Journal of Economic Perspectives* 30(3): 109–132. 10.1257/jep.30.3.109
- McCrary, J. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2): 698–714. 10.1016/j.jeconom.2007.05.005
- O'Malley, M. P., and S. Nelson. 2013. The public pedagogy of student activists in Chile: What have we learned from the penguins' revolution? *Journal of Curriculum Theorizing* 29(2): 41–56.
- Ortúzar, M. S., P. Ayala, C. Flores, and C. Milesi. 2016. Percepciones acerca del Proceso de Búsqueda and Contratación de Docentes en Chile: Nudos críticos e Inequidad del Sistema. *Calidad en la educación* 45:251–287. 10.4067/S0718-45652016000200008
- Papay, J. P., J. B. Willett, and R. J. Murnane. 2011. Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics* 161:203–207. 10.1016/j.jeconom.2010.12.008
- Paredes, R., F. Bogolasky, V. Cabezas, R. Rivero, and M. Zahri. 2013. *Los determinantes del primer trabajo para profesores de Educación Básica en la Región Metropolitana*. FONIDE F611105. Santiago, Chile: Ministry of Education.
- Porter, K. E., S. F. Reardon, F. Unlu, H. S. Bloom, and J. R. Cimpian. 2017. Estimating causal effects of education interventions using a two-rating regression discontinuity design: Lessons from a simulation study and an application. *Journal of Research on Educational Effectiveness* 10(1): 138–167. 10.1080/19345747.2016.1219436
- Reardon, S., N. Arshan, A. Atteberry, and M. Kurlaender. 2010. Effects of failing a high school exit exam on course. *Educational Evaluation and Policy Analysis* 32:498–520. 10.3102/0162373710382655
- Reardon, S., and J. Robinson. 2012. Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness* 5:83–104. 10.1080/19345747.2011.609583

- Robinson, J. 2011. Evaluating criteria for English learner reclassification: A causal-effects approach using a binding-score regression discontinuity design with instrumental variables. *Educational Evaluation and Policy Analysis* 33(3): 267–292. 10.3102/0162373711407912
- Ronfeldt, M., S. Loeb, and J. Wyckoff. 2013. How teacher turnover harms student achievement. *American Educational Research Journal* 50(1): 4–36. 10.3102/0002831212463813
- Rouse, C. E., J. Hannaway, D. Goldhaber, and D. Figlio. 2013. Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy* 5(2): 251–281.
- San Martín, E., R. Rivero, M. Bascopé, and C. Hurtado. 2013. *¿Es la prueba INICIA una medida predictiva de efectividad docente?* FONIDE F711265. Santiago, Chile: Ministry of Education.
- Simon, N. S., and S. M. Johnson. 2015. Teacher turnover in high-poverty schools: What we know and can do. *Teachers College Record* 117(3): 1–36. 10.1177/016146811511700305
- Sims, S. 2016. *High-stakes accountability and teacher turnover: How do different school inspection judgements affect teachers' decisions to leave their school?* Department of Social Science, University College of London, England. Available <https://www.teachertoolkit.co.uk/wp-content/uploads/2016/11/qsswp1614.pdf>.
- Spillane, J. P. 2021. Performance-based accountability systems cross-nationally: Learning by comparing. *Educational Assessment, Evaluation and Accountability* 33:581–585. 10.1007/s11092-021-09370-y
- Wong, V., P. Steiner, and T. Cook. 2013. Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four methods. *Journal of Educational and Behavioral Statistics* 38:107–141. 10.3102/1076998611432172