IDB WORKING PAPER SERIES Nº IDB-WP-1074

The effects of accountability on the allocation of school resources: Regression discontinuity evidence from Chile

Gregory Elacqua
Analía Jaimovich
Alonso Román

Inter-American Development Bank
Education Division

October 2019

The effects of accountability on the allocation of school resources: Regression discontinuity evidence from Chile

Gregory Elacqua
Analía Jaimovich
Alonso Román

Inter-American Development Bank
Education Division

October 2019

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Elacqua, Gregory M., 1972-
The effects of accountability on the allocation of school resources: regression discontinuity evidence from Chile / Gregory Elacqua, Analia Jaimovich, Alonso Román.
p. cm. — (IDB Working Paper ; 1074)
Includes bibliographic references.
1. Educational accountability-Chile-Econometric models. 2. Education-Chile-Finance-Econometric models. 3. Government aid to education-Chile-Econometric models. 4. Educational vouchers-Chile-Econometric models. I. Jaimovich, Analia. II. Román, Alonso. III. Inter-American Development Bank. Education Division. IV. Title. V. Series.
IDB-WP-1074

Contact information: SCL-EDU@iadb.org

The effects of accountability on the allocation of school resources: Regression discontinuity evidence from Chile

August 2019

Gregory Elacqua, Analia Jaimovich & Alonso Román

Abstract

This research examines the effect of accountability threats for low performing schools on resource allocation decisions and provides evidence that schools act with strategic behavior only when the accountability pressure is high. We used a generalization of a traditional regression discontinuity design, taking advantage of the sharp discontinuity in the Chilean accountability system's ranking of schools based on performance measures, and of a unique school level expenditure data set, to make causal estimates of the effect of being ranked as "low-performing" on school spending decisions. The results indicate that, while first time low-performing schools do not change their resource allocation strategy, chronically underperforming schools are more likely to hire external technical pedagogical support and invest in teacher training that may help them boost achievement in the short and medium-term and avoid sanctions.

Keywords: School accountability, school spending, school finance, Chile, vouchers

JEL Classification: I22, I28, H52

1. Introduction

School accountability and high-stakes testing have been at the center of most major educational reform discussions over the last two decades. One innovation of accountability reforms is the use of student outcomes to evaluate teacher and school performance (Elmore et al., 1996; Figlio & Loeb, 2011; O'Day, 2002). By using different types of accountability systems, central governments set performance standards and choose between giving rewards to schools that meet the standards, and/or apply sanctions to the ones that fail to meet them. These rewards and sanctions can be explicit, through bonuses for example, or implicit through community pressure or signaling. Therefore, school accountability can work through direct government action or through the provision of information (Figlio & Loeb, 2011). Schools are often ranked or classified according to their performance and have a specific amount of time to improve their outcomes. If not, they face sanctions that range from mandatory improvement plans to closure.

The implementation of accountability systems has generated a persistent debate on the effects of these systems on student performance. Advocates argue that accountability pressures should have positive effects on academic outcomes in low-performing schools. However, critics maintain that these improvements may not be explained by real progress in student learning, but rather by strategic behavior that schools develop as they internalize the incentives (see for example Koretz, 2017). While this type of discussion has focused mainly on achievement gains and "gaming" accountability incentives, studies on how these interventions modify school resource allocation decisions are scarce (Booher-Jennings, 2005). This paper contributes to this debate by analyzing the effects of accountability pressures on school spending decisions in Chile.

The Chilean education system is an interesting case study to address this kind of questions since it presents a high-stakes accountability scheme within a school choice institutional arrangement. Chile is one of the few countries that, as part of a systemic reform introduced by the military regime in 1981, instituted a universal voucher program. Under this scheme, school quality was supposed to be assured by parent accountability. Similar to what Hirschman (1970) proposed for companies, when schools offer a low-quality education, parents have two options: they can leave the school ("exit") or they can express their dissatisfaction ("voice"). In a competitive schooling market, choice advocates maintain that low-quality schools would disappear, because they will lose students as a result of the exit and voice mechanisms (Hirschman, 1970). However, in the mid-2000s, despite substantial increases in funding and parental choice, education achievement gaps compared to OECD countries continued to persist. In response to persistent low performance, in 2008 the Chilean Congress enacted the *Subvención Escolar Preferencial*

Law (Preferential School Subsidy, Ley SEP) that, among other changes, introduced a national system of accountability for schools. Similar to other school accountability systems, SEP established minimum performance standards and ranked schools based on their performance on a national standardized test and other indicators. It also established sanctions for low-performing schools, including closure when a school did not show adequate improvement.

In this paper, we examine the effect of accountability threats for low performing schools on resource allocation. We use a regression discontinuity for our analysis, leveraging the sharp discontinuity in Chile's designation of schools to *in recovery* status (low-performing) based on performance measures. The identification strategy allows us to make a causal estimate of the effect of being ranked as *in recovery*, in combination with the threat of sanctions and the stigma of being classified as chronically underperforming, on school spending. We find evidence that being assigned to the treatment (*in recovery*) led to strategic behavior by school owners in the allocation of resources only when the accountability pressure is high, i.e. when schools have underperformed systematically over the years. While first time low performing schools do not change their resource allocation strategy, chronically underperforming schools are more likely to hire external technical pedagogical support and invest in teacher training that may help them boost achievement in the short and medium-term.

This paper is organized as follows. Section 2 lays out the theoretical framework and reviews the empirical literature on the effects of accountability pressures on different outcomes. Section 3 describes the school funding and the SEP accountability system in Chile. The next two sections discuss the methodology we employ and our data. Section 6 describes the RD design. Our results and final discussion are presented in sections 7 and 8.

2. Theoretical framework and literature

Accountability mechanisms have been implemented in various educational systems around the world. The No Child Left Behind Law (2001) in the United States, the Education and Inspection Law (2006) in England, and the SEP Law (2008) in Chile are amongst the most-developed accountability systems, and consequently have been the subject of extensive academic research. In all these cases, the government established performance goals and sanctions for schools that fail to meet them. An important aspect of accountability systems is the information content that allows families, teachers and policy makers a more effective way to assess how successful a school has been in meeting the achievement goals (Figlio & Loeb,

2011). Thus, the identification, classification, and subsequent publication of school rankings are all key components of these accountability systems. The objective of these actions is to increase the supervision of low-performing schools by parents and the government and to increase the pressure on schools to improve outcomes (Jacob, 2005). Previous research shows that the mere identification of low-performing schools operates as a social stigma for its principals, teachers, and students, increasing pressure to improve performance (Goldhaber & Hannaway, 2004). Once low-performing schools are identified, different sanctions are often gradually introduced, with the ultimate consequence of school closure (Brady, 2003). The assumption is that closing chronically underperforming schools would operate as an incentive for other low-performing schools to improve under the threat of closure (Smarick, 2010).

Many accountability studies have focused in school performance consequences (Dee & Dizon-Ross, 2017; Hanushek & Raymond, 2005; Dee & Jacob, 2011), and some have analyzed the effect of accountability pressures on school policies and practices. Rouse et al. (2013) and Elacqua et al. (2016) show that schools under high accountability pressures in Florida and Chile modified some of their internal practices and policies in educationally meaningful ways. Rouse et al. (2013) show that these changes explain performance gains in low performing schools in Florida. This is consistent with the findings of other studies in New York City, Chicago, Texas, which show that, after the introduction of accountability mechanisms, low-performing schools improved their test scores (Deming et al., 2016; Jacob, 2005; Rockoff & Turner, 2010).

Critics counter that accountability pressures can also produce undesirable effects. First, given that the performance standards set by the government measure only certain subjects from the curriculum, researchers have documented that schools spend more time on subjects that are included in the accountability index (Hannaway & Hamilton, 2007; Koretz & Barron, 1998). For example, in Kentucky, where students are evaluated in fifth grade, 82% of fifth-grade teachers reported that they increased instruction time for math, compared to 14% of fourth-grade teachers (Stecher & Barron, 2001). Similar results were found in Washington, California, Florida, Georgia, North Carolina, and Pennsylvania (Deming et al. 2016; Hamilton et al., 2007; Stecher et al.,2000; Hannaway and Cohodes, 2007; Ladd and Zelli, 2002).

Accountability pressures have also led teachers to try to “outsmart” standardized tests through various practices. First, some teachers alter the pool of students evaluated. For example, Figlio and Getzler (2006) find that some teachers reclassify low performing students as pupils with learning disabilities so that their scores are not counted in the assessment. Figlio (2006) finds that some schools suspend low-performing students the day of the test. Jacob and Levitt (2003) find that teachers in schools under accountability

pressure have a greater probability of helping students answer the tests. Pedulla et al. (2003) find that teachers provide 12% to 19% more time than stipulated for students to take the tests. There is also evidence that teachers pay more attention to students who are closer to surpassing the performance threshold established by the authorities, disregarding students who are far below or above the threshold (Booher-Jennings, 2005).

In the case of Chile, specifically, most econometric research on the SEP law has focused on its effects on student outcomes. Some studies show that the SEP subsidy has improved student performance (Nielson, 2013; Navarro-Palau, 2017; Bos et al, 2017), while others find that it did not (Aguirre, 2017; Feigenberg et al., 2017). Most of these studies, however, analyze the SEP law's general effects, regardless of the specific mechanism that may drive them. Indeed, the SEP law introduced several measures: an increase in the size of the voucher based on student characteristics, the mandatory development of School Improvement Plans, and an accountability mechanism that ranked schools based on their performance. Most of these studies cannot disentangle which of the specific mechanisms explain the results.

This paper innovates and contributes to the literature in two ways. First, this research contributes to the international literature on school accountability by, instead of focusing on how accountability affects student performance, it focuses on how accountability pressures may alter school spending decisions. To the best of our knowledge, this is the first paper to explore this relationship. This is a particularly relevant issue in school systems like Chile's where school owners and principals have significant autonomy over school budgets. Second, the paper contributes to the research on the SEP law in Chile by focusing specifically on the effects of the accountability mechanism on schools' decisions, isolating its effects from other aspects of the SEP law.

Thus, the paper analyzes the effects of the accountability system under the SEP law on schools' spending decisions. Specifically, we analyze the effect of being classified as a low performing school on schools' budget allocation. Faced with accountability pressures, schools may decide to respond in several ways. For instance, schools may increase investments in teachers or classroom support for teachers or spend less on inputs that may be less relevant to improve student performance. In contrast, they could also focus investments on non-classroom related activities, such as school uniforms or busing students or on expenditures that will improve the school's image such as publicity or safety. The goal of this paper is to gain insight into the decisions school managers make when faced with accountability pressures.

3. The SEP law and school finance and accountability in Chile

3.1. School Funding

The foundations of the current characteristics of the Chilean education system were first established through the reforms that took place in the early 1980s. In 1981, the military regime introduced a sweeping education reform package that included the decentralization of the administration of public schools from the central level (the Ministry of Education) to the local level (municipalities), and the introduction of a per-capita funding scheme. This funding scheme equalized public funding for municipal and private voucher schools, based on the number and attendance levels of students enrolled in the school.[1]

The original design of the per-capita funding scheme was based on a flat subsidy per student enrolled (and their attendance) in the school, measured in units of the *Unidad de Subvencion Escolar* (USE). The per-capita subsidy includes adjustments for school level, modality of education and geographic location. Since the 1990s, several changes were introduced to this funding scheme: private voucher schools were allowed to charge fees in addition to the government funding; higher adjustments for rurality and location were introduced; and special subsidies for students with special education needs, boarding schools, and student academic support were added.

With the exception of a few subsidies, schools were free to use the government funding as they saw fit. There were no specific regulations regarding the use of school resources, and school administrators were free to determine the expenditure structure that better suited each school's needs.[2] However, as a result of the per capita funding scheme, there were large disparities between schools in terms of their actual capacity to modify their expenditure structure. Small schools spent most of their resources on payroll and school operations, while larger schools tended to have greater financial freedom to spend resources on items other than payroll and operations (pedagogic equipment, school improvement services and teacher training).

[1] The Chilean educational system is composed of three types of schools, depending on their type of administration and source of funding: *municipal* (public) schools, which are financed with government subsidies and administered by the local municipal government, whose maximum authority is the mayor; *private voucher* schools, also financed with government subsidies, but administered by a private (for-profit or nonprofit, religious or secular) organization, and finally *private* schools, which are financed and administered privately.

[2] School administrators could profit from the government subsidy. This was changed in 2015, when the Inclusion Law prohibited for-profit schools, and established that all government subsidies had to be exclusively spent on school activities. This Law will also gradually eliminate school fees.

The most significant change to the allocation of government subsidies took place in 2008, when the Preferential School Subsidy Law (law No. 20248, *Subvención Escolar Preferencial* – SEP) was passed. SEP introduced for the first time a differentiated subsidy for disadvantaged students (low SES), based on the recognition that it is a more complex and expensive population to educate. In simple terms, the SEP law introduced an additional per capita subsidy (close to 80% higher than the base voucher) for students classified as vulnerable (*priority* students) who attended municipal or private voucher schools that voluntarily agreed to participate in the program.[3] The SEP law also established an additional subsidy for schools that had a high percentage of priority students (*Subvención por concentración*).[4] By 2015, 100% of municipal schools and 78% of private voucher schools were participating in the program.

The introduction of the SEP subsidy meant a large influx of resources to schools. In 2015, SEP accounted for 16% of total government subsidies to schools. However, unlike the base per capita subsidy, SEP resources were more restricted. The objective of the SEP subsidy was to compensate for the greater costs of educating socioeconomically disadvantaged students, but also to steer school expenditures towards quality improvements. Participating schools had to develop a school improvement plan (*Plan de Mejoramiento Educativo* – PME), and SEP resources had to be exclusively spent on actions included in these plans. These covered the areas of curriculum management, school leadership, school climate, and resource management, with special emphasis on improving priority students' educational outcomes.

More specifically, SEP resources could be spent exclusively on the following categories: (i) Curriculum management: strengthening of the school's educational mission, improvement of pedagogical practices, support for students with SEN, improvement of student assessment systems, additional academic and psychological support for students, and educational fieldtrips; (ii) School leadership: training for management teams, strengthening teachers' councils and relationships with the school community, and investing in civic education programs; (iii) School climate: psychological support and social assistance to students and their families, management of school climate, strengthening of the school council, and

[3] Priority students included: i) students whose family participate in the Chile Solidario (social welfare) system; ii) students whose families are among the most vulnerable third of families; iii) students whose parents or legal guardians are classified in section A of the National Health Fund; and iv) in the case that they are not included in one of the above criteria, the student classification considers household income, parental education, rurality, and poverty levels in the student's municipality. With the passing of the Inclusion Law in 2015, the SEP subsidy was extended to "preferential" students, reaching the lowest 60% of the income distribution.

[4] The original regulations established the preferential school subsidy per priority student from kindergarten to basic education, beginning with K to 4th grade of elementary school, with one additional grade of primary education being added in each subsequent year. In 2011, the SEP subsidy was gradually expanded to secondary education through the enactment of Law No. 20550. By 2016, all grade levels from kindergarten to 4th grade of secondary school were eligible to receive the subsidy.

support for student learning; (iv) Resource management: teacher professional development aimed at strengthening areas of the curriculum in which students have achieved unsatisfactory results, teacher and management team performance incentives related to PME goals, acquisition of technological aides for pedagogical activities (computers, interactive whiteboards, etc), and educational resources (school libraries, laboratories, etc). In 2011, additional expenditure categories were added to the items allowed to be purchased with SEP resources. These included hiring teachers, educational assistants, and other personnel necessary to improve the technical-pedagogical capacities of the school, the increase in working hours of existing teaching staff and assistants, and salary increases. However, it was established that these actions had to be based explicitly on the PME and could not exceed 50% of total SEP resources. Finally, the 2011 modification determined that 10% of SEP resources could be used at the central level for administrators that managed a network of schools. Resources at the central level were to be used in the administration, monitoring, and support of school improvement plans.

3.2. School accountability in the Preferential School Subsidy (SEP Law) policy

The SEP Law was the first initiative that introduced explicit school accountability mechanisms in Chile. In order for schools to receive the additional per capita subsidy for priority students, they have to comply with a number of requirements. The process of admission to SEP includes the signing of the Equality of Opportunities and Educational Excellence Agreement, which is valid for four years. This agreement can be renewed, provided that at least 70% of all SEP resources have been spent, and all expenditures are reported to the relevant authorities.

Through this agreement, school managers commit to documenting the use of resources, establishing goals for students' academic performance, and providing parents with information on school performance.[5] One of the most important requirements is that all schools that participate in the SEP program must develop and carry out a PME improvement plan, led by the school principal with the participation of the rest of the school community. Schools have the option of using technical educational assistance to develop

[5] Through the Equality of Opportunities and Educational Excellence Agreement, school managers assume responsibility for the following elements: i) submitting annually to the Ministry of Education and the school community a report on the use of SEP resources; ii) ensuring the effective functioning of the school council, the teachers' council and the PTA; iii) providing enough teaching hours to ensure curriculum coverage and non-teaching curricular hours; iv) developing a School Improvement Plan (PME), which includes actions in four areas of management; v) meeting student performance goals, especially for priority students; vi) indicating the amount of subsidies or public resources received annually; vii) informing parents and guardians about the existence of the agreement; viii) ensuring that classroom teachers present an annual educational plan for curricular content at the beginning of the school year; and ix) providing opportunities for extra-curricular activities for the students.

and implement their PME, provided directly by the Ministry of Education or by hiring registered external agencies, known as ATE (*Asistencias Técnicas Educativas*). These private agencies provide consulting, training, evaluation, and institutional diagnostic services.[6]

Under the SEP accountability system, schools are ranked into three categories: i) *autonomous* (schools that systematically perform above national standards); *emerging* (schools that do not systematically perform above national standards); and *in recovery* (schools that systematically perform below national standards).[7]

The position in the ranking has consequences on how resources are transferred to schools, the degree of monitoring in the execution of the PME, and, eventually the revocation of the license to operate. Unlike *autonomous* schools, which receive directly the total amount SEP funding they are eligible for, *emerging* schools receive only 1/3 of their allocated SEP funding directly. The remaining funding is transferred only after developing the PME, and the continuation of these transfers is contingent on the correct implementation of the plan, monitored by the Ministry of Education. Similarly, for *in recovery* schools, the Ministry of Education transfers the SEP funds in monthly installments only after submission of the school's PME. The continuation of these transfers is contingent on the correct implementation of the plan.

Accountability pressures in the SEP system are greater for *in recovery* schools. If an *in recovery* school fails to improve performance and move to the *emerging* category within three years, the Ministry of Education will inform the school community and encourage families to consider another schooling option for their children, as well as facilitating transportation to another school. If the school remains in the *in recovery* category for four years, the Ministry revokes the school's license to operate and receive public funding.

In addition to these consequences of accountability, the SEP law establishes that information on the school ranking is to be made public every year, which is intended to influence parental preferences in school choice. Every parent can go online and find their child's school's SEP ranking, mean standardized test scores, teachers and parents´ public complaints, among other indicators. Being classified as a low-performing school could have a negative effect on teacher provision or future enrollments, thus affecting

[6] Until 2015 (prior to the enactment of Law No. 20845), municipal school administrators had to contract ATE services through public bidding, while private voucher school operators could enter into direct contracts. Since 2015, all ATE contracts, both with municipal and private voucher schools, have to go through public tender.

[7] Schools were not ranked in the recovery category during the first four years of the SEP Law. 2012 was the first year that schools were classified in this category.

the amount of resources the school receives through the per capita funding, and eventually, its ability to operate.

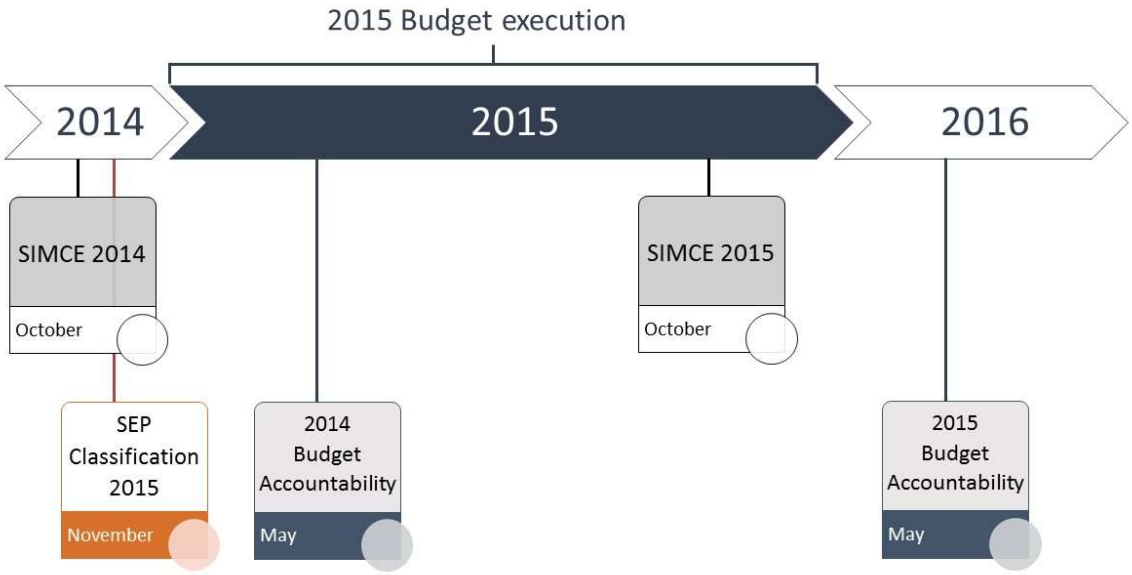
Thus, the SEP accountability system assumes that to avoid closure or negative signaling, *in recovery* school owners will include strategies in their PMEs to rapidly (and sustainably) improve student achievement. We hypothesize that *in recovery* schools will respond to the threats of accountability by altering their expenditure structures with the objective of improving student learning within a relatively short period of time. Low performing (*in recovery*) schools are faced with different options when deciding how to allocate their expenditures: (i) they may focus on long-term improvement, for example changing the composition of their staff to better respond to student needs or invest in human resource development, (ii) they may focus on hiring external pedagogical support that may produce short-term gains (whether sustainable or not), (iii) they may focus on purchasing additional inputs to help teachers in their work, such as ICTs or learning and pedagogical resources, or (iv) they may try to improve performance by altering the composition of their student body, for which they may focus on improving the school's image (through investments in school safety or publicity) or on policies such as school uniforms, meals and transportation. In this paper, we analyze the effects of the accountability pressures on school expenditure decisions. We evaluate the marginal impact of being classified as *in recovery* on school resource allocation decisions.

4. Methodology

Using 2014 budget execution as a baseline, this study captures the effect of the SEP classification 2015 results on budget execution shifts between 2014 and 2015. The schools received the results of the SEP classification in November 2014, which means that, while it is unlikely that they have time to change their 2014 budget decisions, they can respond to pressures by strategically allocating their 2015 resources. 2015 budget execution was reported to the School Inspection and Audit Agency by May 2016.[8] The timeline is shown in **Figure 1**.

[8] Budgets are established each calendar year, and each year's resources must disburse before December 31. The SEP accountability system uses data from the standardized assessments taken in October, and the SEP classification is made public every November.

Figure 1: Study Timeline



4.1. Data sources

We use three different databases from the three national school regulating agencies in Chile: the Ministry of Education (*Ministerio de Educación de Chile*), the School Inspection and Audit Agency (*Superintendencia de Educación)*, and the Agency for Quality Assurance in Education (*Agencia de la Calidad de la Educación*). Our main source of information is the National SEP Classification database, published by the Agency for Quality Assurance in Education since 2013. For 2015, this database contains the school rankings for the 7,460 SEP schools that were subject to be classified in 2015, along with all of the assignment variables that are used in the construction of the ranking. Each year the ranking depends primarily on the school's performance on the Chilean standardized test SIMCE (*Sistema de Medición de la Calidad de la Educación*) in the prior three years, for fourth-grade students. In addition to SIMCE scores and its distribution, the SEP classification considers a set of complementary indicators that measure other aspects of education quality. These indicators are: (i) grade approval and retention rates; (ii) teacher and family integration with the development of the school's educational project; (iii) school capacity to incorporate educational innovation; (iv) quality of working conditions; and (v) results on the national teacher evaluation in the case of public schools. Our second source of information is the National School Income and Expenditure

database, collected by the School Inspection and Audit Agency of Chile. This database contains detailed income and expenditure information for 11,472 schools in 2014 and 11,424 in 2015, including school payroll and private donations. Our third dataset is the official school information record collected by the Ministry of Education, which contains information on school characteristics, such as student enrollment, school location, school curriculum, ownership status, and socio-demographic data, among other data.

4.2. Identification strategy

For our identification strategy, we exploit the fact that the methodology used to rank Chilean schools in the SEP accountability system is based on a school's position relative to a multiple set of variables and their respective thresholds. These variables include national standardized test scores, the number of students tested, the number of available measurements, and a set of indicators that measure other quality dimensions (e.g. student retention rates, student pass rates or teacher evaluation results). The multidimensional characteristic of the accountability ranking allows us to use a multivariate regression-discontinuity design (MRDD), where a combination of cutoffs attained in a number of variables determines treatment status (unlike traditional RDDs, where units are assigned to treatment and control conditions based on a single cutoff score on a continuous variable. See for example Papay et al. 2011; Reardon & Robinson, 2012; Wong et al., 2013).

Methods to estimate average treatment effects with multiple assignment variables are based on regression models such as (Reardon & Robinson, 2012):

$$Y_i = m(R1_i, R2_i, \dots, Rn_i) + \sum_k \tau_k T_i^k + X_i B + e_i, \qquad (1)$$

where $\{R1_i, R2_i, \dots, RJ_i\} \in \boldsymbol{D} \subset \boldsymbol{R}$.

$R1_i, R2_i, \dots, Rn_i$ correspond to the n assignment variables and T_i^k is a dummy variable indicating if unit i is assigned to treatment k. The estimators of treatment effects τ_k differ in two important ways: i) the specification of the m function and ii) the $\boldsymbol{D}$ domain of observations used in estimating the model, which is a subset of the space formed by the n assignment variables ($\boldsymbol{R}$). The inclusion of pretreatment covariates (X_i) may increase the precision of the estimates, but is generally unnecessary, as the model is

well identified without it (Lee & Lemieux, 2010). The choice of the functional form of m may be important, especially when there are few observations near the frontier. In this case, it is necessary to use data further from the cutoff score and make assumptions about the functional form of the average potential outcome, but doing so increases the potential bias in the estimation. In other words, there is a trade-off between bias and precision.

Reardon and Robinson (2012) present five estimation methods: *response surface* RD, *frontier* RD, *fuzzy frontier* RD, *distance-based* RD, and *binding-score* RD. In this paper, we use the binding-score method because it has advantages over other approaches when there are a lower number of observations and when multiple rating scores determine assignment to only two treatment conditions. The main advantage of this approach is that it allows the researcher to parsimoniously collapse scores from multiple assignment rules into a single assignment variable and therefore can use all the observations simultaneously in the estimation. The approach also generalizes well to MRDDs with more than two assignment variables and simplifies the analyses for estimating average treatment effects across multiple discontinuity frontiers, but it requires the assumption that the average treatment effect is the same at each boundary. This method has been used, for example, in the evaluation of NCLB in the United States (e.g. Gill et al., 2009). Other examples are found in Reardon et al. (2010) and Robinson (2011). One disadvantage is that it does not allow the estimation of frontier-specific effects, so we cannot test the existence of heterogeneous treatment effects[9]. Another disadvantage is that pooling units from different frontiers increases the heterogeneity of the outcome at the pooled cutoff, requiring a larger bandwidth for nonparametric estimates and increases the complexity of the functional form around the cutoff (Wong et al., 2013).

The Binding Score method relies on the construction of a new assignment variable Z (*binding-score*) that sharply determines treatment assignment, so it assimilates to a traditional RDD. Let´s suppose that treatment assignment depends on two variables (R and M) and schools are assigned to a single treatment condition T if they score below both cutoffs ($R_i \leq r_c$ and $M_i \leq m_c$), and to the control condition C if they score above either cutoff. 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:

[9] Based on the variables forming the binding score.

$$Z_i = max(R_i^c, M_i^c), \quad (2)$$

where $R_i^c = R_i - r_c$ and $M_i^c = M_i - m_c$. By construction, $T_i = 1$ if $Z_i < 0$ and $T_i = 0$ if $Z_i \geq 0$

In this case, the problem becomes a traditional RDD and all the standard analytic methods can be used, defining Z as the assignment variable and zero as the cutoff. Although this transformation applies to the original assignment variables, Wong et al. (2013) show that this method estimates the same causal effect as alternative methods.

We use the SEP ranking database to construct the binding score of the RDD model. First, we consider the rules under which schools are ranked in the 2015 SEP database. For instance, according to the SEP Law, schools without SIMCE data for 2 or more years or with less than 20 students taking the national test in fourth grade are not ranked, and thus, are excluded from our analytical sample. Second, we determine our treatment and control groups. The treatment group includes schools that are ranked as *in recovery* in 2015. The control group contains those schools that are ranked as *emerging* or *autonomous* in the 2015 classification (*non-recovery*). Third, we limit the school universe to those schools that could be classified in 2015, and that have reported their income and expenses information to the School Inspection and Audit Agency for the years 2014 and 2015. The total sample includes 62 (2,39%) *in recovery* schools and 2.534 (97,61%) *non-recovery* schools for 2015.

Our binding-score (Z_{i201}) is constructed from this final dataset by using the seven rating scores that determine assignment to the *in recovery* category. Details on the construction of the binding score variable are presented in Appendix 1. Table 1 presents the treatment and control group details along with the number of times schools have been classified as *in recovery*.

Table 1: Autonomous, Emergent and in Recovery Schools

	Autonomous	Emergent	In Recovery	Total
Treatment & Control				
SEP Classification 2015	1,049	1,485	62	2,596
	40.4%	57.2%	2.4%	100.0%
In Recovery History				
Never in Recovery	1,040	1,359	0	2,399
	99.1%	91.5%	0.0%	92.4%
In Recovery once	6	72	19	97
	0.6%	4.8%	30.6%	3.7%
In Recovery twice	3	45	12	60
	0.3%	3.0%	19.4%	2.3%
In Recovery three times	0	9	17	26
	0.0%	0.6%	27.4%	1.0%
In Recovery four times	0	0	14	14
	0.0%	0.0%	22.6%	0.5%

Source: Ministerio de Educación de Chile and authors´ calculations

4.3. Variables

To analyze school expenditure structures, we use data reported by schools for 2014 and 2015 captured at the National School Income and Expenditure database. We group expenditures into eight categories, each of which is used as an outcome in our regression models by analyzing the difference in percentages of school total expenditures allocated to the category in 2015, using the year 2014 as a baseline. Table 2 describes the following outcomes:

Payroll expenses refers to the percentage of school total expenditures devoted to teacher and class assistant's payroll, along with all related expenses from hiring and retiring processes, like nursery school expenses or social security and retirement funds.

Teacher training and PME refers to the percentage of school total expenditures focused on quality improvements. These expenditures include teachers' participation in professional development courses, seminars, or coaching; educational software; and external support for the development, implementation, and evaluation of the school's improvement plan (PME). In Chile, this external support is mostly provided by ATEs. In general, the work of the ATEs focuses on providing training for school leadership teams,

assessing school improvement needs, providing advice on the development and implementation of the PMEs, etc. School administrators (private owners in the case of private schools, or municipalities in the case of public schools) can choose any ATE from a national registry of certified agencies. ATEs vary widely in the quality of the services they provide, and there is no specific guidance on the selection of ATE services.

Pedagogical equipment expenses refer to the percentage of total expenditures devoted to purchasing technological aides for pedagogical activities, such as computers, interactive whiteboards, etc.

Learning resources expenses refers to the percentage of total expenditures devoted to the acquisition of school pedagogical inputs such as school libraries, laboratories, evaluation tools, teacher guides, etc.

School transportation refers to the percentage of total expenditures focused on school transportation, which may include the hiring of external school transportation services or the purchasing of school buses. This is an important expenditure item in the case of Chile given that schools receive government subsidies based on student attendance, rather than on student enrollment.

School uniforms refers to the percentage of total expenditures used for school uniforms and clothing accessories like aprons. This expenditure item is optional for schools because the use of uniforms is not mandatory in Chile. Moreover, many of the schools ask the student's family to purchase uniforms. Offering them as a benefit could help retain or attract students.

Table 2: Descriptive Statistics for sample

Variable	2014 Baseline	USD 2014 (2)	Var Mean	Std Dev	N
Outcomes (1)					
School payroll expenses (%)	76.59%	$938,103	1.05%	8.38%	2,596
Teacher training and PME expenses (%)	1.60%	$19,597	-0.44%	2.11%	2,596
Pedagogical equipment expenses (%)	1.38%	$16,903	-0.16%	1.86%	2,596
Learning resources expenses (%)	2.57%	$31,478	-0.11%	2.42%	2,596
Transportation expenses (%)	1.08%	$13,228	-0.10%	1.38%	2,596
School uniform expenses (%)	0.25%	$3,062	0.03%	0.66%	2,596
School safety expenses (%)	0.11%	$1,347	0.00%	0.33%	2,596
Publicity expenses (%)	0.02%	$245	-0.01%	0.08%	2,596

Source: National School Directory of the Education Ministry of Chile and National Income and Expenditure database for 2015, collected by the School Audit Agency of Chile.

(1) Percentages are calculated over Annual School Income (Including Public, Private, and Donations transfers)

(2) USD are calculated over mean exchange rate 2014

School safety refers to the percentage of total expenditures invested in safety measures for the school. This could include security guards, alarm systems and/or cameras among other items. Security has been an increasing concern for families, and our hypothesis is that low performing schools could decide to invest in this category in order to be more competitive without having to substantially change staff and alter management and pedagogical practices.

Finally*, publicity* refers to the percentage of total expenditures invested in advertising. This could include web page development, publicity campaigns, leaflets, and school open houses among others. This category may be important for schools to improve their image in a competitive market.

Table 3: Descriptive Statistics for school characteristics

Variable	2014 Baseline	Std Dev	N
School Characteristics			
Public schools (%)	49.7%		2,596
Rural schools (%)	7.4%		2,596
Adult education (%)	8.6%		2,596
Special needs education (%)	2.6%		2,596
Students enrollment	590.7	392.4	2,596
SEP enrollment (%)	56.8%	17.8%	2,596
Number of classrooms	18.6	9.9	2,596
Number of teachers	34.8	17.1	2,596
Contracted Hours Teachers 2014	1221.8	646.6	2,596
SEP Contracted Hours Teachers 2014	44.9	72.9	2,596
Number of Assistants	22.5	13.25	2,596
Contracted Hours Assistants 2014	866.7	531.8	2,596
SEP Contracted Hours Assistants 2014	121.7	179.7	2,596
Student attendance 2014	83.8%	6.5%	2,594
Student SEP attendance 2014	83.2%	9.9%	2,592
Free disposable income (%)	53.8%	9.9%	2,596

Source: National School Directory of the Education Ministry of Chile and National Income and Expenditure database for 2015, collected by the School Audit Agency of Chile.

In addition to our outcome variables, we analyze several school characteristics. These include dummies indicating whether the school is a private voucher school or a municipal-public school, or whether the school is located in a rural area. We also analyze variables indicating the number of students enrolled in the school, the percentage of students that are *priority* students (low SES), the number of classrooms, the number and working hours of teachers and assistants, student attendance, and a variable indicating the proportion of free disposable income the school has (that is, the amount of income that is not earmarked for a specific

expenditure category) among many other variables. We present a summary of the descriptive statistics for our treatment and control groups in Table 3.

To test whether our treatment and control groups differ significantly in any of these school characteristics, we conduct a difference in proportion and a difference in means tests that compares the differences in each of these variables between *in recovery* (treatment) and *non-recovery* (control) schools within the 0.3 bandwidth. Table 4 indicates evidence of statistically significant differences on percentage of SEP enrollment (71.3% vs. 64%), where schools *in recovery* have a greater proportion of low SES students. No other significant differences were found between the groups. Beyond statistical significance, the data shows that (in means) *in recovery* schools are smaller in enrollment, and therefore, tend to have fewer classrooms and teachers. For income related variables, *in recovery* schools have lower attendance[10] but almost equivalent free disposable income.

Table 4: Testing differences in Recovery and Non-Recovery School groups characteristics

	Recovering		Non-Recovering		Difference in means / Proportion test	
Domain/Variable	Mean	N	Mean	N	t / z	p value
Group Characteristics						
% of Public Schools	69.6%	46	71.2%	118	0.205	0.838
% of Rural Schools	2.2%	46	5.1%	118	0.828	0.407
% of Adult Education	10.9%	46	14.4%	118	0.597	0.550
% of Special (Disable) Education	6.5%	46	2.5%	118	-1.219	0.223
School Characteristics						
Mean of student enrollment	421.04	46	447.42	118	0.588	0.557
Mean of % SEP enrollment	70.5%	46	65.2%	118	-2.123	0.035*
Mean of classrooms	14.98	46	15.46	118	0.374	0.709
Mean of Teachers	29.78	46	30.32	118	0.227	0.821
Mean of Contracted Hours Teachers 2014	1,041.70	46	1,047.08	118	0.060	0.952
Mean of SEP Contracted Hours Teachers 2014	62.61	46	40.78	118	-1.885	0.061
Mean of Assistants	19.41	46	19.58	118	0.101	0.919
Mean of Contracted Hours Assistants 2014	762.000	46	753.695	118	-0.126	0.900
Mean of SEP Contracted Hours Assistants 2014	111.15	46	78.03	118	-1.524	0.130
Mean of attendance 2014	75.3%	46	77.0%	118	1.421	0.157
Mean of SEP attendance 2014	76.4%	46	77.7%	118	1.063	0.289
Mean of Free disposable Income 2015	46.9%	46	49.1%	118	1.163	0.246

* p<0.05 ** p<0.01 *** p<0.001

Source: Authors' calculations based on National School Directory and National School Income and Expense Database for 2015

[10] As explained before, the voucher income system relies on student attendance.

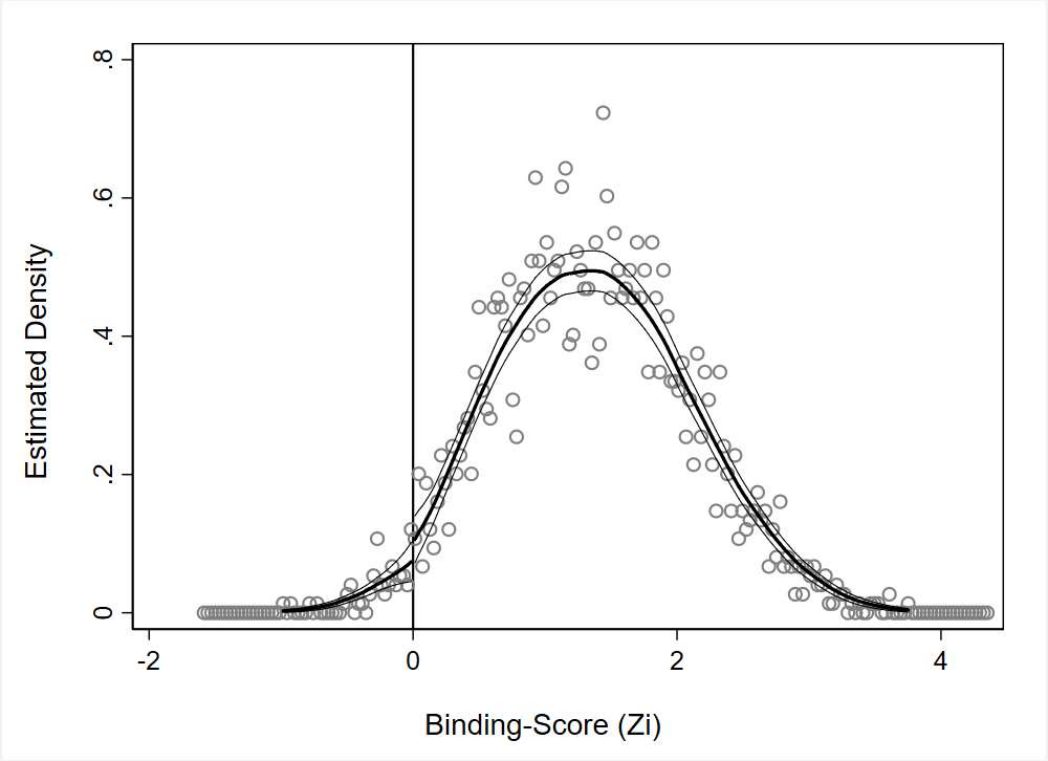
4.4. RD Validity

A key assumption of regression discontinuity analyses is that no agent can manipulate the assignment variable, thus falling on either side of the threshold could be considered random. While it is likely that schools would be motivated to score above the cutoff that places them *in recovery* status, it is unlikely that they can manipulate their ranking. To corroborate this empirically, we explore a standard group of tests for manipulation of the assignment. First, we plot the density function of the binding scores.

Figure ***2*** demonstrates that there is no jump in the density after the cut point of zero. We find no evidence of bunching near the cutoff that could suggest assignment variable manipulation. Along with the two-step procedure test proposed by McCrary (2008) for discontinuity, the second stage estimates 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. Table 5 presents the McCrary test of discontinuity and confirms the graphical evidence displayed in

Figure ***2***. These tests fail to reject the null hypothesis of no discontinuity in our binding score and in every component. In sum, both the smoothness of the assignment variable's distribution and the group and school covariate balance verify the causal assumptions of the RD design.

Figure 2: *McCrary Test for 2015*



Finally, Lee and Lemieux (2010) maintain that researchers should test the continuity of the baseline covariates as an important part of assessing the validity of an RD design. We test for discontinuities in pre-existing school´s characteristic prior to the classification, among them school SEP enrollment, student attendance and free disposable income, and found no discontinuities[11].

Table 5: McCrary Test

Variable	t	p value
Main		
Binding Score 2015	1.076	0.282
Detail		
$psimce_{2013}$	0.368	0.713
$psimce_{2012}$	0.410	0.682
$psimce_{2011}$	0.471	0.638
$p250_{2013}$	-0.595	0.552
$p250_{2012}$	0.459	0.646
$p250_{2011}$	0.518	0.604
Education Quality Index (ICE)$_{2015}$	1.447	0.148

* p<0.05 ** p<0.01 *** p<0.001

Note: Each variable is centered on its respective cutoff and divided by its standard deviation

Source: Authors' calculations

5. Results

We report regression discontinuity estimates of the effect of the 2015 *in recovery* classification on the decisions about school expenditure allocation across spending categories, within a bandwidth of 0.3 standard deviations relative to the binding score of zero that determined treatment status. We present the results with and without preexisting covariates in the following model that includes interactions with past SEP classifications:

$$Y = \alpha + \tau_1 T + \beta_1 (Z - z_c) + \beta_2 T(Z - z_c) + \tau_2 T_{-1} + \tau_3 T T_{-1} + XB + \varepsilon \qquad (3)$$

Where Y is the expense outcome, T takes the value of one if the school is classified as *in recovery* in 2015 and zero for non-*recovery*, T_{-1} is one if the school was classified as *in recovery* in 2012, 2013 or 2014 and

[11] Tests and regressions available upon request

zero otherwise, $(Z - z_c)$ represents the distance from the school to the threshold of our assignment variable constructed with the binding-score method, X represents the covariates, and ε is an error term with a normal distribution. We include as covariates a dummy for municipal-public schools (given management differences between public and private voucher schools) and the percentage of SEP enrollment at school.

Table 6 presents the regression results for being classified as an *in recovery* school in the 2015 SEP classification, taking into account whether this was the first time the school was classified as *in recovery*, or if it had been previously ranked in this category. The first three rows for each expenditure category show results for our model without covariates, and the next three rows include them. For each expenditure category and model, table 6 reports the effect of being *in recovery* 2015, of being *in recovery* at least one other time, and the aggregation of these effects. For this last result, significance is calculated through a Wald test where the null hypothesis that the sum of the effects is equal to zero is tested.

One of our main findings, that is consistent in subsequent analyses, is that being ranked as *in recovery* for the first time does not appear to change the way schools allocate their resources. For our eight expenditure outcomes there are no significant changes in the percentage of budget allocation to each spending category once the schools were publicly classified as *in recovery* in 2015. Nonetheless, schools that were previously classified as *in recovery* in either 2012, 2013 or 2014 responded to accountability pressures by investing in strategies to improve learning. The results show these schools are more likely to invest in teacher training programs and hire external support for class or school management (expenses contained in our "Teacher Training & PME" outcome). For the model without covariates, the budget shift is close to 1.6% of total expenditures and about 1.8% for the model with covariates.

On the other hand, there is a significant negative effect on school payroll expenses when the school has been ranked more than once as *in recovery*. However, in contrast to the impact found for "Teacher Training & PME", there is no significant joint effect of being classified as *in recovery* in 2015. Along with these results, we do not find any other statistically significant budget allocation change at the school level. The results suggest that *in recovery* schools do not react by investing in other quality measures such as pedagogical equipment or learning resources; or in student wellbeing measures, such as transportation or clothing. Finally, *in recovery* schools also do not appear to increase spending on safety and publicity measures.

Table 6: RD Estimates of Treatment Effect on School budget Allocation Decisions across expenditure categories *Model for 2014 - 2015 difference* (0.3 Sd Bandwidth) [(1)]

	School Payroll	Teacher training & PME	Pedagogical equipment	Learning resources	Transportation	Clothing	Safety	Publicity
No covariates								
In recovery 2015	0.00645	-0.000384	0.00131	-0.00635	0.00443	-0.000447	0.00116	0.000125
	(0.0264)	(0.00933)	(0.00626)	(0.00956)	(0.00529)	(0.00274)	(0.00200)	(0.000250)
In recovery 2015 and before	-0.0454*	0.0163*	-0.00284	-0.00188	-0.00371	0.0000994	0.00180	-0.000276
	(0.0263)	(0.00929)	(0.00624)	(0.00952)	(0.00527)	(0.00273)	(0.00199)	(0.000249)
Total Effect [(3)]	-0.03895	0.015916*	-0.00153	-0.00823	0.00072	-0.0003476	0.00296	-0.000151
With covariate [(2)]								
In recovery 2015	0.00402	0.00248	0.00155	-0.00430	0.00478	-0.000276	0.000842	0.000168
	(0.0267)	(0.00922)	(0.00634)	(0.00953)	(0.00535)	(0.00277)	(0.00202)	(0.000248)
In recovery 2015 and before	-0.0449*	0.0155*	-0.00292	-0.00250	-0.00383	0.0000369	0.00188	-0.000292
	(0.0264)	(0.00912)	(0.00628)	(0.00943)	(0.00529)	(0.00274)	(0.00199)	(0.000246)
Total Effect [(3)]	-0.04088	0.01798**	-0.00137	-0.0068	0.00095	-0.0002391	0.002722	-0.000124
N	164	164	164	164	164	164	164	164

Standard errors in parentheses

* p<0.10 ** p<0.05 *** p<0.01

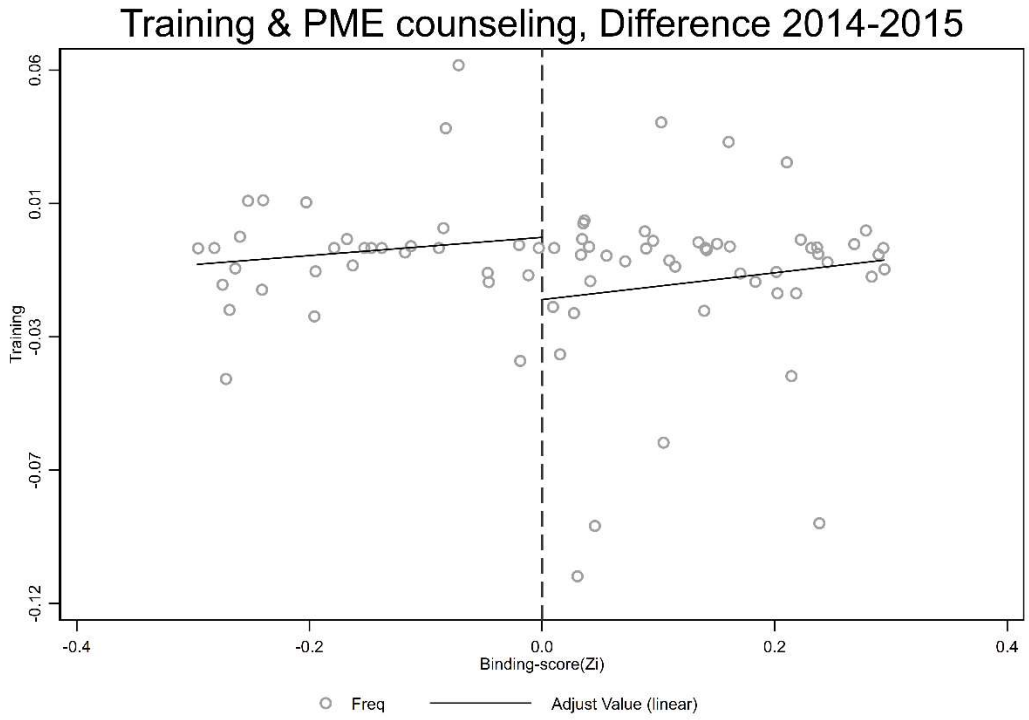
(1) Outcomes are 2014 - 2015 difference expenses in percentages over Annual School Income (Including Public, Private, and Donations transfers).

(2) The regression with covariate included dummies for Public School and % SEP enrollment

(3) Effect sum between "in Recovery 2015" and "in Recovery 2015 and before". Significance Wald test for the hypothesis that the sum of both effects is 0.

Figure 3 illustrates the relationship between different SEP classifications at the threshold (Z=0) and our statistically significant main finding within a bandwidth of 0.3 standard deviations. It shows the binned raw data along with the estimated regression line for a simple linear RD specification using only schools previously classified as *in recovery* between 2012 and 2014. Therefore, the figure captures partly the essence of our interactive exercise.

***Figure 3**:* Teacher Training and PME expenses adjustment between 2014 and 2015



5.1. Robustness checks

We test the robustness of our findings by using four different methods: *Alternative functional form, outlier removal, alternative bandwidth RDs, and use of false positives*.

The first exercise entails incorporating higher order terms of the assignment variable (Z). We present polynomial order 2 regressions in table 7. These regressions are consistent with our linear results, with

significant results in teacher training and external counseling expenses. The effect in this robustness exercise is higher than in our chosen specification (close to 2%).

Our second robustness check consists of eliminating common outliers from the regressions. There is a persistent discussion in the literature on the handling of outliers and their consequences on findings (Cortina 2002, Aguinis et al 2013). Even though we present our main results including outliers, we test here if our findings are driven by observations that could be considered outliers. For this analysis we identify three outliers among the different outcomes in the 0.3 SD bandwidth. In order to maintain consistency between outcomes, we take these observations out of our sample, so that we keep the same number of observations for the eight regressions we run (one per spending category). Table 8 presents the results without outliers for the models with and without covariates. The main results are neither affected in magnitude nor in significance.

Our third robustness check tests our findings using alternative bandwidths. Table 9 shows our results for local linear regressions using subsamples of schools defined by bandwidths from 0.2 SD to 0.4 SD from the threshold. The results indicate that, regardless of the chosen bandwidth, the schools decide to invest in teacher training and external consultancies once they are classified repeatedly as *in recovery*. Alternative bandwidths are evaluated for other expenditure variables, not reported here, and show patterns consistent with the previously reported findings.

Finally, our fourth robustness check tests for jumps at non-discontinuity points of our model. Imbens and Lemieux (2008) specify this approach of "testing zero effect in settings where it is known that the effect should be zero". We implement this by testing for jumps in two subsamples within the *non-recovery* schools, at a 0.3 SD and 0.5 SD distance from the original cutoff. Table 10 shows that there are no significant jumps in the outcome variables studied.

Table 7: Robustness check with polynomial order 2 RD . *Model for 2014 - 2015 difference* (0.3 Sd Bandwidth) [1]

Polynomial order 2 [2]	**School Payroll**	**Teacher training & PME**	**Pedagogical equipment**	**Learning resources**	**Transportation**	**Clothing**	**Safety**	**Publicity**
No covariates								
In recovery 2015	-0.000522	0.00619	-0.00874	-0.0102	0.00428	-0.000878	0.000118	0.0000739
	(0.0352)	(0.0124)	(0.00826)	(0.0127)	(0.00699)	(0.00366)	(0.00266)	(0.000332)
In recovery 2015 and before	-0.0450*	0.0157*	-0.00321	-0.00199	-0.00308	0.00000213	0.00191	-0.000296
	(0.0265)	(0.00930)	(0.00623)	(0.00960)	(0.00527)	(0.00275)	(0.00200)	(0.000250)
Total Effect [4]	-0.045522	0.02189*	-0.01195	-0.01219	0.0012	-0.00087587	0.002028	-0.0002221
With covariates [3]								
In recovery 2015	-0.00144	0.00664	-0.00869	-0.00998	0.00430	-0.000889	0.0000477	0.0000706
	(0.0354)	(0.0121)	(0.00831)	(0.0126)	(0.00702)	(0.00367)	(0.00266)	(0.000328)
In recovery 2015 and before	-0.0446*	0.0149	-0.00333	-0.00263	-0.00322	-0.0000622	0.00198	-0.000312
	(0.0266)	(0.00915)	(0.00626)	(0.00950)	(0.00529)	(0.00277)	(0.00201)	(0.000247)
Total Effect [4]	-0.04604	0.02154*	-0.01202	-0.01261	0.00108	-0.0009512	0.0020277	-0.0002414
N	164	164	164	164	164	164	164	164

Standard errors in parentheses

* p<0.10 ** p<0.05 *** p<0.01

(1) Outcomes are 2014 - 2015 difference expenses in percentages over Annual School Income (Including Public, Private, and Donations transfers).

(2) Consider a polynomial order two in the assignment variable Z

(3) The regression with covariate included dummies for Public School and % SEP enrollment

(4) Effect sum between "in Recovery 2015" and "in Recovery 2015 and before". Significance Wald test for the hypothesis that the sum of both effects is 0.

Table 8: Robustness check of RD without outliers. *Model for 2014 - 2015 difference* (0.3 Sd Bandwidth) [(1)]

Without outliers	**School Payroll**	**Teacher training & PME**	**Pedagogical equipment**	**Learning resources**	**Transportation**	**Clothing**	**Safety**	**Publicity**
No covariates								
In recovery 2015	0.0118	-0.000800	0.000717	-0.00823	0.000401	0.00000901	0.00116	0.000127
	(0.0261)	(0.00945)	(0.00631)	(0.00875)	(0.00408)	(0.00275)	(0.000958)	(0.000253)
In recovery 2015 and before	-0.0509*	0.0168*	-0.00236	-0.000367	-0.00159	-0.000180	0.000638	-0.000283
	(0.0259)	(0.00940)	(0.00628)	(0.00870)	(0.00406)	(0.00273)	(0.000953)	(0.000252)
Total Effect [(3)]	-0.0391	0.016*	-0.001643	-0.008597	-0.001189	-0.00017099	0.001798*	-0.000156
With covariates [(2)]								
In recovery 2015	0.0110	0.00261	0.000813	-0.00636	-0.000316	0.000377	0.000929	0.000186
	(0.0265)	(0.00937)	(0.00642)	(0.00881)	(0.00413)	(0.00278)	(0.000965)	(0.000252)
In recovery 2015 and before	-0.0509*	0.0156*	-0.00241	-0.00108	-0.00140	-0.000329	0.000705	-0.000310
	(0.0261)	(0.00924)	(0.00632)	(0.00868)	(0.00407)	(0.00274)	(0.000952)	(0.000248)
Total Effect [(3)]	-0.0399	0.01821**	-0.001597	-0.00744	-0.001716	0.000048	0.001634*	-0.000124
N	161	161	161	161	161	161	161	161

Standard errors in parentheses

* p<0.10 ** p<0.05 *** p<0.01

(1) Outcomes are 2014 - 2015 difference expenses in percentages over Annual School Income (Including Public, Private, and Donations transfers).

(2) The regression with covariate included dummies for Public School and % SEP enrollment

(3) Effect sum between "in Recovery 2015" and "in Recovery 2015 and before". Significance Wald test for the hypothesis that the sum of both effects is 0.

Table 9: RD Estimates using alternative bandwidths. Model for 2014 - 2015 difference[1]

	School Payroll			Teacher training & PME		
Bandwidth	0.2 Sd	0.3 Sd	0.4 Sd	0.2 Sd	0.3 Sd	0.4 Sd
No covariates						
In recovery 2015	0.00591	0.00645	0.00370	0.00451	-0.000384	0.00458
	(0.0329)	(0.0264)	(0.0252)	(0.0107)	(0.00933)	(0.00847)
In recovery 2015 and before	-0.0446	-0.0454*	-0.0379	0.0189*	0.0163*	0.0119
	(0.0332)	(0.0263)	(0.0245)	(0.0108)	(0.00929)	(0.00825)
Total Effect [3]	-0.03869	-0.03895	-0.0342	0.02341**	0.015916*	0.01648**
With covariates [2]						
In recovery 2015	0.00454	0.00402	0.00351	0.00505	0.00248	0.00751
	(0.0331)	(0.0267)	(0.0254)	(0.0105)	(0.00922)	(0.00834)
In recovery 2015 and before	-0.0460	-0.0449*	-0.0399	0.0206*	0.0155*	0.0103
	(0.0334)	(0.0264)	(0.0247)	(0.0107)	(0.00912)	(0.00811)
Total Effect [3]	-0.04146	-0.04088	-0.03639	0.02565**	0.01798**	0.01781**
N	101	164	231	101	164	231

Standard errors in parentheses

* p<0.10 ** p<0.05 *** p<0.01

(1) Outcomes are 2014 - 2015 difference expenses in percentages over Annual School Income (Including Public, Private, and Donations transfers).

(2) The regression with covariate included dummies for Public School and % SEP enrollment

(3) Effect sum between "in Recovery 2015" and "in Recovery 2015 and before". Significance Wald test for the hypothesis that the sum of both effects is 0.

Table 10: Robustness Check with false treatment schools. Model for 2014 - 2015 difference[1]

	School Payroll		Teacher training & PME	
Treatment Threshold (Z)	Z = 0.3 Sd	Z = 0.5 Sd	Z = 0.3 Sd	Z = 0.5 Sd
No covariates				
In recovery 2015	-0.0264	0.0332	0.00152	-0.0120
	(0.0330)	(0.0464)	(0.0113)	(0.0139)
In recovery 2015 and before	0.0115	-0.0110	-0.00952	0.00504
	(0.0192)	(0.0219)	(0.00653)	(0.00654)
Total Effect [3]	-0.0149	0.0222	-0.008	-0.00696
With covariates [2]				
In recovery 2015	-0.0273	0.0322	0.00331	-0.0121
	(0.0331)	(0.0464)	(0.0113)	(0.0139)
In recovery 2015 and before	0.00997	-0.00759	-0.00987	0.00540
	(0.0192)	(0.0220)	(0.00654)	(0.00659)
Total Effect [3]	-0.01733	0.02461	-0.00656	-0.0067
N	343	480	343	480

Standard errors in parentheses

* p<0.10 ** p<0.05 *** p<0.01

(1) Outcomes are 2014 - 2015 difference expenses in percentages over Annual School Income (Including Public, Private, and Donations transfers).

(2) The regression with covariate included dummies for Public School and % SEP enrollment

(3) Effect sum between "in Recovery 2015" and "in Recovery 2015 and before". Significance Wald test for the hypothesis that the sum of both effects is 0.

6. Discussion

One of today's most controversial topics in education reform discussions is school accountability. Advocates argue that schools under accountability pressure have strong incentives to adjust internal practices and policies to improve student performance. Critics have countered that accountability pressures also produce undesirable effects such as teaching to the test, altering the composition of the testing pool, the overemphasis of tested material, and cheating by teachers. Accountability opponents have also argued that low performing schools will tend to focus on quick solutions that generate rapid improvements (e.g. test taking strategies) rather than on educational investments that produce longer term gains (e.g. teacher development). Skeptics are also concerned that, faced with accountability pressures, schools will have incentives to undertake "glitzy" reforms that focus on publicity and improving the school's image. This is especially relevant in systems of school choice where parents may easily choose to exit low performing schools.

While scholars have developed a substantial body of empirical research that has examined the effects of accountability on student achievement and school "gaming" of accountability incentives (e.g. Figlio & Loeb, 2011), there has been little attention paid to changes in school resource allocation resulting from school accountability. This is an important oversight since there is evidence that some resources are more likely to improve student performance than others. For example, there is a growing evidence that effective teachers can dramatically improve student achievement (e.g. Araujo et al., 2016). This dearth of research is mainly due to the lack of school level expenditure data. Our study seeks to contribute to this debate by analyzing a unique school expenditure data set in Chile, coupled with administrative data in a school accountability system.

Our results indicate that low-performing schools respond timidly to the accountability pressures generated by the SEP Law. The findings show that, despite high fixed costs (teacher salaries, facilities, etc.), *in recovery* schools strategically focus a larger share of their variable spending on certain expenditures, *but only when the accountability pressure is high*. First-time *in recovery* schools do not show a resource allocation pattern different from similar schools just above the threshold. It is only when low-performing schools are ranked *in recovery* more than once that they change their resource allocation strategy, being more likely to allocate resources to professional development and to external technical assistance than similar schools just above the threshold and reducing expenses on teacher payroll.

Thus, recurrent low-performing schools, compared to their counterfactual, are more likely to focus spending on items that may potentially be linked to quality improvements. School administrators under systematic pressure are focusing resources on measures such as professional development and external technical assistance that aim to improve student performance in the short or medium-term. Interestingly, we also find that *in recovery* schools are not more likely than *emerging* schools to focus resources on inputs such as learning and pedagogical resources, student well-being (transportation and uniforms) and inputs that may improve the school's image with parents such as publicity or school security. Recurrent low-performing schools seem to be responding in a way that is consistent with the design of the SEP accountability system.

First-time low performing schools, in contrast, are not reacting to the accountability pressures. Whether this slow reaction is due to poor management capacity, or a perception of low risk of closing due to a single low performance classification is beyond the scope of this paper. Nevertheless, this finding highlights the importance of the design of accountability systems, because the deadlines, types of sanctions, communication strategy, and the assumptions made about the school improvement process are key in determining how schools under threat will target their resources. The literature on school improvement emphasizes the fact that low performing schools do not improve overnight; they take sometimes years to boost achievement levels. The slow reaction of first-time *in recovery* schools in the case of Chile calls attention to the need to critically analyze the design of accountability systems to ensure that these schools are not losing valuable time doing more of the same. Carefully targeted external support programs for first time low performing schools that recommend more effective resource allocation may be among the policy options to support these schools in this process.

References

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. Journal of Management, 39(6), 1490-1528.

Araujo, M.C., P. Carneiro, Y. Cruz & N. Schady (2016) Teacher quality and learning outcomes in kindergarten. *Quarterly Journal of Economics*, March.

Booher-Jennings, J. (2005). Below the Bubble: "Educational Triage" and the Texas Accountability System. American Educational Research Journal, 42(2), 231-268.

Murnane, R., Waldman, M., Willett, J., Bos, M.S., & Vegas, E. (2017) The consequences of educational voucher reform in Chile, IDB working paper series N. 833.

Brady, R. (2003). Can Failing Schools be fixed? Washington, DC: Thomas B. Fordham Foundation.

Cortina, J. M. (2002). Big things have small beginnings: An assortment of “minor” methodological misunderstandings. Journal of Management, 28(3), 339-362.

Dee, T., & Dizon-Ross, E. (2017). School performance, accountability and waiver reforms: Evidence from Louisiana (No. w23463). National Bureau of Economic Research.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. Journal of Policy Analysis and management, 30(3), 418-446.

Deming D., S. Cohodes, J. Jennings & C. Jencks (2016) School Accountability, Postsecondary Attainment, and Earnings, The Review of Economics and Statistics, MIT Press, vol. 98(5), pages 848-862,

Elacqua, G., M. Martinez, H. Santos & 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 Vol. 27, Iss. 3, 2016

Elmore, R. F., Abelman, C. H., & Fuhrman, S. H. (1996). The new accountability in state education reform: From process to performance. In H. F. Ladd (Ed.), Holding schools accountable: Performance-based reform in education (pp. 65-98). Washington, DC: Brookings Institution.

Figlio, D. N. (2006). Testing, crime and punishment. Journal of Public Economics, 90, 837-851.

Figlio, D. N., & Getzler, L. S. (2006). Accountability, Ability and Disability: Gaming the System? In T.J. Gronberg, & D.W. Jansen (Eds.), Improving School Accountability (Advances in Applied Microeconomics, Volume 14) (pp.35-49). Bingley, UK: Emerald Group Publishing Limited

Figlio, D.N., & Loeb, S. (2011). School Accountability. In E. A.Hanushek, S. Machin, & L. Woessmann (Eds.), Economics of Education (pp. 383-421). The Netherlands: North Holland.

Figlio, D. N., & Lucas, M. E. (2004). What's in a Grade? School Report Cards and the Housing Market. The American Economic Review, 94, 591-604.

Gill, B., Lockwood, J., Martorell, F., Setodji, C. M., Booker, K., Vernez, G., Birman, B., & Garet, M. S. (2009). An Exploratory Analysis of Adequate Yearly Progress, Identification for Improvement, and Student Achievement in Two States and Three Cities. (Technical Report U.S. Department of Education).

Goldhaber, D., & Hannaway, J. (2004). Accountability with a kicker: Observations on the Florida A+ accountability plan. Phi Delta Kappan, 85, 598-605.

Hamilton, L., Stecher, B., Marsh, J., McCombs, J., Robyn, A., Russel, J. N., & Barney, H. (2007). Implementing standards-based accountability under No Child Left Behind: Responses of superintendents, principals, and teachers in three states. Santa Monica, CA: RAND.

Hannaway, J., & Cohodes, S. (2007). Trouble Even in Choice Paradise: NCLB in Miami-Dade County Public Schools. In F. M. Hess, & C. E. Finn (Eds.), No remedy left behind: Lessons from a half-decade of NCLB (pp. 244-264). Washington DC: American Enterprise Institute Press.

Hannaway, J., & Hamilton, L. (2008). Performance-based accountability policies: Implications for school and classroom practices. Washington, DC: Urban Institute and RAND Corporation. Retrieved from http://www.urban.org/sites/default/files/alfresco/publication-pdfs/411779-Accountability-Policies.PDF

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management, 24(2), 297-327.

Hirschman, A. O. (1970). Exit, voice, and loyalty: Responses to decline in firms, organizations, and states. Cambridge, MA: Harvard University Press.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. Journal of econometrics, 142(2), 615-635.

Jacob, B. A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in Chicago Public Schools. Journal of Public Economics, 89, 761-796.

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. The Quarterly Journal of Economics, 118, 843-877.

Koretz, D. (2017) The testing charade. Pretending to make schools better. Chicago: University of Chicago Press.

Koretz, D. M., & Barron, S. I. (1998). The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS).

Ladd, H. F., & Zelli, A. (2002). School-based accountability in North Carolina: The responses of school principals. Educational Administration Quarterly, 38, 494-529.

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. Journal of Economic Literature, 48, 281–355

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics, 142, 698-714.

O'Day, J. (2002). Complexity, Accountability, and School Improvement. Harvard Educational Review, 72, 293-329.

Papay, J. P., Willett, J. B., & Murnane, R. J. (2011). Extending the regression-discontinuity approach to multiple assignment variables. Journal of Econometrics, 161, 203-207.

Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers. Boston, MA: National Board on Education Testing and Public Po

Reardon, S. F., Arshan, N., Atteberry, A., & Kurlaender, M. (2010). Effects of failing a high school exit exam on course taking, achievement, persistence, and graduation. Educational Evaluation and Policy Analysis, 32, 498–520. doi:10.3102/0162373710382655

Reardon, S. F, & Robinson, J. (2012). Regression discontinuity designs with multiple rating-score variables. Journal of Research on Educational Effectiveness, 5, 83–104. doi:10.1080/19345747.2011.609583

Robinson, J. P. (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, 267–292. doi:10.3102/0162373711407912

Rockoff, J., & Turner, L. (2010). Short-run impacts of accountability on school quality. American Economic Journal: Economic Policy, 2, 119–147. doi:10.1257/pol.2.4.119

Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. American Economic Journal: Economic Policy, 5(2), 251-81.

Smarick, A. (2010). The turnaround fallacy. Education Next, 10, 20–26. Retrieved from http://educationnext.org/theturnaround-fallacy/

Stecher, B. M., & Barron, S. (2001). Unintended consequences of test-based accountability when testing in "milepost" grades. Educational Assessment, 7(4), 259-281.

Stecher, B. M., Barron, S. L., Chun, T., & Ross, K. (2000). The effects of the Washington state education reform on schools and classrooms. Santa Monica, CA: RAND Corporation.

Wong, V., Steiner, P., & Cook, T. (2013). Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four methods. Journal of Educational and Behavioral Statistics, 38, 107–141. doi:10.3102/1076998611432172

Appendix 1. Construction of the Binding Score

Table A1 shows the seven variables that define if a school is classified as *Recovery* or *Not In-Recovery* for 2015. Using these variables, we are able to establish a unique continuous rating score (Z_{iy}) that determine the year's final school classification. In order to have all variables on a same scale, each variable was centered on the respective predefine cutoff and then divided by their standard deviation. For 2015 we transform its first variable $psimce_{2013}$ to $psimce^{ZC}_{2013}$:

$$psimce^{ZC}_{2013} = \frac{psimce_{2013} - 220}{\sigma_{psimce_{2013}}}$$

For 2015, we construct the first rule to be pre-classified as Recovery, which is that in two years the school average SIMCE score must be under the cutoff value, and that less than 20% of their students reach a higher score of 250. Therefore, we calculate first the maximum between $psimce$ and $p250$ for each year:

$$Z^1_{2015} = \max\left(psimce^{ZC}_{2013}, p250^{ZC}_{2013}\right)$$

$$Z^2_{2015} = \max\left(psimce^{ZC}_{2012}, p250^{ZC}_{2012}\right)$$

$$Z^3_{2015} = \max\left(psimce^{ZC}_{2011}, p250^{ZC}_{2011}\right)$$

So, as the variables are centered to their respective cutoff, if the school doesn´t meet any of the two rules in a year, the constructed value will be negative $Z^i_{2015} < 0$. To capture the "two bad year" rule, we build a fourth binding value that takes the second maximum between Z^1_{2015}, Z^2_{2015} and Z^3_{2015}. For 2015 is:

$$Z^4_{2015} = second\max\left(Z^1_{2015}, Z^2_{2015}, Z^3_{2015}\right)$$

Thus, Z^4_{2015} indicates if a school is classified as recovery according to SIMCE results. If $Z^4_y < 0$ then $psimce_{yt} < 220$ and $p250_{yt} < 0.2$ in two years, and therefore the school meets the requirements to be pre-classified as recovery. The opposite is true when $Z^4_y \geq 0$.

The final classification rule incorporates the Education Quality Index. If a school`s Index is below the 10th percentile, then it will be classified as Recovery, thus:

$$Z_{2015} = min\left(Z^4_{2015}, ICE^{ZC}_{2015}\right)$$

This variable (binding-score) perfectly determines treatment assignment. If $Z_i < 0$, i school is classified as Recovery. If $Z_i \geq 0$ it will be classified as Non-Recovery.

Table A1: Variables that define if a school is classified as Recovery or Non-Recovery in year 2015

Variable	Description	Cutoff
Binding Score$_y$		
psimce$_{2013}$	School average SIMCE fourth grade year 2013 score	220
psimce$_{2012}$	School average SIMCE fourth grade year 2012 score	220
psimce$_{2011}$	School average SIMCE fourth grade year 2011 score	220
p250$_{2013}$	School average proportion of students who have scored over 250 points in SIMCE fourth grade year 2013 score	20%
p250$_{2012}$	School average proportion of students who have scored over 250 points in SIMCE fourth grade year 2012 score	20%
p250$_{2011}$	School average proportion of students who have scored over 250 points in SIMCE fourth grade year 2011 score	20%
Education Quality Index (ICE)$_{2015}$	Index that combines average SIMCE score of previous 3 years[1] (70%) with complementary indicators[2] (30%)	10th percentile

Notes: The SIMCE variables takes scores in Math, Language and Science tests.

[1] Average of years 2013, 2012 and 2011

[2] Complementary indicators are: student´s approval and retention rates, teacher`s and family involvement is school project, school´s educational innovation, teacher´s working conditions, and public teacher`s evaluation.