FACULTAD ECONOMÍA Y NEGOCIOS
UNIVERSIDAD DE CHILE
DEPARTAMENTO DE ECONOMÍA

SERIE DE DOCUMENTOS DE TRABAJO

SDT 513

The Impact of Grade Retention on Juvenile Crime

Autores:
Juan Díaz
Nicolás Grau
Tatiana Reyes
Jorge Rivera

Santiago, Enero de 2021

sdt@econ.uchile.cl
econ.uchile.cl/publicaciones

THE IMPACT OF GRADE RETENTION ON JUVENILE CRIME

Juan Díaz* Nicolás Grau† Tatiana Reyes‡ Jorge Rivera†

This version: January, 2021

Abstract

This paper studies the causal effect of grade retention in primary school on juvenile crime in Chile. Implementing a fuzzy regression discontinuity design, we find that repeating an early grade in primary school decreases the probability of committing a crime as a juvenile by 14.5 percentage points. By estimating and simulating a dynamic model, we show that the RD result is mainly driven by two mechanisms related to the timing of grade retention. First, grade retention in early grades decreases the probability of grade retention in late primary school grades. Second, late grade retention in primary education has a positive and more relevant effect on crime than the direct effect in early grades. Our findings support the argument that, conditional on the decision to keep grade retention as an ongoing policy, the optimal implementation at the margin is to retain students in early grades in order to avoid retention in later ones.

Keywords: Juvenile Crime, Grade Retention, Regression Discontinuity. **JEL Classification: I21 and K42**.

We thank Tomás Cortés and Isidora Lara for excellent research assistance. We also thank Aureo De Paula, Claudio Ferraz, Guido Imbens, José Zubizarreta, Miguel Urquiola, Zelda Brutti, as well as seminar participants at Universidad de Chile, Sao Paulo School of Economics, Universidad Diego Portales, the Workshop on Human Capital (RIDGE Forum 2016), LACEA-LAMES 2016, Workshop on Applied Economics of Education (Catanzaro, 2016), and the 2016 North American Summer Meeting of the Econometric Society, for valuable comments and suggestions. We thank the staff at the Public Defender's Office and Chile's Ministry of Education. Jorge Rivera thanks the Complex Engineering Systems Institute, ISCI (ICM-FIC: P05-004-F, CONICYT: FB0816) partial financial support. Juan Díaz and Nicolás Grau thank the Centre for Social Conflict and Cohesion Studies (ANID/FONDAP/15130009) for financial support. All remaining errors are our own.

*Department of Management Control and Information Systems, Faculty of Economics and Business, University of Chile. E-mail: juadiaz@fen.uchile.cl.

†Department of Economics, Faculty of Economics and Business, University of Chile. E-mails: ngrau@fen.uchile.cl and jrivera@econ.uchile.cl.

‡Department of Economics, University of California, Berkeley. E-mail: tatiana_reyes@berkeley.edu.

1 Introduction

Does grade retention in school increase the likelihood that young people will engage in criminal activity? From an opportunity cost point of view, it seems reasonable to argue that students may be more prone to pursue non-educational activities if they are not promoted to the next grade (Lochner (2004)). Conversely, repeating a grade might strengthen knowledge and improve discipline with potential positive effects on a particular student's outcomes. Thus, instead of representing a "cost" for students, grade retention could be viewed as an "opportunity" that may help a young person become more competitive in the classroom and discourage divergence to non-educational activities (Jacob (2005)). The latter scenario is particularly relevant if early grade progression for students at the margin of minimum learning requirements can increase the probability of grade retention in the future. The ambiguity surrounding the potential effect of grade retention on crime is at the core of the well-known controversy relating to grade retention[1]

Settling this issue in an empirical way is particularly important in developing countries where the rates of both grade repetition and juvenile crime are much higher than those observed in developed countries. In 2012 the average rate of grade retention in primary education was 5.1% in developing countries but 1.4% in developed countries (UNESCO Institute for Statistics). In Chile, the average rate of grade retention, although below the average for developing countries, has been increasing over the last 20 years, rising from 3.1% to 3.8% between 1999 and 2012. Regarding crime levels in a broad sense, Chile has a higher incarceration rate than OECD countries, detaining 266 inmates per $100,000$ as opposed to 145.5.[2]

Despite the vast literature linking grade retention and juvenile crime,[3] evidence of

[1]The grade retention controversy exists because of ambiguous, and even contradictory, evidence regarding the effect that this measure has on some academic and socio-emotional outcomes of students, see Holmes et al. (1989) and Jimerson (2001); see also Reschly and Christenson (2013) for a fresh look at this subject.

[2]European Institute for Crime Prevention and Control, affiliated with the United Nations (2010).

[3]For instance, Burdick-Will (2013), Fagan and Pabon (1990), and Hirschfield (2009), among others, have shown how criminal activities affect some schooling outcomes, a sort of inverse of the problem studied here. The effect of compulsory schooling laws on crime has been investigated by Lochner and Moretti (2004), Brugård and Torberg (2013), and Machin et al. (2011), among others. Other contributions have investigated how school starting age may affect crime (see Landersø et al. (2016) and references therein).

a causal effect between them is scarce and, as discussed later in this section, does not exist for developing economies. One factor that may explain the lack of evidence is the difficulty in finding an adequate empirical setting and dataset to overcome the potential endogeneity produced by the fact that the latent outcome —namely, criminal activity that would be observed in the absence of grade retention —and the propensity to fail a grade are simultaneously determined.

To fill the gap, this paper estimates the causal effect of grade retention on juvenile crime using a regression discontinuity (RD) approach. More specifically, we rely on the discontinuity in the probability of grade retention generated by the most commonly applied rule used to determine grade retention decisions in Chile —namely, that the grade must be repeated when a student scores below 4 in two or more subjects and has an average score lower than 4.95 across all subjects.[4] Although students can be retained due to more than one rule, the conditions of the rule selected for this paper were fulfilled in 84% of the cases of total grade retentions in Chile in 2007 (the year considered in our estimation sample).

We use an exceptional database from Chile, which matches individual academic records for all students (1st to 12th grade) with youth and adult criminal prosecution information, also on an individual basis, between 2007 and 2019. Our estimation sample includes students who attended 2nd and 3rd grade in 2007 and had not previously been the subject of grade retention. We restrict the sample to 2007 because it is the earliest year we can observe the annual average score for all subjects taken by students, which is required to evaluate the implementation of the retention rule. Moreover, we focus our attention on early grades because we see no evidence of manipulation in terms of grade retention decisions (i.e., sorting around the threshold for grade retention).[5]

Our main finding is the robust evidence of a (local) negative causal effect of grade retention on juvenile crime. We implement the standard fuzzy RD procedure developed by Calonico et al. (2014b) and Calonico et al. (2019). We find that repeating a grade decreases the probability of committing a crime as a juvenile by 14.5 percentage points (pp) and by 10.7 pp in the case of a severe crime. Reassuringly, the results are not sensitive to the bandwidth choice or the implementation of higher order polynomials;

[4]For reference, the potential scores range from 1 to 7 (awarded in increments of 0.1).

[5]The 1st grade is not considered because Chilean law gives much more agency to teachers on grade retention decisions at this level.

we also find no effect when we use placebo cutoff values. Additionally, we examine the effect of grade retention on dropping out of school, finding that grade retention in 2nd or 3rd grade decreases the probability of dropping out by 31 pp.[6]

Given that we also find that retention during the early grades improves future grade point average (GPA) and decreases the probability of future grade retention, we complement the RD analysis by estimating a (semi-structural) dynamic model.[7] In this model, grade retention in the early grades can directly and indirectly (via future GPA and future grade retention) affect crime. By estimating this model we show that the results from our RD estimation are not driven by a direct and relevant negative effect of grade retention on crime, but they are driven by a combination of a negative effect of grade retention in early primary grades on grade retention in late primary grades, with an increasing impact of grade retention on crime as students progress through the primary grades. Thus, our findings support the idea that, conditional on the decision to keep grade retention as an ongoing policy, the best implementation of this policy for those students around the threshold of the retention rule is to be retained in early grades in order to avoid retention in later ones.[8]

There is a growing literature that examines test-based promotion policies on measures of academic performance and crime. Greene and Winters (2009), looking at 3rd grade students in Florida, show that retained students slightly outperformed students that were socially promoted (i.e., who should have been retained under the policy in place at the time). Jacob and Lefgren (2009), looking at students from Chicago, find a differential effect between grade retention in the 6th and 8th grades. Their results show a substantial increase in the probability of dropping out of high school if a student is subject to retention in the 8th grade. To the best of our knowledge, there are two papers more closely related to our investigation —namely, Eren et al. (2017) and Eren et al. (2018). Both estimate the impact of grade retention (after the offer of a summer school program) on juvenile and adult delinquency (as well as other outcomes) in Louisiana. They assemble a novel dataset after merging administrative information on educational outcomes with the criminal records of students attending schools in Louisiana. Then,

[6] The issues related to grade retention and school dropout have been investigated by Roderick (1994) and Manacorda (2012). See King et al. (2015) for a comprehensive literature review.

[7] Future GPA is defined as a student's average GPA between 4th and 8th grade.

[8] There are many other educational policies that can be useful to prevent juvenile crime. Given what we can learn from our model, here we only focus on the timing of the grade retention.

taking advantage of the test-based grade promotion policy, the authors build a fuzzy RD design where the forcing variable is a standardized test score that determines whether or not a student is promoted. Eren et al. (2017) conclude that there is no effect of this test-based grade retention policy for 4th grade students; for students attending the 8th grade the policy has a small negative impact at most. Eren et al. (2018) show that when looking at a longer period (i.e., criminal convictions until age 25) being retained in the 8th grade has large effects on the likelihood of being convicted of a crime (and the number of convictions). The results presented by Eren et al. (2018) are consistent with our results as well as other evidence showing that the effect of grade retention varies as a function of when students are retained (Ou and Reynolds (2010), Fruehwirth et al. (2016)). Our dynamic model addresses the differing results from the literature related to the impact of grade retention on academic performance and crime by stressing the crucial role of grade retention timing as well as the impact of the treatment on the probability of being treated in the future.

A similar study was carried out by Cook and Kang (2016) who merge administrative data for academic performance with the criminal records of students attending public schools in North Carolina. They exploit a sharp RD design generated by the specific date that establishes the minimum age for school enrollment ("the cut date") and assess its effect on a number of educational outcomes as well as on crime committed as a juvenile. They highlight two main findings. First, that middle school students born just after the cut date (i.e., the oldest children in each grade) are more likely to outperform (in mathematics and reading) those born just before the cut date (i.e., the youngest children in each grade), and the oldest children are also less likely to be involved in juvenile delinquency. Second, those children born just before the cut date are more likely to drop out of school and commit a severe offence. Finally, Depew and Eren (2016), exploiting the same discontinuity as in Eren et al. (2017). and using the same student data from Louisiana, find that delaying school entry by one year decreases the frequency of juvenile delinquency for young black females.

In the context of the existing literature this paper makes two main contributions. To the best of our knowledge, this is the first paper that studies the causal effect of grade retention on crime in a developing country. To the extent that grade retention and juvenile crime are much more prevalent in these countries, this is a relevant contribution.

Moreover, by estimating a dynamic model, we show that our finding of a negative effect of grade retention on crime, which is also present in the literature, can be explained by a dynamic effect of grade retention in early primary grades on the probability of grade retention in late primary grades. In other words, it is not due to a direct negative effect of grade retention on crime. As we discuss in the conclusion, this result may be useful in order to better understand the heterogeneity in the results observed in the literature and, hence, it may be very relevant for policy design.

This paper is organized as follows: In Section 2 we describe the institutional background and the main features of both the educational and criminal datasets, and we present the evidence regarding the discontinuity created by the retention rule. In Section 3 we present our empirical strategy and study its validity. In Section 4 we outline our main results. Section 5 introduces the dynamic model and discusses its simulation results. And, finally, Section 6 concludes.

2 Institutional background, data, and retention rules

In this section, we describe the Chilean school system and juvenile criminal justice system, then we describe the characteristics of our dataset. Finally, we explain how the grade retention rule operates, which is critical in order to understand the source of the exogenous variation in our empirical strategy.

2.1 School system

In Chile, primary education lasts from the 1st to the 8th grade and secondary education from the 9th to the 12th grade.[9] Primary education has a unified curriculum that consists of a set of minimum subjects; in order to progress to the next grade, a student must attain a certain level of academic knowledge. The Ministry of Education provides guidelines for grade retention. The guidelines state that a student should be retained if their GPA or attendance rate falls below certain level (described below). In this paper we study the effect of grade retention at 2nd or 3rd grade on juvenile crime. We select these particular grades for study because in later grades there is evidence of scores manipulation (see Solis (2017)), which raises questions about the utilization of a regression discontinuity

[9]Secondary education (or until age 21) became mandatory in 2003.

approach, and because the law gives more agency to teachers regarding grade retention decisions at the first grade level.[10]

Each grade in primary education is comprised of approximately 250,000 students who can attend public, private subsidized, or private unsubsidized schools. The first two types of school account for 93% of the total enrollment (in similar proportions) and receive the same per student subsidy.[11] In this paper we consider students from all school types.

Dropout in Chile is low compared with Latin-American countries or the US. According to official statistics for 2012, it was 3.7% and it dropped to 2.2% in 2019. In Chile, most dropout happens before completing secondary education and towards the beginning of high school. In our dataset, we define dropout as one when the student does not enroll in 12th grade until 3 years after his/her expected graduation date. This definition overestimates dropout because a student who was enrolled in 2nd grade may not be enrolled until 12th grade for several reasons: migration, change to alternative types of education like adult education, educational lag bigger than three years, among others. There are no reasons to expect this measurement error to be non-randomly distributed in our population. However, all things considered, the analysis and quantitative interpretation of the dropout results should be taken with caution.

To measure the learning process, knowledge acquisition, and school performance of students there is a system of national standardized testing (SIMCE) in which all students in 4th grades must participate.[12] The government uses the results from the SIMCE tests to allocate resources and to inform the public about the quality of schools by listing school-level results in major newspapers. Since all schools, including public schools, are funded on the basis of a per student formula there is significant pressure to produce good SIMCE test results. This, in turn, creates an impetus for selecting and expelling students as well as for increasing grade retention in order to improve academic performance.

[10]The law establishes that the school principal in concordance with the academic coordinator and the teaching staff (after their consultation) can waiver the attendance condition. Additionally, if the student does not fulfill the requirements to be promoted, or fails an important subject, the school has to evaluate the reasons and social and emotional condition of the student. They would need to make the case, with all the relevant documentation, to proceed in a different way than what the rule requires.

[11]More details about Chilean education system can be found in Gauri (1999) and Grau et al. (2018).

[12]See Meckes and Carrasco (2010) for details.

2.2 Juvenile criminal justice system

The juvenile criminal justice system in Chile was reformed in 2005 (Act N^o 20084) and came into effect in 2007. Inspired by the United Nations Convention on the Rights of the Child, it is based on the principles of an exceptional and moderate application of criminal law and the use of confinement only as *ultima ratio* (see Langer and Lillo (2014)). This reform made three major changes to the previous system. It reduced the age of criminal responsibility from 16 to 14. It ended the ambiguity of the previous system whereby adolescents could be treated as adults or juveniles depending on the considerations of the judge. And, for convicted juvenile defendants, it reduced the punishment by one grade relative to the corresponding adult sentence.[13] Furthermore, the new juvenile criminal justice system was implemented as Chile was undertaking a radical reform of its criminal justice system as a whole, which began in 2000 and was completed in 2005. This broad reform replaced the inquisitorial model, a written system that had been in place for more than a century, with an oral, public, and adversarial procedure.[14] As part of the reform, several new institutions were created including the Public defender's office (PDO) and the Public Prosecutor's Office. The PDO provides free legal representation to almost all individuals who have been accused of committing a crime, and it collects information on all defendants that use their services, both juveniles and adults, which includes detailed information on the particular crime in question. Our data on juvenile criminal activity comes from PDO records.

In this paper we measure crime, our dependent variable, as being prosecuted. We consider two types of crimes: *all crimes*, an indicator variable that takes the value of one when the juvenile was prosecuted during the ten years following 2007 (the year that defines treatment);[15] and *severe crimes*, an indicator variable that takes the value of one when the juvenile was prosecuted for a severe crime during the period already described. Following previous literature (see Cortés et al. (2020)), we define severe crime as a type of crime for which the pretrial detention rate is greater than 3%. Table 7 (Appendix A) presents the distribution of juvenile crimes across different types of crime. Figure 6 (Appendix A) shows, among repeaters and non repeaters at early primary school grades,

[13] See Couso and Duce (2013) for a detailed description of this reform.

[14] See Blanco et al. (2004) for a detailed description of the criminal justice system reform in Chile.

[15] Because our estimation sample is comprised of students with different ages (i.e., we consider 2nd and 3rd grade students), we follow those students who attended 2nd grade in 2007 for one more year.

the fraction of students prosecuted for the first time at different ages.

2.3 Data

We assemble our administrative dataset using data from the Ministry of Education and the PDO. For youths not legally represented by a PDO attorney (i.e., they have a private attorney), we observe the alleged crime but we do not observe the final verdict. That said, less than 3% of prosecuted youths in our dataset are represented by a private attorney. In this paper, we use PDO records for juvenile criminal cases prosecuted during the period between January, 2008 and December, 2018.

The information collected from the Ministry of Education is an administrative panel dataset for every student in Chile between 2002 and 2019. The dataset indicates the school attended each year, the grade level (and whether the grade was repeated), the student's attendance rate, some basic demographic information, and (for 2007 only) the student's annual average score for each subject (cumulative GPA).[16] The cumulative GPA for each subject is critical information in the context of our RD approach because it is needed to build a more continuous measurement for the average across all subjects.[17] From this panel we build the other dependent variables considered in this paper: future grade retention, defined as a student's average GPA between 4th and 8th grade, and dropout. The latter is defined as permanently absent without graduation from 12th grade. We merge this panel with the data from the SIMCE test, which is taken annually by all 4th grade students. When students take the SIMCE test a survey is administered to their parents. From these surveys, we obtain information about both parent's education level and family income.

2.3.1 Retention rules and discontinuity

In Chilean primary education students are taught around 10 subjects per year and are scored between 1 and 7 for each subject with an increment of 0.1. In this context, there

[16] To distinguish between two distinct uses of the word "grade" —that is, between a level and an academic performance —grade performance will, hereafter, be referred to as "score".

[17] For other years, we only have the average across all subjects officially reported by the Ministry of Education. The problem with this measurement is that it is approximated and, consequently, using this level of aggregation for our estimation would mean comparing students with an average of 4.4 with students with an average of 4.5, for example.

are three important rules for grade retention. Students have the right to progress to the next grade, unless: (1) their attendance rate is below 85%; (2) they score below 4 for one subject and have an average score across all subjects lower than 4.45; or (3) they score below 4 on two or more subjects and have an average score across all subjects lower than 4.95. In 2007 489,168 students attended the 2nd and 3rd grades (the two grades included in our estimation sample) and 20,309 repeated the grade. Of those students, 2,634 (13%) were retained after scoring above 4 in all the subjects (probably due to a low attendance rate), 548 (2.7%) repeated the grade after scoring below 4 in only one subject, and 17,127 (84.3%) were retained after scoring below 4 in two or more subjects. Although students can be retained due to more than one rule, given the distribution of cases the most relevant is last rule described, where the threshold is 4.95 and the condition to be applied is scoring below 4 in at least two subjects. Therefore, in this paper we exploit the discontinuity of treatment probability around a GPA of 4.95 as exogenous variation in the probability of grade retention.[18]

Given our research question and the characteristics of the selected retention rule, the estimation sample has the following characteristics. We focus our attention on the students who attended the 2nd and 3rd grade in 2007. As stated, we restrict our sample to those grades because for later grades we observe some evidence of manipulation in grading decision around the 4.95 threshold (for more detail regarding this manipulation see Solis (2017)) and because the law gives more agency to teachers on grade retention decisions at the first grade level. For obvious reasons, we only consider students affected by the aforementioned retention rule —namely, those scoring below 4 in two or more subjects. In order to exclude schools where no student scores below the threshold, we only consider schools where at least one student scores less than 4.95 on average across all subjects. Finally, we focus on students who had not previously repeated a grade. In the last section of the paper we develop and estimate a dynamic model that, besides helping us understand our results, allows us to understand the effect of more than one

[18]The school principal, in concordance with the academic coordinator and the teacher staff (after their consultation), can waiver the attendance condition. Additionally, if a student does not fulfil the requirements to be promoted, or fails an important subject, the school has to evaluate the reasons for this academic performance and social and emotional condition of the student. They would need to make the case, with all the relevant documentation, to proceed in a different way respect to the rules. This is a costly process. That said, we also present the reduced form estimations (i.e., the sharp RD estimation), which are not affected by this discretion.

grade retention.[19]

2.3.2 Estimation sample

The final dataset includes $13,072$ students. The overall impact of sample restrictions are observed in Table 1, which shows how the estimation sample is different to the full sample (i.e., the population) in terms of a set of observables. As can be observed, in terms of individual characteristics the estimation sample has a greater percentage of males and students from less educated families, and the students perform less well at school, including a greater grade retention rate. The schools in the estimation sample report lower average standardized test scores and lower average education for both parents. Furthermore, the students in the estimation sample have a higher probability of committing crime in the future, approximately twice the probability of the population. The differences between the full and estimation sample demonstrate that by restricting our attention on students close to the retention cutoff, the estimation sample is comprised of low performance students. This is also observed in the table. The differences also emphasize that, as is usually the case when the causal effect is estimated using an RD empirical strategy, our results and their causal interpretation only have local validity.

[19] As our treatment is grade retention in 2007 we should exclude students who were prosecuted before 2008; however, the ages of students in our 2007 estimation sample means that this is not a binding restriction.

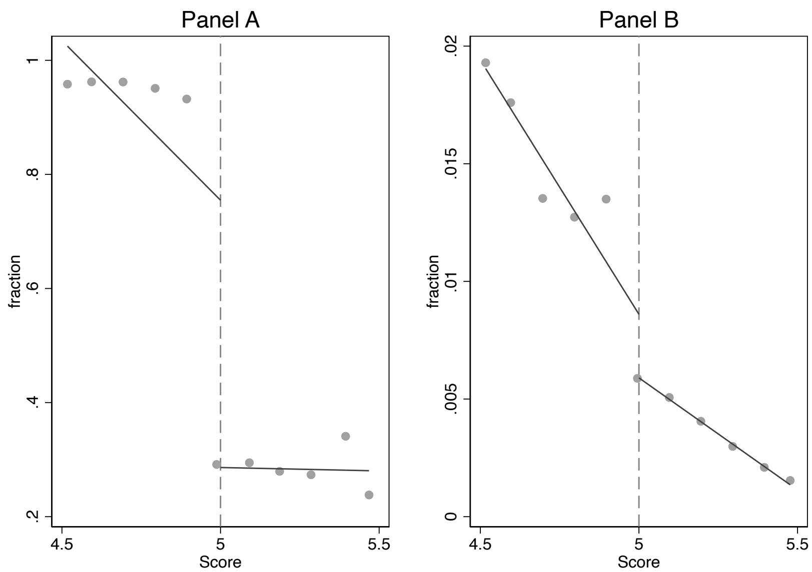
Table 1: Estimation sample versus full sample

Variables	Full Sample [$n = 489,168$] Mean	Std. Dev.	Estimation Sample [$n = 13,072$] Mean	Std. Dev.
Panel A. Individual Characteristics: Estimation Sample				
Male (%)	51.5%	50.0	60.1%	49.0
Father's Education	11.2	3.9	9.2	3.7
Mother's Education	11.0	3.7	9.1	3.6
Attendance (2006) (%)	93.1%	10.0	92.0%	6.6
GPA (2006)	6.1	0.7	5.2	0.5
Panel B. School Characteristics: Estimation Sample.				
Grade Retention Rate (last 3 years)	0.05	0.07	0.06	0.06
Average Math Standardized Score	0.0	0.6	-0.2	0.5
Average Verbal Standardized Score	0.0	0.5	-0.2	0.4
Average Income Decile	5.4	2.0	4.6	1.5
Average Father's Education	11.0	2.6	9.9	2.0
Average Mother's Education	11.2	2.3	10.2	1.9
Average Expectation of Child Education	15.4	1.7	14.7	1.5
Panel C. Outcome & Other Variables				
All Crimes (%)	6.6%	24.9	13.0%	33.7
Severe Crime (%)	5.2%	22.3	10.4%	30.6
Future Grade Retention (%)	33.8%	47.3	64.5%	47.9
Dropout (%)	24.4%	43.0	50.1%	50.0
GPA From 4th to 8th Grade	5.6	0.5	5.2	0.4
GPA (2007)	59.1	6.7	44.2	4.1
Grade Retention (2007)	0.04	0.20	0.93	0.26

Notes: The estimation sample considers students who attended 2nd or 3rd grade in 2007, who scored below 4 in two or more subjects, from schools with at least one student who scored less than 4.95 on average across all subjects, and who had not been subject to a previous grade retention. The full sample includes all the students who attended 2nd or 3rd grade in 2007.

To explore the discontinuity due to the aforementioned retention rule, panel (a) of Figure 1 presents the probability of grade retention around the 4.95 threshold for those students who belong to the estimation sample. The figure shows a discrete and relevant change in the probability around the threshold. More specifically, this probability increases by 61.7 pp for those individuals marginally below the 4.95 threshold. To show that the rule is only binding under its specific conditions, in panel (b) we show the same exercise only considering those students who score below 4 in one or no subjects. In these cases, the grade retention probability is continuous around the threshold.

Figure 1: First Stage



Notes: This figure presents a binscatter plot of the fraction of students in 2nd and 3rd grade not promoted in 2007, with a linear fit at each side of the cutoff. Panel A includes students from the estimation sample with two or more subjects below 4. Panel B includes students without two or more subjects below 4.

3 Empirical Approach

Exploiting the retention rule outlined in the previous section, we estimate the effect of grade retention on juvenile crime using the method to implement a fuzzy RD approach developed by Calonico et al. (2014b) and Calonico et al. (2019). In this context, the bandwidth selection is calculated by minimizing an approximation to the asymptotic mean squared error of the point estimator and removing the bias due to the curvature of the regression function.[20]

Let n denote the number of students in the sample. For individual i, let Z_i be the GPA score in 2007, the running variable in our application, whose cutoff level is denoted by $\bar{z}$ (which in our scenario is 4.95); and let W_i be the treatment indicator that takes the value one if the ith student repeats the grade in 2007; and let Y_i be the binary outcome that takes the value of one when the individual committed a crime as a juvenile and zero

[20]We implement this approach by using the Stata routines developed by Calonico et al. (2014a) (using updated code as of 2020).

otherwise. Finally, X_i is a set of covariates. Given an optimal bandwidth h we calculate the RD estimation, τ_{FRD}, as:

$$\widehat{\tau}_{FRD}(h) = \frac{\widehat{\tau}_Y(h)}{\widehat{\tau}_W(h)},$$

$$\widehat{\tau}_Y(h) = \widehat{\alpha}_{Y,-}(h) - \widehat{\alpha}_{Y,+}(h), \qquad \widehat{\tau}_W(h) = \widehat{\alpha}_{W,-}(h) - \widehat{\alpha}_{W,+}(h),$$

where, for $J = Y, W$, the estimators $\widehat{\alpha}_{J,-}$ and $\widehat{\alpha}_{J,+}$ come from a standard local linear RD estimator:

$$\begin{pmatrix} \hat{\alpha}_{J,-} \\ \hat{\alpha}_{J,+} \\ \hat{\beta}_{J,-} \\ \hat{\beta}_{J,+} \\ \hat{\gamma}_J \end{pmatrix} = \underset{\alpha_{J,-},\alpha_{J,+},\beta_{J,-},\beta_{J,+},\gamma_J}{\arg\min} \sum_{i=1}^{n} [J_i - \mathbb{1}(Z_i < \bar{z}) \cdot (\alpha_{J,-} + \beta_{J,-} \cdot (Z_i - \bar{z})) -$$

$$\mathbb{1}(Z_i \geq \bar{z}) \cdot (\alpha_{J,+} + \beta_{J,+} \cdot (Z_i - \bar{z})) - \gamma_J \cdot X_i]^2 \cdot \frac{K\left(\frac{Z_i - \bar{z}}{h}\right)}{h},$$

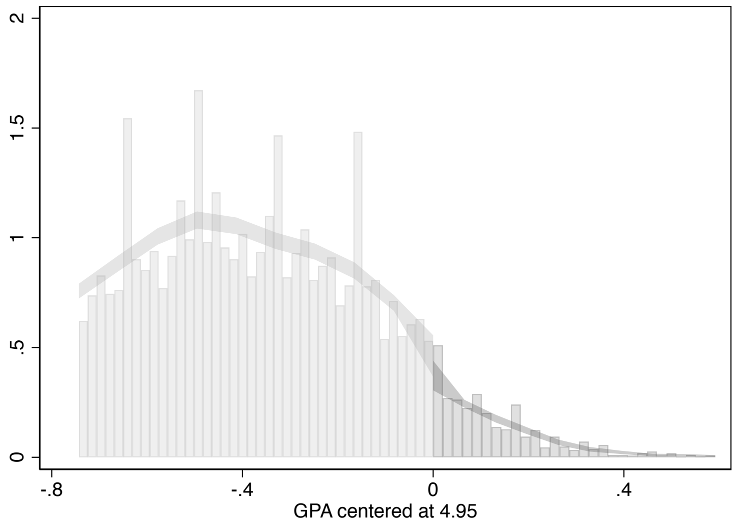
where $K(\cdot)$ is a kernel function. We cluster errors at school level.

3.1 Validity of the RD Design

We explore the validity of the RD design by performing the two most common tests for this purpose: (i) we study the continuity of the density of the running variable at the cutoff and (ii) we examine whether covariates (i.e., observed variables measured before 2007) are similar between estimation sample students who are below and above the cutoff.

For (i) we implement the density test developed by Cattaneo et al. (2017). In concrete terms, we assess whether the density of the GPA (the running variable in our application) is a continuous function at 4.95 (the cutoff). As shown in Figure 2, the test result reveals that the null hypothesis stating that the density of the GPA is a smooth function at 4.95 cannot be rejected with a value of the robust bias-corrected statistic of 1.096. This implies a p-value of 0.27. This finding provides evidence in favor of the validity of our RD design because it suggests that the GPA is not determined by strategic behavior or manipulation at the cutoff.

Figure 2: DENSITY TEST



Notes: The plot shows the density test proposed by Cattaneo et al. (2017). We implement this test using the Stata command rddensity. The value of the robust bias-corrected statistic of this test is equal to 1.096. This implies that we cannot reject the null hypothesis of continuity of the density function of the running variable at the cutoff (p-value equal to 0.27).

For (ii) we estimate the effect of being marginally below the cutoff on several covariates for students at the cutoff. Specifically, we consider 12 covariates in total, which are related to students' academic achievements and socio-economic backgrounds as well as certain characteristics of the students' schools.

As can be observed in Table 2, all covariates are similar between students who are marginally below and above the cutoff, except in the case of 2006 GPA where the difference between the two groups is statistically significant. Importantly, the significant difference between the two groups at the cutoff found in 1 of the 12 covariates can be explained by chance in a setting of multiple comparisons rather than by a systematic difference between the two groups. Given that students who are marginally below and above the cutoff are similar in covariates, these results provide evidence that supports the validity of our RD design as any difference in the observed outcome of interest (juvenile crime, dropout, or future grade retention) between the two groups of students at the cutoff can be attributed to the treatment (i.e., being retained in 2007).

Table 2: Differences in covariates at the cutoff for retention

Variable	RD Estimator	Robust Inference p-value	Robust Inference C.I.	Number of Observations
Panel A. Individual Characteristics				
Male	0.07	.4	[-0.09 0.24]	13,029
Father's Education	0.39	.59	[-1.37 2.40]	7,981
Mother's Education	0.45	.46	[-1.08 2.37]	8,331
Attendance (2006)	1.26	.12	[-0.45 3.68]	12,567
GPA (2006)	0.18	.01	[0.04 0.36]	12,567
Panel B. School Characteristics				
Grade Retention Rate (last 3 years)	0.00	.41	[-0.03 0.01]	12,889
Average Math Standardized Score	-0.03	.94	[-0.21 0.19]	12,815
Average Verbal Standardized Score	0.06	.48	[-0.12 0.24]	12,817
Average Income Decile	0.23	.22	[-0.19 0.80]	11,262
Average Father's Education	0.06	.64	[-0.53 0.87]	11,280
Average Mother's Education	0.30	.15	[-0.18 1.14]	11,281
Average Expectation of Child Education	0.14	.34	[-0.26 0.76]	11,271

Notes: The table presents the results based on the methods for estimation and inference of a sharp RD developed by Calonico et al. (2014b) and Calonico et al. (2019). The model is estimated without covariates, and the dependent variables are the 12 covariates presented in the first column. Differences in sample sizes are due to missing values for some of the covariates. In an ideal RD context, all the point estimates should not be statistically significant. The robust inference considers the bias term coming from the approximation error that does not vanish from the asymptotic distribution of the RD estimator.

4 Results

In this section, we present our findings on the impact of grade retention on juvenile crime, juvenile severe crime, and dropping out of school. These results are based on the methods for estimation, inference, and bandwidth selection for fuzzy RD designs developed by Calonico et al. (2014b) and Calonico et al. (2019). We also provide several robustness analyses that reinforce the validity of our RD design and, therefore, the plausibility of our findings.

4.1 Impact on crime, severe crime, and dropout

In Table 3, we present our estimations for the impact of grade retention on juvenile crime considering all crimes and severe crimes only, and dropout, including and not including covariates in the estimation. The covariates considered include attendance rate in 2006, GPA in 2006, gender, school characteristics, grade dummies, school grade retention rate in the previous three years, school average SIMCE score for mathematics, school average score for the SIMCE verbal test, school average years of education for both the father and mother, school average family income decile decile, and school average expectation of childhood education. All the estimations are presented with errors clustered at a school level. As seen in Table 3, we estimate that repeating a grade in 2007 decreases the probability of committing a crime as a juvenile by 14.5 pp when we do not include covariates and 17.7 pp when we include them. In the case of severe crimes, the effect of repeating a grade is -10.7 pp when we do not include covariates and -13.3 pp when we include them. All these effects are statistically significant at the 1% and 5% levels for all crimes and severe crime, respectively.

Table 3: Effect of grade retention on juvenile crime and dropout

	All Crimes		Severe Crime		Dropout	
	Without Covs. (1)	Including Covs. (2)	Without Covs. (3)	Including Covs. (4)	Without Covs. (5)	Including Covs. (6)
RD Estimator	-.144***	-.176***	-.107**	-.133**	-.306***	-.408***
	(.045)	(.045)	(.041)	(.043)	(.076)	(.078)
Mean Variable	.132		.106		.508	
Std. Dev. Variable	.339		.308		.5	
Robust Inference						
p-value	0.01	0.00	0.03	0.02	0.00	0.00
C.I.	[-.249 -.038]	[-.293 -.059]	[-.205 -.009]	[-.246 -.02]	[-.485 -.127]	[-.621 -.196]
Effective Obs.						
Left	1,611	1,356	1,425	1,200	1,612	1,357
Right	567	483	537	457	567	483
Optimal Bandwidth[a]	.168	.168	.161	.161	.171	.171

Notes: This table presents the results for the impact of grade retention on juvenile crime and dropout, based on the methods for estimation and inference for fuzzy RD designs developed by Calonico et al. (2014b) and Calonico et al. (2019). The reported means and standard deviations are control group statistics. The specifications (2), (4) and (6) include the following covariates: attendance rate in 2006, GPA in 2006, gender, school characteristics and grade dummies. The robust inference considers the bias term coming from the approximation error that does not vanish from the asymptotic distribution of the RD estimator. Standard errors are in parentheses and clustered at the school-level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

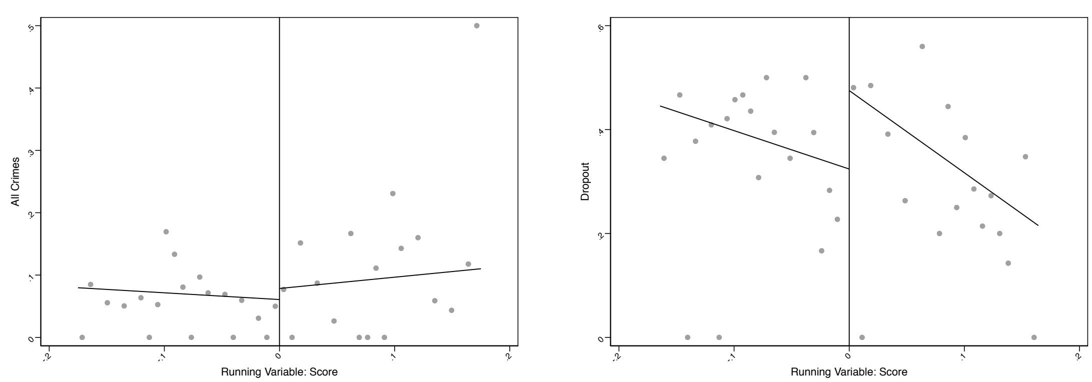
There are two aspects that must be stressed regarding these results. First, these are sizable effects. For the control group the crime rate is 13% for all crimes and 10.4% for severe crimes only. This means that the effect of grade retention on crime (without covariates) represents a decrease of 112%, whereas for severe crime the effect corresponds to a decrease of 102%. Second, although the effects are a little higher in absolute value and estimated with basically the same precision when we include covariates, the results are reasonably stable to the inclusion of additional control variables, which is consistent with the evidence shown in the previous section on covariates balance at the cutoff.

Given the richness of our panel dataset we can also examine the effect of grade retention on other outcomes. This allows us to present a more complete picture of what happens to the students' trajectories after they repeat the 2nd or 3rd grade. Specifically, for the probability of dropping out of school, we find that not being promoted to the next grade in 2007 decreases the probability of dropping out of school by 31.1 pp without covariates and by 41.1 pp with covariates. All these effects are statistically significant at the 1% level. To assess the size of these effects, it should be noted that 50.1% of the control group drop out of school at some point. It is remarkable, again, how stable the point estimates are to the inclusion of covariates in the estimation. And it is this stability that reinforces our confidence in the validity of the RD approach in this context.[21]

To present the RD results graphically, in Figure 3 we show the outcome values and an estimation of the regression functions via local linear regressions around the threshold all crimes and dropping out of school. As can be seen in Figure 3 in each plot there is a jump in the outcome variable at the cutoff for grade retention (at zero), which reinforces the plausibility of the findings presented in Table 3.[22]

Figure 3: Graphic results for crime and dropout

(a) All Crimes (b) Dropout



Notes: This figure shows the outcome values and an estimation of the regression functions via local linear regressions around the threshold for grade retention for all crimes and dropping out of school.

[21] Table 8 (Appendix B) presents the reduced form estimations (i.e., sharp RD), for all the dependent variables analyzed in this paper.

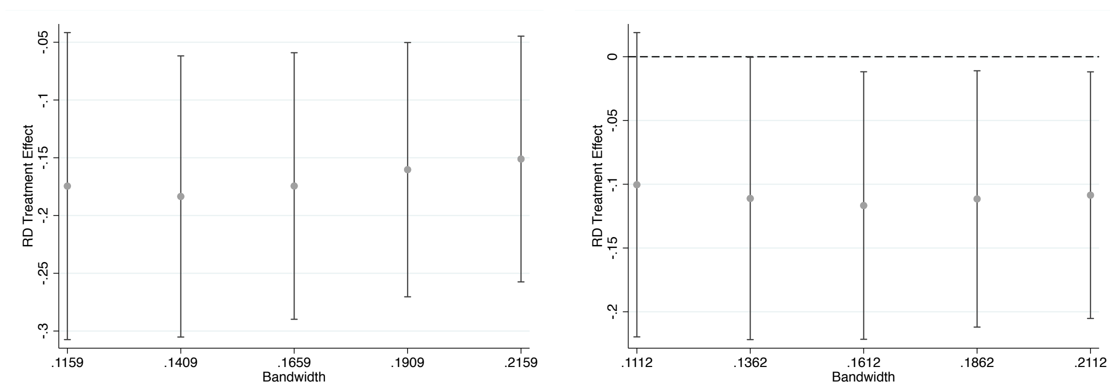
[22] Figure 7 (Appendix B) shows the same plot but for severe crimes.

4.2 Robustness Analysis

We begin our robustness check by studying the extent to which point estimates are sensitive to the bandwidth choice. In our main estimation, following Calonico et al. (2014a), this is calculated by minimizing an approximation to the asymptotic mean squared error of the point estimator and removing the bias due to the curvature of the regression functions. In Figure 4, we show our estimates for all crimes and severe crimes considering five possible bandwidths used in our main specification, with the middle bandwidth being the optimal value. These figures show that the estimation results regarding crime are not sensitive to bandwidth choice. Moreover, Figure 8 (Appendix C) shows that the dropout point estimate is also not sensitive to bandwidth choice.

Figure 4: Sensitivity to bandwidth: all crimes and severe crimes

(a) All Crimes (b) Severe Crimes



Notes: This figure shows the fuzzy RD estimations for the impact of grade retention on juvenile crime (using the methods developed by Calonico et al. (2019)), considering five different values of the bandwidth (the middle estimate is the optimal bandwidth). The point estimates are the dots and the confidence intervals at 95% are the brackets.

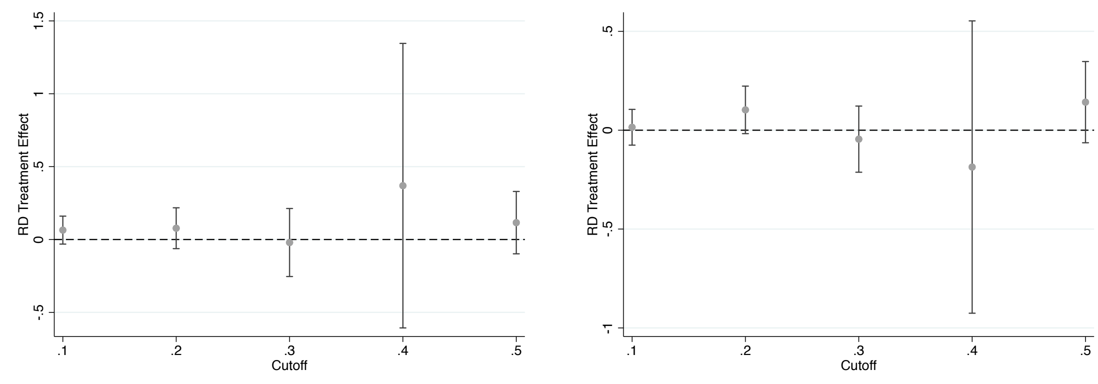
Our second robustness analysis performs estimation and inference of treatment effects in our RD setting but uses artificial (or placebo) cutoff values. Naturally, if our design is valid, we would expect that no significant treatment effects should appear at any artificial cutoff. In Figure 5 we present the results of the estimation, for all crimes and severe crimes, of the impact of being below the artificial cutoff considering five possible artificial cutoffs.[23] As expected, these figures show that the effects are not statistically

[23]Remember that the actual cutoff involved in the grade retention rule is 0 as we calculate it using the running variable centered at 4.95.

significant regardless of the artificial cutoff employed. For the case of dropping out of school, the effects are also not significant for distinct choices of placebo cutoffs, as can be observed in Figure 9 (Appendix C). These results reinforce the validity of our RD design.

Figure 5: Placebo tests: all crimes

(a) All Crimes (b) Severe Crimes



Note: This figure shows the sharp RD estimations for the impact of being below the cutoff for juvenile crime (following the method developed by Calonico et al. (2019)), for different values of cutoff. The point estimates are the dots and the confidence intervals at 95% are the brackets.

For our final robustness check we follow a global polynomial approach up to the fifth order. This approach makes assumptions about the global shape of the regression function at each side of the cutoff using all the observations. A global polynomial approach used to be more common; however, as Cattaneo et al. (2019) pointed out, it has been recognized that this approach does not deliver point estimators and inference procedures with good properties for the RD treatment effect. Acknowledging these limitations, we estimate five different models as a robustness exercise. In Table 4 we show the results from this approach. In short, the point estimates are consistent with our main results across the different specifications and most are statistically significant. Notwithstanding that for the three outcomes (all crimes, severe crimes, and dropout) the estimates are smaller than our preferred specification, they provide supporting evidence for the robustness of our results.

Table 4: Effect of grade retention on juvenile crime and dropout: global polynomials

	All Crimes		Severe Crimes		Drop out	
Order of polyn.	est.	(s.e.)	est.	(s.e.)	est.	(s.e.)
1	-.043	.026	-.033	.023	-.009	.04
2	-.06	.032	-.05	.028	-.043	.051
3	-.088	.041	-.072	.036	-.115	.064
4	-.066	.047	-.044	.042	-.189	.075
5	-.094	.053	-.059	.047	-.125	.086

Notes: This table presents the results for the impact of grade retention on juvenile crime and dropout, using the estimation sample described in Table 1 and following a global polynomial approach up to the fifth order.

5 Dynamics

Even though our RD empirical strategy delivers local and causal estimates for the effect of grade retention on juvenile crime, the results are not conclusive regarding the effect on juvenile crime of eliminating grade retention, not even for the compliers. In other words, the problem is not just the local nature of our estimates. There are two important reasons for this. On the one hand, grade retention can be an incentive to increase student academic effort, something that is not captured by our empirical approach. On the other hand, grade retention today can reduce the probability of grade retention in the future. Thus, students who belong to the control group in our RD design can be part of the treatment group in the future (in later grades). And to the extent that there is heterogeneous effect of grade retention on juvenile crime across grades, the negative effect that we find using a RD estimation is also consistent with positive effects of grade retention on juvenile crime in both the present (early primary grade) and the future (late primary grade), but where the effect in the future is larger than the effect in the present. While we do not have the data to study concerns relating to incentives, this section is devoted to addressing concerns about the dynamics.

We start this analysis by documenting the effect of grade retention in the 2nd or 3rd grade (2007) on future GPA, defined as a student's average GPA between 4th and 8th grade, and on future grade retention, defined as an indicator variable that takes the value of one if a student is subject to at least one grade retention between 2008 and 2014. In practice, this analysis involves running the same model specification that produced the results presented in Table 5 but changing the dependent variable from crime to future GPA or grade retention. Table 5 also shows that grade retention increases future GPA by 0.3 points (0.86 standard deviations) and decreases the probability of future grade retention by 43.2 pp. As the table shows, these point estimates are quantitatively relevant, statistically significant, and robust to the inclusion of covariates.

Table 5: EFFECT OF GRADE RETENTION ON EDUCATIONAL OUTCOMES

	Future Grade Retention		GPA From 4th to 8th Grade	
	Without Covs. (1)	Including Covs. (2)	Without Covs. (3)	Including Covs. (4)
RD Estimator	-0.427***	-0.492***	0.030***	0.032***
	(.072)	(.074)	(.005)	(.005)
Mean Variable	.645		5.163	
Std. Dev. Variable	.479		.351	
Robust Inference				
p-value	0.00	0.00	0.00	0.00
C.I.	[-.602 -.253]	[-.708 -.277]	[.018 .043]	[.018 .047]
Effective Obs.				
Left	1,617	1,361	1,275	1,077
Right	569	485	485	416
Optimal Bandwidth[a]	.177	.177	.163	.163

Notes: This table presents the results for the impact of grade retention on education outcomes, based on the methods for estimation and inference for fuzzy RD designs developed Calonico et al. (2019). The specifications (2), (4) and (6) include the following covariates: attendance rate in 2006, GPA in 2006, gender, school characteristics and grade dummies. The robust inference considers the bias term coming from the approximation error that does not vanish from the asymptotic distribution of the RD estimator. Standard errors are in parentheses and clustered at the school-level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

This result confirms that —relative to treatment RD group —our control group has a greater probability of being treated in the future. Thus, once again using the discontinuity in the probability of grade retention in 2007, we develop a dynamic model that allows us to identify and estimate the effect of grade retention at different grades (i.e., in different years). The parameters of this model are identified under assumptions that are more demanding than those needed in the RD approach but are still reasonable. We estimate this model using the same estimation sample from our RD estimation —namely, students who were attending 2nd or 3rd grade in 2007.

5.1 Model setting

The model has three periods. The first period ($t = 1$) is the year 2007 and corresponds to the students attending 2nd or 3rd grade. The second period ($t = 2$) is the group of years between 2008 and 2015 and corresponds to the students attending a grade between the 2nd and the 8th. The particular grade depends on the year, the grade the students attended in 2007, and the number of grade retentions sustained. Finally, the third period ($t = 3$) corresponds to the students being aged between 14 and 17 years, when an individual can commit a crime and be punished as a juvenile (C).

An individual i is characterized by the vector X_i, $G3_i$ and τ_i, where X_i and $G3_i$ are observable by the econometrician and τ_i is an unobservable variable with finite support, $\tau_i \in \{1, 2, ..., K\}$ (i.e., unobserved types). In the estimation, we consider three unobserved types ($K = 3$). X includes gender, the average education of both parents, and an indicator variable that takes the value of one when educational information regarding the parents is missing from the SIMCE survey. $G3$ is an indicator variable that takes the value of one if the student is attending the 3rd grade in 2007; it is included to capture the fact that the results can be different depending on the starting point. In the first two periods, student i attends schools j (which can be different between $t = 1$ and $t = 2$). The schools are characterized by the vector of characteristics W_j^t, which considers the average education of fathers and mothers of students at the school, the average scores for mathematics and Spanish from the SIMCE test, and an indicator variable for public schools. The academic performance at period t is characterized by GPA^t and the grade retention indicator R^t ($t \in \{1, 2\}$). GPA^2 is defined as GPA when a student repeated a grade for the first time during the second period or the student's lowest GPA during $t = 2$ if they were not subject to grade retention during this period. Along similar lines, and given that it is possible for a student to attend different schools during the second period, we define the second period school as the one where the student repeated a grade for the first time during $t = 2$ or, if the student was not subject to grade retention during that period, the school where they had the lowest average score across all subjects.

The dynamic model is given by the following equations:

$$GPA_{ij}^1 = \alpha_{\tau_i}^1 + X_i\alpha_x^1 + G3_i\alpha_{gr3}^1 + W_j^1\alpha_w^1 + \varepsilon_{ij}^1. \tag{1}$$

$$GPA^2_{ij} = \alpha^2_{\tau_i} + X_i\alpha^2_x + G3_i\alpha^1_{gr3} + W^1_j\alpha^2_w + R^1_i\alpha^2_{R,\tau_i} + \varepsilon^2_{ij}. \quad (2)$$

$$R^1_i = 1\Big(\gamma^1_{\tau_i} + 1(GPA^1_{ij} < 4.95)\gamma^1_1 + (GPA^1_{ij} - 4.95)\gamma^1_2 + (GPA^1_{ij} - 4.95)^2\gamma^1_3 + G3\gamma^1_4 \geq \eta^1_i\Big). \quad (3)$$

$$R^2_i = 1\Big(\gamma^2_{\tau_i} + R^1_i\gamma^2_1 + GPA^2_{ij}\gamma^2_3 + (GPA^2_{ij})^2\gamma^2_4 + 1(GPA^2_{ij} < 4.5)\gamma^2_5 + 1(GPA^2_{ij} < 5)\gamma^2_6 + G3\gamma^2_7 \geq \eta^2_i\Big). \quad (4)$$

$$C_i = 1\Big(\beta_{\tau_i} + R^1_i\beta_{R1} + R^2_i\beta_{R2} + R^1_iR^2_i\beta_{RR} + G3_i\beta_{g3} + GPA^1_{ij}\beta_{G1} + GPA^2_{ij}\beta_{G2} + X_i\beta_X + W^1_j\beta_{W1} + W^2_j\beta_{W2} \geq \eta^3_i\Big). \quad (5)$$

We assume that all shocks are normally and independently distributed with mean zero and variances equal to $\sigma^2_{\varepsilon 1}$, $\sigma^2_{\varepsilon 2}$, $\sigma^2_{\eta 1}$, $\sigma^2_{\eta 2}$, and $\sigma^2_{\eta 3}$, respectively.

There are some features of this model and the identification of its parameters that are worth highlighting. First, although shocks are independent, the model allows for correlation across dependent variables, conditional on observables, through the unobserved heterogeneity (τ_i). This approach is similar to allowing correlation among shocks and has the advantage that the unobserved type can also accommodate heterogeneity in the impact of the covariates.[24] In fact, we allow for heterogeneity in the impact of grade retention on GPA in the second period (via α^2_{R,τ_i}). This heterogeneity is useful because the effect can be due to a positive or negative effect of R^1_i on future academic performance or to GPA manipulation impacting grade retention probability.. Yet, these two

[24]For example, given two random variables $\tilde{\varepsilon}_i = \varepsilon_i + \beta_{\tau_i}$ and $\tilde{\eta}_i = \eta_i + \alpha_{\tau_i}$, where ε_i and η_i are independent and τ_i is an unobserved heterogeneity with two types: $\tau_i \in \{0, 1\}$ and $Pr(\tau_i = 1) = \pi$. Then, $Cov(\tilde{\varepsilon}_i, \tilde{\eta}_i) = \pi(1-\pi)(\beta_1 - \beta_0)(\alpha_1 - \alpha_0)$.

mechanisms cannot be separately identified. Second, given the evidence supporting our RD strategy, we assume there is no manipulation in the first period and as a consequence γ_1^1 identifies the exogenous variation in R_i^1, which enables the identification of α^2_{R,τ_i}, γ_1^2, and β_{R1}. To the extent that unobserved type is equivalent to allowing for correlation across shocks and that the discontinuity works as an instrumental variable, the causal effect of grade retention in the first period on juvenile crime and on grade retention in the second period are identified given the same exclusion restriction that supports identification in the case of a bivariate probit model (see Li et al. (2019) and Han and Vytlacil (2017)). Third, there are two sources of exogenous variation for R_i^2: the discontinuity in the probability of R_i^1 (which impacts R_i^2) due to the grade retention rule in the first period (which is not manipulated); and the discontinuity in the probability of R_i^2 due to the grade retention rule in the second period. For the second period, we do not observe a perfectly continuous measure for the average score across all subjects (i.e., it has an increment of 0.1), but we include two indicator variables that take the value of one when the average score is below each threshold. In this period, the manipulation is captured by α^2_{R,τ_i}. Fourth, the sources of exogenous variation for R_i^1 and R_i^2 allow us to identify the parameters of interest: β_{R1}, β_{R2}, and β_{RR}.

We estimate this model considering two samples. The first (the *full estimation sample*) is the same sample used in the RD estimation (see Table 1). The second (the *restricted estimation sample*) is the sample that restricts the full estimation sample to those students who are effectively considered in the RD estimation —namely, those with an average score between $4.95 - 0.16$ and $4.95 + 0.16$ —where 0.16 is the RD optimal bandwidth. Although in practice both estimation samples deliver similar results, we prefer the second sample because it makes the exogenous nature of the variation in R_i^1 more reliable.

5.1.1 Estimation

Let Ω be the set of parameters to estimate, such that $\Omega = \{\alpha^1, \alpha^2, \gamma^1, \gamma^2, \beta, \sigma\}$, the likelihood contribution of individual i, whose unobserved type is τ_i, is equal to:[25]

[25] $Z^R_{i,1} = \gamma^1_{\tau_i} + 1(GPA^1_{ij} < 4.95)\gamma^1_1 + (GPA^1_{ij} - 4.95)\gamma^1_2 + (GPA^1_{ij} - 4.95)^2\gamma^1_3 + G3\gamma^1_4$; $Z^R_{i,2} = \gamma^2_{\tau_i} + R^1_i\gamma^2_1 + GPA^2_{ij}\gamma^2_3 + (GPA^2_{ij})^2\gamma^2_4 + 1(GPA^2_{ij} < 4.5)\gamma^2_5 + 1(GPA^2_{ij} < 5)\gamma^2_6 + G3\gamma^2_7$; and $Z^C_i = \beta_{\tau_i} + R^1_i\beta_{R1} + R^2_i\beta_{R2} + R^1_iR^2_i\beta_{RR} + G3_i\beta_{g3} + GPA^1_{ij}\beta_{G1} + GPA^2_{ij}\beta_{G2} + X_i\beta_X + W^1_j\beta_{W1} + W^2_j\beta_{W2}$.

$$L_i(\Omega|GPA,X,G3,W,R,C;\tau_i)=\phi\left(\frac{GPA^1_{ij}-\alpha^1_{\tau_i}-X_i\alpha^1_x-G3_i\alpha^1_{gr3}-W^1_j\alpha^1_w}{\sigma_{\varepsilon1}}\right)\frac{1}{\sigma_{\varepsilon1}}$$
$$\phi\left(\frac{GPA^2_{ij}-\alpha^2_{\tau_i}-X_i\alpha^2_x-G3_i\alpha^2_{gr3}-W^1_j\alpha^2_w-R^1_i\alpha^2_{R,\tau_i}}{\sigma_{\varepsilon2}}\right)\frac{1}{\sigma_{\varepsilon2}}\Phi\left(\frac{Z^R_{i,1}}{\sigma_{\eta1}}\right)^{R^1_i}$$
$$\left[1-\Phi\left(\frac{Z^R_{i,1}}{\sigma_{\eta1}}\right)\right]^{1-R^1_i}\Phi\left(\frac{Z^R_{i,2}}{\sigma_{\eta2}}\right)^{R^2_i}\left[1-\Phi\left(\frac{Z^R_{i,2}}{\sigma_{\eta2}}\right)\right]^{1-R^2_i}\Phi\left(\frac{Z^C_i}{\sigma_{\eta3}}\right)^{C_i}\left[1-\Phi\left(\frac{Z^C_i}{\sigma_{\eta3}}\right)\right]^{1-C_i}.$$

Let π_k be the unconditional probability that an individual is type k, then the likelihood function is given by:

$$L(\Omega|GPA,X,G3,W,R,C)=\prod_{i=1}^{N}\left(\sum_{k=1}^{K}\pi_k L_i(\Omega|GPA,X,G3,W,R,C;\tau_i=k)\right). \quad (6)$$

The estimated parameters are the Ω and $\{\pi_k\}_{k=1}^K$ that maximize L.[26] As is common in these types of models, the standard errors are calculated using the approximation of the Hessian given by the mean of the outer product of the scores.

5.2 Results

In Appendix D we present the point estimates and their standard errors for the two estimation samples. Before performing two exercises to study the effect of grade retention on juvenile crime, there are three aspects of our estimation results worth noting. First, all the signs are as expected and most of the point estimates are statistically significant. Second, the results from these two estimation samples are qualitatively similar. Third, in the case of the restricted sample the three unobserved types collapse into one, which reinforces the idea that when we focus on the students who are at the margin of grade retention (our preferred specification) there are no unobserved differences among the students below and above the threshold.

We run two simulation exercises to assess the effect of grade retention on crime. The first exercise is to simulate the marginal effect of grade retention in the first period and in

[26]We minimize $-L(\Omega|GPA,X,G3,W,R,C)$ using the Matlab solver *fminsearch*.

the second period on juvenile crime, without dynamic. Thus, we use the point estimates of γ^2 to simulate the effect of grade retention in the first period on juvenile crime, without considering its indirect effect through academic performance and grade retention in the second period. In other words, we only use equation (5) in the simulation. In terms of the effect of grade retention in the second period on crime, our model does not have dynamics.

Table 6 (panel A) shows the following results from the first exercise. Using the restricted estimation sample we see that grade retention in the first period decreases the probability of crime by 2.1 pp. And that grade retention in the second period increases the probability of crime by 3.8 pp for those students who were subject to a grade retention in the first period and by 4.6 pp for those students who were retained for the first time. For the full estimation sample, first period grade retention increases the probability of crime by 0.01 pp and second period grade retention increases the probability of crime by 4.9 pp (for those retained in the first period) and 5.3 pp (for those retained for the first time).[27] These results confirm our concerns about the dynamics in the sense that the negative and strong effect we get from our RD estimation is mainly driven by the difference between the effect of grade retention in the first period versus the second period rather than a relevant negative effect of grade retention on juvenile crime.

To be sure about this interpretation, we run the second exercise which —by considering the dynamics —seeks to replicate our RD estimation by simulating our estimated model. We take the following steps: First, for each student we simulate GPA in the first period, using equation (1). Second, we define as marginal those students whose simulated first period GPA is between $4.95 - 0.16$ and $4.95 + 0.16$. We only keep the sample of simulated marginal students. Third, we randomly assign the treatment of grade retention in the first period to one half of the simulated marginal students. Fourth, for the treated and control simulated groups, we simulate equations (2), (4), and (5). We do so considering all the dynamics —namely, the simulation of R^1 affects GPA^2, the simulation of R^1 and GPA^2 affect R^2, and the simulation of R^1, GPA^2, and R^2 impact C. Finally, given all these simulations, we can evaluate the (full dynamic) impact of R^1

[27] In a previous version of this paper, we estimated the effect of grade retention (at 3rd-5th grades) on juvenile crime, finding an increase of 4.6 pp. To the extent that 3rd-5th grades are close to our second-period definition, it is remarkable how similar are the magnitudes from our simulation with the effects found in the previous paper.

on GPA^2, R^2, and C.

Table 6 (panel B) shows the results from this model-based RD simulation. Using the restricted estimation sample and taking into account all the dynamics of the model, we see that first period grade retention decreases the probability of juvenile crime by 5 pp, increases second period GPA by 0.12 points, and decreases second period grade retention probability by 26.8 pp. In the case of the full estimation sample, the figures are 1.4 pp, 0.08 points, and 14.6 pp, respectively. Notice that these numbers are qualitatively (and to lesser degree quantitatively) equivalent to our RD estimations, particularly in the case of the restricted sample. This similarity is reassuring as these marginal effects are moments that we do not directly use in estimating our model. Therefore, the simulations from this second exercise strongly support the interpretation that the results from our RD estimation are not driven by a direct and relevant negative effect of grade retention on juvenile crime, but they are mainly driven by a combination of a negative effect of grade retention in early primary grades on grade retention in later primary grades, with an increasing impact of grade retention on juvenile crime as students progress through primary grades.

Table 6: Model Simulations

	Estimation sample	
	Restricted	Full
Panel A: Direct effects		
Grade retention in low grades on crime	-0.021	0.001
Grade retention in high grades on crime (first repetition)	0.046	0.053
Grade retention in high grades on crime (second repetition)	0.038	0.049
Panel B: Direct + indirect effects		
Grade retention in low grades on future academic performance	0.120	0.080
Grade retention in low grades on future grade retention	-0.268	-0.146
Grade retention in low grades on crime	-0.050	-0.014

Notes: This table presents the results from two simulation exercises, considering two estimation samples: Restricted (GPA between $4.95-0.16$ and $4.95+0.16$, $N = 1,787$) and Full ($N = 11,813$). Panel (A) shows the results from the simulation of the marginal effect of grade retention in low grades and in high grades on crime, without dynamic. Low grades is our first period and high grades our second period (both in primary school). Panel (B) shows the simulation of the effect of grade retention in the low grades but considering the dynamic, namely, the simulation of R^1 affects GPA^2, the simulation of R^1 and GPA^2 affect R^2, and the simulation of R^1, GPA^2, and R^2 impacts C.

6 Conclusion

This is the first paper that estimates a causal effect of grade retention on juvenile crime in a developing country. We implement the standard fuzzy RD approach developed by Calonico et al. (2014b) and Calonico et al. (2019) by exploiting a discontinuity in the probability of not being promoted to the next grade that is produced by a grade retention rule in the Chilean educational system. Our results show that repeating a grade —for the first time —in the 2nd or 3rd grade decreases the probability of a student committing a crime as a juvenile crime by 14.5 pp and by 10.7 pp for a severe crime.

In addition to this empirical approach, we estimate a semi-structural dynamic model that is crucial to correctly interpreting the RD results in order to guide a policy discussion. More specifically, model simulations show that the decrease in the probability of juvenile crime is not because of a negative and relevant direct effect of grade retention on juvenile crime but is due to the impact of grade retention on future grade retention probability. Hence, our RD results are driven by grade retention timing, given that the grade retention in the later grades of primary education has a positive and much more relevant effect on crime than the direct effect in early grades. The insights produced by this dynamic model may be very useful for understanding the heterogeneity that we observe in the literature regarding the effect of grade retention on juvenile crime and how this effect depends on the timing of the retention.

The evidence from this paper calls into question the appropriateness of grade retention as a public policy. And this concern becomes even more relevant in the context of Chile, a developing country with high rates of grade retention. If policymakers continue to support this practice, our results indicate that the optimal policy is to retain students in early grades when their performance is around the threshold as a way to decrease the probability of grade retention in late primary school grades.

That said, any interpretation of our findings should consider that we do not take into account other aspects of this policy. Our approach is silent, for example, on how the threat of retention could serve as an incentive for all students to exert more effort (see, for instance, Koppensteiner (2014)). Therefore, our results should be considered as only one part of the story and a call for a more comprehensive evaluation of grade retention as a recurrent educational policy.

References

Blanco, R., R. Hutt, and H. Rojas (2004): "Reform to the Criminal Justice System in Chile: Evaluation and Challenges," *Loyola University Chicago International Law Review*, 2, 253–269.

Brugård, K. H. and F. Torberg (2013): "Post-compulsory education and imprisonment," *Labour Economics*, 97, 97–106.

Burdick-Will, J. (2013): "School violent crime and academic achievement in Chicago," *Sociology of education*, 86, 343–361.

Calonico, S., M. D. Cattaneo, M. Farrell, and R. Titiunik (2019): "Regression Discontinuity Designs Using Covariates," *Econometrica*, 101, 442–451.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a): "Robust data-driven inference in the regression-discontinuity design," *Stata Journal*, 14, 909–946.

——— (2014b): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

Cattaneo, M. D., N. Idrobo, and R. Titiunik (2019): *A practical introduction to regression discontinuity designs: Foundations*, Cambridge University Press.

Cattaneo, M. D., M. Jansson, and X. Ma (2017): "Simple Local Polynomial Density Estimators," working paper., University of Michigan.

Cook, P. J. and S. Kang (2016): "Birthdays, Schooling, and Crime: Regression-Discontinuity Analysis of School Performance, Delinquency, Dropout, and Crime Initiation," *American Economic Journal: Applied Economics*, 8, 33–57.

Cortés, T., N. Grau, and J. Rivera (2020): "Juvenile Incarceration and Adult Recidivism," Working paper, Universidad de Chile.

Couso, J. and J. Duce (2013): *Juzgamiento penal de adolescentes*, LOM ediciones.

Depew, B. and O. Eren (2016): "Born on the wrong day? School entry age and juvenile crime," *Journal of Urban Economics*, 96, 73–90.

Eren, O., B. Depew, and S. Barnes (2017): "Test-based promotion policies, dropping out, and juvenile crime," *Journal of Public Economics*, 153, 9–31.

Eren, O., M. F. Lovenheim, and N. H. Mocan (2018): "The effect of grade retention on adult crime: Evidence from a test-based promotion policy," Tech. rep., National Bureau of Economic Research.

European Institute for Crime Prevention and Control, affiliated with the United Nations (2010): "International Statistics on Crime and Justice," HEUNI Publication Series 64, United Nations.

Fagan, J. and E. Pabon (1990): "Contributions of delinquency and substance use to school dropout among inner-city youths," *Youth and Society*, 21, 306.

Fruehwirth, J. C., S. Navarro, and Y. Takahashi (2016): "How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects," *Journal of Labor Economics*, 34, 979–1021.

Gauri, V. (1999): *School Choice In Chile: Two Decades of Educational Refor*, Pitt Latin American Studies, University of Pittsburgh Press.

Grau, N., D. Hojman, and A. Mizala (2018): "School closure and educational attainment: Evidence from a market-based system," *Economics of Education Review*, 65, 1 – 17.

Greene, J. P. and M. A. Winters (2009): "The effects of exemptions to Florida's test-based promotion policy: Who is retained?: Who benefits academically?" *Economics of Education Review*, 28, 135–142.

Han, S. and E. J. Vytlacil (2017): "Identification in a generalization of bivariate probit models with dummy endogenous regressors," *Journal of Econometrics*, 199, 63 – 73.

Hirschfield, P. (2009): "Another way out: The impact of juvenile arrests on high school dropout," *Sociology of Education*, 82, 368–393.

Holmes, C. T. et al. (1989): "Grade level retention effects: A meta-analysis of research studies," *Flunking grades: Research and policies on retention*, 16, 33.

Jacob, B. A. (2005): "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago," *Journal of Public Economics*, 89, 761–796.

Jacob, B. A. and L. Lefgren (2009): "The Effect of Grade Retention on High School Completion," *American Economic Journal: Applied Economics*, 1, 33–58.

Jimerson, S. R. (2001): "Meta-analysis of grade retention research: Implications for practice in the 21st century," *School psychology review*, 30, 420.

King, E. M., P. F. Orazem, and E. M. Paterno (2015): "Promotion with and without learning: Effects on student enrollment and dropout behavior," *The World Bank Economic Review*, lhv049.

Koppensteiner, M. F. (2014): "Automatic grade promotion and student performance: Evidence from Brazil," *Journal of Development Economics*, 107, 277 – 290.

Landersø, R., H. S. Nielsen, and M. Simonsen (2016): "School Starting Age and the Crime-age Profile," *The Economic Journal*, dOI: 10.1111/ecoj.12325.

Langer, M. and R. Lillo (2014): "Reforma a la justicia penal juvenil y adolescentes privados de libertad en Chile: Aportes empíricos para el debate," *Política criminal*, 9, 713 – 738.

Li, C., D. Poskitt, and X. Zhao (2019): "The bivariate probit model, maximum likelihood estimation, pseudo true parameters and partial identification," *Journal of Econometrics*, 209, 94 – 113.

Lochner, L. (2004): "Education, Work, And Crime: A Human Capital Approach," *International Economic Review*, 45, 811–843.

Lochner, L. and E. Moretti (2004): "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports," *American Economic Review*, 94, 155–189.

Machin, S., O. Marie, and S. Vujić (2011): "The crime reducing effect of education," *The Economic Journal*, 121, 463–484.

Manacorda, M. (2012): "The Cost of Grade Retention," *The Review of Economics and Statistics*, 94, 596–606.

Meckes, L. and R. Carrasco (2010): "Two decades of SIMCE: an Overview of the National Assessment System in Chile," *Assessment in Education: Principles, Policy and Practice*, 17, 233–248.

Ou, S.-R. and A. J. Reynolds (2010): "Grade retention, postsecondary education, and public aid receipt," *Educational Evaluation and Policy Analysis*, 32, 118–139.

Reschly, A. L. and S. L. Christenson (2013): "Grade retention: Historical perspectives and new research," *Journal of school psychology*, 51, 319–322.

Roderick, M. (1994): "Grade retention and school dropout: Investigating the association," *American Educational Research Journal*, 31, 729–759.

Solis, A. (2017): "The Effects of Grade Retention on Human Capital Accumulation," Working Paper Series 2017:15, Uppsala University, Department of Economics.

Appendix

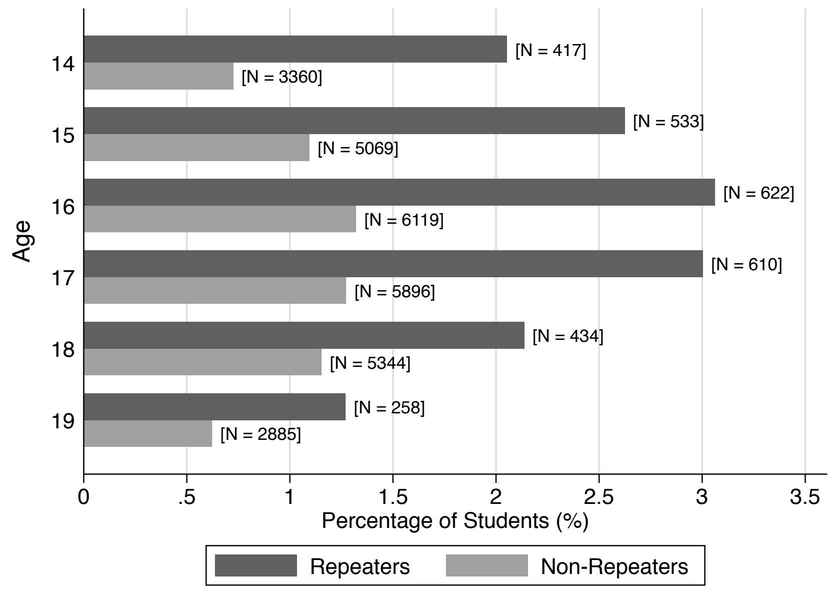
A Descriptive statistics for juvenile crime

Table 7: JUVENILE CRIME DISTRIBUTION

Crime Category	Freq.	Percentage	Pretrial Detention (%)
Theft	1057	24.34	0.76
Non-violent Robbery	662	15.25	11.03
Other Crimes Agains Property	567	13.06	4.23
Robbery	522	12.02	29.89
Injuries	394	9.07	1.78
Crimes Against Sexual Freedom and Privacy	343	7.90	3.50
Other Crimes	274	6.31	1.46
Offenses	169	3.89	0.59
Crimes Against Drug Laws	123	2.83	7.32
Crimes Against Special Laws	67	1.54	29.85
Traffic Law Crimes	52	1.20	0.00
Sex Crimes	45	1.04	8.89
Crimes Against Public Faith	18	0.41	5.56
Homicides	18	0.41	44.44
Intellectual and Industrial Property Crimes	13	0.30	7.69
Financial and Tax Crimes	9	0.21	0.00
Crimes Against Millitary Laws	5	0.12	20.00
Negligent Offense	3	0.07	0.00
Facts of Criminal Relevance	1	0.02	0.00

Notes: This plot shows distribution of crimes for students who attended 2nd or 3rd grade in year 2007. Severe crimes are the ones in which case the pretrial detention rate is greater than 3%.

Figure 6: % OF STUDENTS WHO WERE CRIMINALLY PROSECUTED BY AGE



Notes: This table shows, among repeaters and non repeaters at 2nd and 3rd grades (2007), the fraction of students prosecuted for the first time at different ages.

B More results

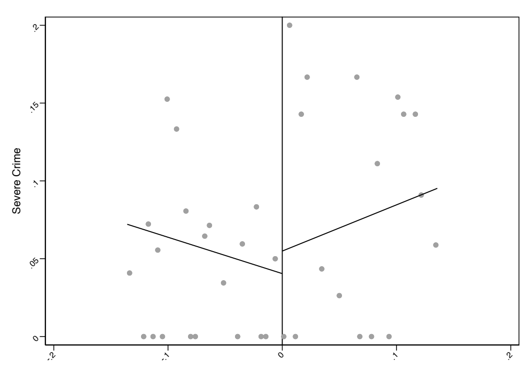
Table 8: Effects of Grade Retention: reduced form

	All Crimes	Severe Crimes	Dropout	Future Grade Retention	GPA From 4th to 8th Grade
	(1)	(2)	(3)	(4)	(5)
RD Estimator	.071**	.056*	.225***	.293***	-.198***
	(.027)	(.025)	(.058)	(.052)	(.04)
Mean Variable	.132	.106	.508	.645	5.163
Std. Dev. Variable	.027	.025	.058	.052	.04
Robust Inference					
p-value	0.02	0.05	0.00	0.00	0.00
C.I.	[.01 .132]	[0 .113]	[.1 .35]	[.179 .408]	[-.284 -.112]
Effective Obs.					
Left	1,845	1,618	829	1,062	898
Right	610	571	415	470	407
Optimal Bandwidth[a]	.193	.177	.104	.128	.12

Notes: This table presents the results for the impact of grade retention on the 5 listed outcomes, based on the methods for estimation and inference for sharp RD designs developed by Calonico et al. (2019). Note that in this case the sign is reverse because now we are accounting for the effect of crossing the score cutoff (which decrease the probability of grade retention). Therefore the interpretation here is: by getting a average score above the cutoff, and therefore decreasing their probability of grade retention, the students are, on average, 7 pp more likely to commit a crime, 22 pp more likely to dropout and 29 pp more likely of being retained a grade in the future. The robust inference considers the bias term coming from the approximation error that does not vanish from the asymptotic distribution of the RD estimator. Standard errors are in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 7: Graphic results for severe crime

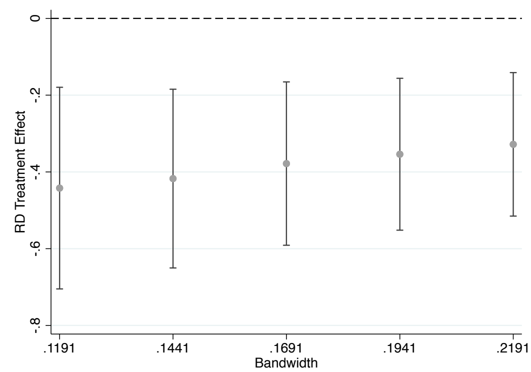
(a) Severe Crimes



Notes: This figure shows the outcomes values and an estimation of the regression functions via local linear regressions around the threshold for grade retention for the case of severe crimes.

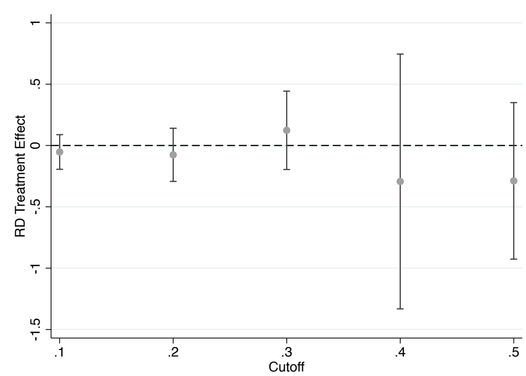
C Robustness Analysis

Figure 8: Sensitivity to Bandwidth: Dropout



Notes: This figure shows the fuzzy RD estimations for the impact of grade retention on juvenile crime (using the methods developed by Calonico et al. (2019)), for different values of the bandwidth (the third estimate is the one with the optimal bandwidth). The point estimates are the dots and the confidence intervals at 95% are the brackets.

Figure 9: Placebo Tests: Dropout



Notes: This figure shows the sharp RD estimations for the impact of being below the cutoff on juvenile crime (using the methods developed by Calonico et al. (2019)), for different values of the cutoff. The point estimates are the dots and the confidence intervals at 95% are the brackets.

D Model's estimated parameters

D.1 Restricted sample

Table 9: GPA estimated parameters (Restricted Sample)

	1st period GPA		2nd period GPA	
	α	S.E.	α	S.E.
Male	-0.007	0.004	-0.123	0.020
3rd Grade in 2007	-0.017	0.005	-0.086	0.026
Missing mother educ	0.029	0.006	-0.065	0.028
Mother educ. >14 years	0.011	0.009	0.036	0.044
Sch. Average: Father educ.	-0.001	0.003	0.009	0.013
Sch. Average: Mother educ.	0.005	0.003	0.002	0.014
Sch. Average: Math SIMCE	-0.0003	0.0001	0.003	0.001
Sch. Average: Spanish SIMCE	0.0001	0.0001	-0.001	0.001
Public School	0.005	0.005	0.088	0.023
Repeated in period one (type I)	.	.	0.132	0.027
Repeated in period one (type II)	.	.	-1.146	0.179
Constant Type I	4.883	0.029	4.076	0.142
Constant Type II	4.892	0.036	3.072	0.195
Log(Standard Error)	-2.526	0.023	-0.891	0.016

Notes: This table presents the point estimates and standard errors for the parameters of equations (1) and (2), using the restricted sample. The standard errors are calculated using the approximation of the Hessian given by the mean of the outer product of the scores.

Table 10: GRADE RETENTION, CRIME, AND TYPES ESTIMATED PARAMETERS (RESTRICTED SAMPLE)

	Grade Retention			Crime	
	γ	S.E.		β	S.E.
First Period:			Repeated in 1st period	-0.141	0.203
$1(GPA < 4.95)$	1.545	0.149	Repeated in 2nd period	0.318	0.205
$GPA - 4.95$	-3.183	0.893	Repeated both periods	-0.058	0.230
$(GPA - 4.95)^2$	24.475	5.926	Male	0.504	0.106
3rd Grade in 2007	-0.316	0.085	3rd Grade in 2007	0.132	0.121
Constant Type I	-0.409	0.117	Missing mother educ	0.017	0.133
Constant Type II	-1.400	0.287	Mother educ. >14 years	0.344	0.242
			First Period variables:		
Second Period:			GPA	-0.442	0.809
Repeated in 1st period	-0.673	0.084	Sch. Average: Father educ.	0.020	0.099
GPA	-1.410	1.267	Sch. Average: Mother educ.	0.079	0.096
GPA^2	0.130	0.137	Sch. Average: Math SIMCE	-0.010	0.006
$1(GPA < 4.45)$	0.755	0.129	Sch. Average: Spanish SIMCE	0.003	0.005
$1(GPA < 4.95)$	1.124	0.160	Public School	0.240	0.151
3rd Grade in 2007	-0.102	0.071	**Second Period variables:**		
Constant Type I	2.834	2.974	GPA	-0.631	0.140
Constant Type II	0.422	2.446	Sch. Average: Father educ.	-0.004	0.101
			Sch. Average: Mother educ.	-0.197	0.099
Type distribution:			Sch. Average: Math SIMCE	0.000	0.006
Type I parameter	4.209	0.263	Sch. Average: Spanish SIMCE	0.002	0.005
Type I probability	0.99	.	Public School	-0.169	0.160
Type II probability	0.01	.	Constant Type I	5.818	4.114
			Constant Type II	5.252	4.125

Notes: This table presents the point estimates and standard errors for the parameters of equations (3), (4), and (5), using the restricted sample. The standard errors are calculated using the approximation of the Hessian given by the mean of the outer product of the scores.

D.2 Full sample

Table 11: GPA estimated parameters (Full Sample)

	1st period GPA		2nd period GPA	
	α	S.E.	α	S.E.
Male	-0.083	0.007	-0.116	0.009
3rd Grade in 2007	-0.009	0.008	-0.087	0.011
Missing mother educ	0.033	0.009	-0.038	0.011
Mother educ. >14 years	0.002	0.018	0.077	0.024
Sch. Average: Father educ.	0.012	0.005	-0.012	0.006
Sch. Average: Mother educ.	0.011	0.005	0.026	0.006
Sch. Average: Math SIMCE	0.001	0.0002	0.002	0.0004
Sch. Average: Spanish SIMCE	0.001	0.0002	-0.0001	0.0004
Public School	-0.015	0.007	0.029	0.009
Repeated at period one (type I)	.	.	0.170	0.157
Repeated at period one (type II)	.	.	0.080	0.019
	0.308			
Constant Type I	3.349	0.047	3.680	0.047
Constant Type II	3.955	0.046	4.019	0.061
Log(Standard Error)	-1.209	0.011	-0.775	0.005

Notes: This table presents the point estimates and standard errors for the parameters of equations (1) and (2), using the full sample. The standard errors are calculated using the approximation of the Hessian given by the mean of the outer product of the scores.

Table 12: GRADE RETENTION, CRIME, AND TYPES ESTIMATED PARAMETERS (FULL SAMPLE)

	Grade Retention			Crime	
	γ	S.E.		β	S.E.
First Period:			Repeated in 1st period	0.005	0.111
$1(GPA < 4.95)$	1.853	0.085	Repeated in 2nd period	0.269	0.127
$GPA - 4.95$	-1.321	0.152	Repeated both periods	-0.021	0.131
$(GPA - 4.95)^2$	-0.426	0.125	Male	0.475	0.036
3rd Grade in 2007	-0.289	0.046	3rd Grade in 2007	0.053	0.037
Constant Type I	-0.125	0.274	Missing mother educ	0.015	0.039
Constant Type II	-0.323	0.061	Mother educ. >14 years	0.040	0.088
			First Period variables:		
Second Period:			GPA	-0.059	0.097
Repeated in 1st period	-0.347	0.050	Sch. Average: Father educ.	0.021	0.028
GPA	0.034	0.280	Sch. Average: Mother educ.	0.010	0.028
GPA^2	0.016	0.034	Sch. Average: Math SIMCE	-0.002	0.002
$1(GPA < 4.45)$	1.083	0.042	Sch. Average: Spanish SIMCE	0.000	0.002
$1(GPA < 4.95)$	1.177	0.064	Public School	0.040	0.043
3rd Grade in 2007	-0.069	0.026	**Second Period variables:**		
Constant Type I	-1.313	0.569	GPA	-0.386	0.036
Constant Type II	-1.727	0.572	Sch. Average: Father educ.	-0.030	0.027
			Sch. Average: Mother educ.	-0.040	0.028
Type distribution:			Sch. Average: Math SIMCE	-0.001	0.002
Type I parameter	3.296	0.141	Sch. Average: Spanish SIMCE	-0.001	0.002
Type II parameter	4.882	0.139	Public School	0.080	0.043
Type I probability	0.17	.	Constant Type I	2.077	0.465
Type II probability	0.82	.	Constant Type II	1.935	0.553

Notes: This table presents the point estimates and standard errors for the parameters of equations (3), (4), and (5), using the restricted sample. The standard errors are calculated using the approximation of the Hessian given by the mean of the outer product of the scores.