

DISCUSSION PAPER SERIES

IZA DP No. 13770

**The Effect of Grade Retention on Adult
Crime: Evidence from a Test-Based
Promotion Policy**

Ozkan Eren
Michael F. Lovenheim
H. Naci Mocan

OCTOBER 2020

DISCUSSION PAPER SERIES

IZA DP No. 13770

The Effect of Grade Retention on Adult Crime: Evidence from a Test-Based Promotion Policy

Ozkan Eren

University of California, Riverside

Michael F. Lovenheim

Cornell University and NBER

H. Naci Mocan

Louisiana State University, NBER and IZA

OCTOBER 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Effect of Grade Retention on Adult Crime: Evidence from a Test-Based Promotion Policy*

This paper presents the first analysis in the literature of the effect of test-based grade retention on adult criminal convictions. We exploit math and English test cutoffs for promotion to ninth grade in Louisiana using administrative data on all public K-12 students combined with administrative data on all criminal convictions in the state. Our preferred model uses the promotion discontinuity as an instrument for grade retention, and we find that being retained in eighth grade has large long-run effects on the likelihood of being convicted of a crime by age 25 and on the number of criminal convictions by age 25. Effects are largest for violent crimes: the likelihood of being convicted increases by 1.05 percentage points, or 58.44%, when students are retained in eighth grade. Our data allow an examination of mechanisms, and we show that the effects are likely driven by declines in high school peer quality and educational investments that result in lower non-cognitive skill acquisition. We find little effect on juvenile crime. Using the method proposed by Angrist and Rokkanen (2015), we also estimate effects of grade retention away from the promotion cutoff and show that our results are generalizable to a larger group of low-performing students. Our estimates indicate that eighth grade test-based promotion cutoffs lead to nontrivial private and social costs in terms of higher levels of long-run criminal convictions that are important to consider in the development and use of these policies.

JEL Classification: I2, K4

Keywords: grade retention, crime, grade promotion policies, regression discontinuity, education, discontinuity high school, middle school

Corresponding author:

Naci Mocan
Louisiana State University
Department of Economics
3039 Business Education Complex
Baton Rouge, LA 70803
USA
E-mail: mocan@lsu.edu

* We would like to thank seminar participants at University of Illinois-Urbana Champaign, National University-Singapore, Singapore Management University, University of California-Irvine, University of South Carolina, CESifo Area Conference on the Economics of Education, and the Association for Education Finance and Policy annual meeting for helpful comments and suggestions. Data used in this study was provided by the Louisiana Department of Education, Louisiana Office of Juvenile Justice, and Louisiana Department of Public Safety and Corrections, Adult Services. Access was provided by Louisiana State University. All results, conclusions and errors are our own.

1 Introduction

Education policies alter the skills and knowledge of young adults, which influence not only individual earnings but also socially relevant outcomes such as civic engagement, health behaviors, and criminal activity later in life.¹ These social returns to education are critical to understand, because much of the argument for the large role played by the government in funding and delivering education services is predicated on the existence of such positive externalities. While a sizable literature exists that examines the private returns to education investments and policies in the form of long-run labor market outcomes, there is a paucity of evidence on the existence and magnitude of social returns (Grossman 2006; Oreopoulos and Salvanes 2011).

This paper helps fill the gap in our understanding of the social returns to education policy by estimating the effect of a test-based grade retention policy for promotion to 9th grade on adult criminal convictions in Louisiana. Crime is one of the most important social outcomes to examine because it generates enormous costs to society.² Investigation of the determinants of criminal activity also is important because delinquency and ensuing incarceration have long-run negative effects on a number of outcomes, ranging from employment opportunities to family formation and health (Johnson and Raphael 2009; Charles and Luoh 2010; Agan and Starr 2017).

Education can lead to reductions in adult criminal behavior by increasing the opportunity cost of engaging in crime (Lochner 2004) as well as by altering time discounting (Becker and Mulligan 1997) and by building non-cognitive skills that contribute to better self-control and decision-making as adults (Fudenberg and Levine 2006). Despite the strong theoretical arguments linking education to crime, empirical work is sparse because of the difficulty in overcoming selection into criminal activity: those who obtain less schooling tend to have unobserved characteristics that also make them more likely to commit a crime. Existing work with credible identification strategies has investigated the impact of education, as measured by the number of years of schooling, on criminal propensity (e.g., Lochner and Moretti 2004; Machin et al. 2011;

¹See Card (2001) and Oreopoulos (2007) for estimates of the effect of education on wages. For estimates of the impact of education on civic participation, production of health, and crime, see Dee (2004), Lochner and Moretti (2004), Lleras-Muney (2005) and Chou et al. (2010).

²Ferraz and Soares (2018) calculate that the expenditures on the criminal justice system and the cost of victimization add up to \$450 billion in the U.S., and the estimates of Anderson (1999) imply that in 2018 the total cost of criminal activity, including lost productivity and diminished quality of life, is \$1.7 trillion.

Hjalmarsson et al. 2015). In this paper, we instead estimate the effect on crime of a specific education policy that governs promotion to 9th grade. Specifically, we exploit a grade retention policy in Louisiana in a regression discontinuity framework to examine the extent to which having been held back in 8th grade impacts criminal delinquency up to age 25. Our analysis thus makes contributions both to the literature on the impact of human capital formation on crime and to the literature that investigates the consequences of test-based promotion policies.

The test-based accountability program we analyze was adopted in 1998 and requires 8th grade students to achieve a certain level of proficiency on both the state math and English exams in order to be promoted to 9th grade. Students who score below the cutoff on either test are offered remedial summer classes at no cost, and they retake the failed exam(s) in the summer regardless of whether they enroll in summer school. If a student fails to reach the passing threshold on one of the summer exams, she is retained in 8th grade.

Louisiana was an early adopter of this test-based accountability program, and these programs have grown markedly in popularity in the past several decades with the rise of accountability policies more broadly. Typically, accountability policies are aimed at teachers (e.g., teacher merit pay) or schools (e.g., No Child Left Behind). Test-based promotion policies, on the other hand, are aimed directly at students, and currently 16 states as well as many large school districts incorporate end-of-year exams into the decision to promote the students to the next grade level (Zinth 2005; Marsh et al. 2009; Workman 2014).³ By ending “social promotion,” the goal of these policies is to provide incentives for students to pass exams that policymakers hope will ultimately increase learning. A growing literature examines the effect of these policies, and of grade retention more generally, on measures of academic performance and juvenile crime (e.g., Greene and Winters 2007; Jacob and Lefgren 2009; Manacorda 2012; Eren et al. 2017; Schwerdt et al. 2017; Mariano et al. 2018). The effect of these policies on longer-run adult outcomes has remained unexamined prior to this analysis.

We employ two detailed administrative datasets from Louisiana to estimate the effect of being retained in 8th grade due to failing the summer promotion exam. Our data combine ad-

³Among these states and school districts, the promotion and retention policies vary along several key dimensions including, but not limited to, grade level, subjects tested and waiver criteria. For example, Arizona and Florida target their promotion policies at grade 3, while Georgia and Texas include a wider range and focus on grades 5 and 8. Despite their popularity, the findings on the efficacy of test-based promotion policies on long-run outcomes is far from conclusive. This mixed evidence is not solely a function of variation in design features of promotion policies, because policies implemented at similar grades on identical subjects yield different long-run impacts (e.g., Jacob and Lefgren (2009) and Eren et al. (2017)).

ministrative records on all K-12 students from the Louisiana Department of Education (LDOE) with administrative crime records on individuals up to age 25 from the Louisiana Department of Public Safety and Corrections, Adult Services. These data allow us to link together the universe of K-12 students in Louisiana with the universe of adult criminal convictions. Because of our focus on adult crime among those 25 years and younger, we focus on 8th graders from the 1998-1999 through 2000-2001 school years (the 1984-1987 birth cohorts). Our analysis is further restricted to those who fail one of the initial exams (in March of the relevant academic year). Thus, the effects we estimate are relative to a low-performing set of students on the margin of failing to be promoted to 9th grade. We also employ the method developed by Angrist and Rokkanen (2015) to extrapolate effects away from the cutoff, which reveals that the results are generalizable to a larger group of low-performing students.

In our preferred IV model that uses the promotion cutoff as an instrument for grade retention, we find that being retained in 8th grade increases the likelihood of being convicted of any crime by a non-statistically-significant 1.25 percentage points. This is a 10.85% increase relative to the mean. The effect is driven by violent crime: 8th grade retention increases the likelihood of violent crime conviction as a young adult by a statistically significant 1.05 percentage points (58.44%). We find no effect on the likelihood of committing a property or drug related crime, although the point estimates for drug crimes are positive and sizable in magnitude. As we discuss below, we do find positive and statistically significant effects on these other crime categories when extrapolating away from the promotion cutoff. We also find positive effects on robbery convictions. Although robbery is motivated by monetary gain like property crime, robbery is a violent crime. Examining the number of convictions at first conviction instead of just the probability of conviction shows that retention increases the number of criminal convictions by 0.025 (18.38%), which is composed primarily of a 0.017 (85.00%) violent crime increase that is significant at the 5 percent level.

The rich data on educational outcomes allow us to examine several potential mechanisms. We show that grade retention leads students to be exposed to lower-quality education environments, worsens non-cognitive outcomes, and lowers educational attainment. Specifically, failing the 8th grade promotion exam leads students to subsequently attend high schools with lower-

performing and more disadvantaged students. We also find that retention in 8th grade increases the number of days absent three years later by almost two days (11.14% relative to the mean) and increases the number of disciplinary incidents by at least 1.1 percentage points. Grade retention additionally leads to a 7.17 percentage point reduction in the likelihood of high school completion and a 0.264 reduction in years of high school completed. Estimates of the effect on juvenile crime are positive but are small in magnitude and are not statistically different from zero. Using results from the literature, we argue that reduced educational attainment and any juvenile crime effects are too small to explain our results. Instead, the evidence on mechanisms suggests that failing to be promoted to 9th grade on time reduces investment in non-cognitive skills, which manifests itself in terms of lower school attendance and worse behavioral problems as students age. The result is a higher probability of being convicted of a crime as an adult, and this effect is not simply a reflection of a higher likelihood of being in the criminal justice system as a juvenile. Thus, retaining students in 8th grade due to not passing this high stakes exam leads to substantially worse outcomes among a subset of these students in their young adulthood.

Finally, we show that the effects are largest for 20+ year olds and that the impact of grade retention on crime generalizes to students away from the test passing threshold. Using the technique pioneered by Angrist and Rokkanen (2015), we estimate effects of retention on in-framarginal students who pass the exam and who fail the exam. The results indicate sizable positive effects of retention on crime for both groups. Indeed, we find positive and statistically significant effects away from the cutoff for violent, property and drug-related crime. These findings suggest that our regression discontinuity estimates understate the effects of 8th grade retention on crime for a broader population of low-scoring students and for a broader set of crimes. They also indicate that the passing threshold cannot be simply moved up or down to eliminate the crime effect; the effect of grade retention on crime is evident at any reasonable retention cutoff in the test score distribution.

This paper is the first to estimate the causal effect of grade retention, due to test-based accountability policies, on adult criminal convictions. We thus make contributions to several strands of the literature. First, this paper adds to a growing body of research that examines

the effect of grade retention on student outcomes. Most studies that credibly overcome selection concerns investigate test-based student accountability policies in a regression discontinuity framework, as we do here. Jacob and Lefgren (2009) provide the earliest such analysis focused on the test-based accountability regime in Chicago. They find that test-based grade retention in the 6th grade has no effect on high school dropout but that 8th grade test-based retention significantly reduces the likelihood of high school completion. Florida’s test-based retention program focusing on 3rd grade students also has been analyzed by several researchers. Students who barely fail the promotion cutoff perform better on standardized exams at first, but this effect fades out by high school. They have higher GPAs as well, but the likelihood of graduating is unaffected (Schwert et al. 2017; Greene and Winters 2007). Retaining students in early grades may have fundamentally different long-run effects than retaining them in higher grades. This may explain the differences in findings across these studies.⁴

Eren et al. (2017) make use of the same Louisiana policy and employ similar data as we do in this paper to estimate the effect of summer school on high school dropout. Their supplementary analysis of grade retention shows that grade retention in the 8th grade has at most a small impact on juvenile crime convictions.⁵ This finding, which is also confirmed by our paper, underscores the importance of looking at longer-run outcomes like adult crime. A mounting body of evidence suggests that short-run effects of educational interventions differ substantively from longer-run effects, though we are the first to show that this is the case with crime outcomes.⁶ Examining only juvenile crime would lead one to falsely conclude that grade retention has at most a small effect on criminal activity. Our paper demonstrates that the long-run effects on crime are much larger, which occurs because retention leads students to accumulate less non-cognitive skill over time. Furthermore, the effect of grade retention on adult crime we identify is not the result of a spillover effect; i.e. it is not due to the continuation of crime commission that started as a juvenile (before the age of majority). Our paper highlights the importance of directly examining longer-run outcomes, such as adult criminal convictions,

⁴Ou and Reynolds (2010) and Fruehwirth et al. (2016) provide evidence that the effect of grade retention varies as a function of when students are retained.

⁵Other work has found evidence of a link between high school attainment and juvenile crime: Anderson (2014) shows that increasing the high school dropout age leads to a reduction in juvenile arrest rates.

⁶For example, researchers have shown that short-run effects do not predict long run effects when examining head start (Ludwig and Miller 2007), class size (Chetty et al. 2011), school choice (Deming et al. 2013; Beuermann and Jackson 2019), teachers unions (Lovenheim and Willén 2019), and Medicaid (Cohodes et al. 2016).

using a design that can plausibly account for selection bias. To our knowledge, we are the first to be able to estimate this parameter.

Second, our analysis contributes to a small literature on the effect of education on crime.⁷ Lochner and Moretti (2004) use state changes in compulsory schooling laws to estimate the effect of educational attainment on both arrests and incarceration. They find that an extra year of completed schooling reduces incarceration rates for males ages 20-60.⁸ Similarly, Machin et al. (2011) exploit a 1972-73 increase in the minimum schooling age in England and Wales. They find that a one-year increase in average schooling levels reduces conviction rates for property crimes without having any statistically significant impact on violent crime conviction rates. Hjalmarsson et al. (2015) use a Swedish compulsory school reform to identify the effect of years of education on crime. Although baseline results indicate that more schooling reduces criminal convictions, there is non-negligible heterogeneity along the dimensions of gender, age, and crime type.

While similar in many ways to the parameter we estimate, students who do not complete high school because of compulsory schooling laws likely differ from those failing the promotion exam in important dimensions related to cognitive and non-cognitive skills. Additionally, as we show in this paper, grade retention can impact crime through mechanisms other than altering the number of years of schooling. This underscores the importance of providing direct evidence on how test-based promotion policies impact adult crime.⁹

Further evidence on the effect of education on crime comes from the school choice literature: Deming (2011) shows that an increase in school quality due to winning a school choice lottery in Charlotte-Mecklenberg leads to lower levels of crime 7 years after assignment. These estimates are identified off of clearly exogenous lottery-based variation, but the affected students tend to be higher-achieving than those who are on the margin of being retained in our setting.¹⁰ Thus, the school choice results do not necessarily provide insights into potential effects of test-based grade retention on adult criminal activity.

⁷See Lochner (2011) for a more detailed review of the literature on the link between schooling and crime.

⁸Oreopoulos and Salvanes (2011) reproduce these results using an expanded sample and a slightly altered specification.

⁹Stephens and Yang (2014) cast doubt on the validity of the difference-in-difference approach surrounding changes in compulsory schooling laws based on evidence that states that change these policies experience different secular trends from those that do not. However, they lack the data to explicitly examine the common trends assumption with respect to crime outcomes.

¹⁰Cullen et al. (2006) show similar estimates among high school students who win a school choice lottery in Chicago. They use self-reported crime data, however, and do not have information on criminal activity after high school. Jackson (forthcoming) also presents evidence that single-sex schooling in Trinidad and Tobago reduces juvenile arrests.

Our main findings indicate that 8th grade retention increases the likelihood of being convicted of a violent crime and the number of violent crimes of which individuals are convicted. Thus, the test-based promotion policy causes harm to a subset of students, and the effect persists into adulthood. Because violent crimes have large externalities associated with them, this effect reduces or may reverse any social benefits to these policies. We calculate that increase in violent crimes, driven by 8th grade retention, implies a social cost between \$2.6 to \$18.4 million over the three cohorts of students we examine. Whether test-based promotion has other offsetting benefits that may make it cost-beneficial is beyond the scope of our analysis, but the size of the crime effect is such that any benefits would have to be large to overcome the costs from more crime.

Another policy implication relates to where the cutoff for promotion is located. At the current location of the pass-fail cutoff, there is a long-run effect on criminal behavior, and our estimates based on the Angrist and Rokkanen (2015) extrapolation method suggest we would obtain similar effects with a different cutoff. Our results do not address whether there is an optimal cutoff that balances any positive effects of grade promotion with the effects on crime and high school dropout. This is an important public policy question highlighted by our findings. Finally, our results point to specific types of students who are being harmed by this policy. These students could be targeted with additional resources, which could perhaps mitigate the effects we identify.

The rest of this paper is organized as follows: Section 2 presents the policy background, Section 3 describes the data, and Section 4 shows the estimating equation and discusses identification. Results are shown in Section 5, and Section 6 concludes.

2 Policy Background

The Louisiana School and District Accountability system was adopted by the state's Board of Elementary and Secondary Education (BESE) in June 1998. The adoption of this policy was four years before the federal No Child Left Behind Act; Louisiana was an early mover in terms of adopting accountability policies. A core component of the policy focused on school accountability, with the state setting ten and twenty year goals for all public schools and requiring schools

to demonstrate progress toward these goals. As part of the overall accountability system, the BESE also included a student-focused accountability provision that was designed to end “social promotion,” whereby students were passed to the next grade regardless of their school performance. Under the new test-based promotion policy, students in eighth grade are required to score at predefined levels on the Louisiana Educational Assessment Program (LEAP) tests for both English Language Arts (ELA) and math to advance to the next grade.¹¹

LEAP tests are criterion-referenced tests and are designed to directly align with the state content standards. A student’s LEAP score can be expressed as either a continuous scale score ranging from 100 to 500 points or as a discrete achievement level ranging from unsatisfactory to advanced. Students must score at least Approaching Basic in both subjects to advance to the next grade. This is equivalent to 269 and 296 scale points in ELA and math LEAP tests, respectively.¹² In addition to LEAP tests, students in earlier middle school grades were given Iowa Tests of Basic Skills (ITBS). The ITBS is a low-stakes norm-referenced test for which scores are compared to a national norm group.¹³ All tests are first administered in mid-March.

Students who fail to achieve the promotional standards in March are required to retake each failed subject exam in July. The school districts must offer, at no cost, a minimum 50 hours per subject of summer remediation in ELA and math to students who fail to meet the passing standards in March. The school districts are given the flexibility to determine curriculum used in summer remediation classes, but the summer programs are monitored by the state on a regular basis. Evidence from the annual summer school remediation reports and monitoring visits suggest that teachers in summer schools are proficient in the content area in which they are teaching and that they are using a variety of creative teaching strategies including but not limited to small group instruction, use of hands-on materials, and problem solving teams. However, there are large classrooms in some summer programs (around 20 students per classroom), which has raised concerns about their effectiveness (Pastorek 2010).

Students are not required to attend summer school programs to be eligible for the July testing. Those who pass the July exams move on to the next grade. Students who again fail

¹¹Similar test-based promotion policies are implemented in fourth grade as well.

¹²Raw scores are transformed to scaled scores in a three stage process. They are first mapped onto the Item Response Theory scale. They are then converted to a reporting scale and finally, they are equated to reflect the differences in item difficulty.

¹³Beginning in Spring 2006, the state replaced ITBS with *integrated* Louisiana Educational Assessment Program (*i*LEAP), a low-stakes test consisting of both criterion-referenced and norm-referenced components.

are required to repeat the grade unless they are permitted to move to the next grade through a waiver. A school system may grant waivers for students following extenuating circumstances, such as physical illness or court-ordered custody issues. Similarly, students may receive waivers because of local education agency errors or other unique situations not covered under extenuating circumstances. The state also may grant a waiver to LEP students.¹⁴ Conversely, students can fail a grade if state-mandated attendance requirements are not met, irrespective of LEAP scores. These institutional features lead us to estimate fuzzy regression discontinuity models, since failing or passing one of the July LEAP exams does not necessarily determine grade retention.

3 Data

We use data from a merge of administrative education records and administrative criminal records that allow us to observe rich information about the entire population of Louisiana public school students and those convicted of a crime. The data for this study come from two different sources. The first is administrative records from the Louisiana Department of Education (LDOE). These administrative data include student demographic information such as gender, race and free/reduced lunch status, as well as scores from the LEAP and ITBS tests. Unique state identification numbers allow us to track all the students through their tenure in the public school system and therefore, we are able to identify schools in which each student was enrolled. The data also include whether a student has dropped out without receiving a diploma, the number of days absent in each school year, and the number of disciplinary incidents. Note that because we observe all public school enrollments in the state, attrition only arises if students leave the state, attend a private school, or are home schooled. It is plausible that students who fail the July exam respond by leaving the state or enrolling in a private school, but as we show below, failure is uncorrelated with the likelihood of attriting from the public education sample. Furthermore, only leaving the state causes attrition in our crime estimates, as private school and home schooled students are included in the conviction data.

¹⁴Unfortunately, there were not well defined waiver policies over our sample period. Thus, while we can observe non-compliance in the data, we do not know why a student is non-compliant.

The second source of data is the Louisiana Department of Public Safety and Corrections, Adult Services. These data span from 1996 through 2012 and are merged with the LDOE data using common state identification numbers from both datasets. The crime data include basic demographic information, the exact type of crime committed, and sentence type (incarceration or probation) for each conviction. Importantly, only those convicted of a crime are in this dataset;¹⁵ those who are arrested but never charged or found not guilty are not recorded. Thus, these data include the most serious offenses that led to criminal convictions. While 91% of property crime and 87% of drug-related crime convictions are given probation, only 55 percent of violent crime convictions receive probation. Thus, serving prison time is common for violent crime convictions. The high rate of probation for drug-related offenses suggest that most of these offenses are related to drug use rather than drug dealing.

Our outcome of interest is adult conviction at age 25 or earlier. We use this age cutoff to be consistent with the literature (Aizer and Doyle 2015), because most crime is committed by those 25 and younger (Hirschi and Gottfredson 1983), and because of the timing of our education data. In order to measure criminal activity without any censoring, we limit our focus to eighth graders from the 1998-1999 to the 2000-2001 academic years, roughly corresponding to the cohorts born between 1984 and 1987. Also, our analysis sample is composed solely of students who fail to meet the March promotional cutoff and took the July exams. This is important because summer school can have independent effects on student outcomes (Matsudaira 2008). That both retained and promoted students had equal access to summer school means our estimates are net of any summer school effect. Our final analysis sample consists of 22,929 unique student observations. The total number of students in this sample roughly corresponds to 15% of the universe of eighth graders from the 1998-1999 to the 2000-2001 academic years.

Table 1 presents descriptive statistics for the July testing sample. We show tabulations for the full sample as well as by retention status. Because we are focused on low-performing students, the sample disproportionately consists of disadvantaged students and average test scores are lower than the state average. For example, 45% of the sample is on free/reduced

¹⁵One concern is whether people who were retained because they failed the promotion exam are more likely to leave the state as adults. While we cannot directly test for such attrition, that we do not observe any differential attrition out of the education sample due to failing the promotion exam suggests that it is not a concern in the crime data either. Tabulations from the American Community Survey show that about 5% of those born between 1980 and 1985 in Louisiana moved out of the state between ages 18-25. This rate is just above 2% for those with at most a high school degree.

prices lunch and the mean student in the full sample scores 0.86 and 0.69 standard deviations below the state average in math and ELA, respectively. Seventy-six percent of the sample is black. Disproportionate representation of black students and low test scores are even more pronounced for the sample of retained students, as shown in column (5). The math exam is more of an obstacle for promotion than the ELA test, as only 5% of the retained sample is retained solely because of ELA failure. Note that about 5 percent of the retained students in eighth grade during our sample period passed the July exams but were retained for other reasons, such as not meeting the minimum school attendance requirements. This is why the 8th Grade LEAP test outcomes only sum to 95% in column (5). Furthermore, about 35% of those promoted have also failed one or both of the tests of the July exam but were given exemptions for various reasons and were allowed to advance (see Section 2).

The mean dropout rate for the full sample is quite high, at 45.8%.¹⁶ This reflects the low academic achievement level of the summer school group. Among those retained, the dropout rate is even higher, at 53.5%. The much higher dropout rate among those who are retained likely reflects the lower academic achievement level of this group as well as any causal effect of retention. Our regression discontinuity approach is designed to overcome this selection problem. We also find that 4 percent of these students were convicted of a crime as a juvenile. Not surprisingly, juvenile conviction rates are also higher for those who were retained in eighth grade.

The final set of descriptive statistics in Table 1 relate to adult crime. We show that 11.5 percent of the students are convicted as adults by age 25. The age at first-time adult conviction is almost 21, which underscores the importance of examining outcomes into the mid-20s. About 2 percent are convicted of a violent crime, 4 percent of a property crime, 5 percent of drug-related crime, and 1 percent of another type of crime.¹⁷ Note that because an individual may have been convicted of more than one crime, the sum of all crime categories is slightly greater than the overall conviction rate. Second-degree battery, simple robbery and armed robbery are the most common violent crime convictions, making up more than 41 percent of violent offenses. Simple burglary, theft and illegal possession of stolen things make up 57 percent of

¹⁶Drop out statistics do not include students who transferred to private/home school.

¹⁷This last category (other crimes) is a very heterogeneous group and includes offenses ranging from hit and run driving to aggravated incest, from operating a vehicle intoxicated to perjury.

all property crime convictions, and almost all drug related convictions fall under the category of possession and distribution of drugs. Interestingly, while those who are promoted are less likely to be convicted of a crime by age 25, the differences across the groups is small as are the distributions of crime types.

4 Empirical Methodology

4.1 Reduced Form and IV Models

The goal of this analysis is to estimate the effect of grade retention on crime during young adulthood. This parameter is difficult to identify, because students who are retained have lower academic achievement, are more likely to be minorities, and tend to come from lower-income households.¹⁸ These characteristics also are correlated with the likelihood of having a criminal conviction as an adult. Even a rich conditioning set is unlikely to fully control for selection into retention, since students who are retained probably have unobserved skills and traits (e.g., patience) that make them more likely to commit a crime. This makes the assumptions underlying selection-on-observables models very strong.

In order to obtain credible estimates of the effect of grade retention on adult crime convictions, we rely on the exogenous variation generated by the accountability policy in Louisiana in a regression discontinuity framework. Assuming that there is local randomness to test scores, all observed and unobserved characteristics of students should move smoothly through the July test score cutoff. Evidence consistent with this assumption is presented in Section 4.2. We estimate reduced form models that show the effect of barely failing one of the July exams. We also estimate two stage least squares models in which we use the discontinuity created by the accountability policy as an instrument for grade retention among students who failed the March exam. Note that because we use the sample of students who failed the March exam, our estimates should not be confounded by the presence of summer school: those on both sides of the July cutoff had equal access to summer school, and it is unlikely that summer school

¹⁸In our setting, minority students are either black or Hispanic. However, because of the demographic makeup of Louisiana, virtually all minority students in our sample are black.

takeup was differential across the cutoff.¹⁹ The reduced-form model is as follows:

$$Convicted_i = \gamma_0 + \gamma_1 FP_i + f(Index_i; \delta) + X_i' \theta + \mu_i, \quad (1)$$

where $Convicted_i$ is either an indicator variable that takes the value of one if student i is convicted of a crime by the age of 25 or is the count of first-time convictions by age 25 (including zeros). The first-time conviction count is the number of crimes for which an individual was convicted when he was found guilty by a court the first time as an adult (until age 25). This can be greater than one because people sometimes are convicted of multiple offenses at the same time. Our two outcome measures thus capture somewhat different margins of criminal engagement. The variable FP_i is an indicator that takes the value of one if student i scored below Approaching Basic in either of the July ELA or Math LEAP exams. That is, FP_i takes the value of one if the minimum of the difference between subject-specific July LEAP scores and their respective relevant cutoffs is negative. $Index_i$ is the running variable and denotes the minimum of the subject-specific distances from the respective cutoffs:

$$Index_i = \min[S_i^{Math} - Cutoff^{Math}, S_i^{ELA} - Cutoff^{ELA}], \quad (2)$$

where S_i and $Cutoff$ are the LEAP score and the relevant cutoff in math and ELA. The functional form between $Index_i$ and adult criminal convictions is described by the polynomial function $f(\cdot)$. Our baseline models employ a linear spline in $Index_i$, as suggested by Gelman and Imbens (2017), and we show that the results are robust to using quadratic splines and to the use of local linear estimation. The vector X_i is a set of observed covariates that include indicators for gender, race, free/reduced lunch and immigrant status as well as year fixed effects. In some specifications, we also control for 8th grade school fixed effects, for composite test scores in ELA and math from the 7th grade ITBS exams,²⁰ and for the 8th grade March exam. These test scores are standardized against the statewide mean and standard deviation, separately by

¹⁹Unfortunately, we do not have data on which students took up summer school to directly test this assertion because the administrative data do not include any information on summer school attendance. However, aggregate information from the annual summer remediation reports (Pastorek 2010) and discussions with the LDOE administrators indicate that the participation rate was more than 90 percent among eligible students for both subjects. While summer school could increase the likelihood of passing the test, it is doubtful that summer school participation leads to heaping right over the passing threshold. As long as summer school participation moves smoothly through the July cutoff, summer school is not a confounding factor in this analysis.

²⁰These exam scores are only available beginning with the 1999-2000 academic year.

test year and subject. The key identifying assumption underlying the identification of γ_1 in equation (1) is that unobserved characteristics move smoothly through the cutoff, conditional on the controls for $Index_i$. Under this assumption, for students near the cutoff, the estimate of γ_1 can be interpreted as the intent to treat (ITT) effect of grade retention.

Since we observe the actual retention status of the students in the sample, we also estimate a fuzzy regression discontinuity (RD) model. Using the threshold indicator as an instrument for actual grade retention yields the local average treatment effect (LATE), which is the impact for those students whose retention decisions are causally affected by the state’s promotion policy. Thus, this parameter identifies the effect of grade retention among those who were retained because they failed a July exam and complied with the decision. The two stage least squares models is as follows:

$$Retention_i = \alpha_0 + \alpha_1 FP_i + f(Index_i; \pi) + X'_i \phi + \eta_i \quad (3)$$

$$Convicted_i = \beta_0 + \beta_1 \widehat{Retention}_i + f(Index_i; \lambda) + X'_i \psi + \varepsilon_i, \quad (4)$$

where $Retention_i$ is an indicator variable that takes the value of one if student i is retained in eighth grade and all other variables are as previously defined. Standard errors in the IV and reduced form models are clustered by 8th grade school.²¹ The baseline bandwidth we use is 30 index points on either side of the cutoff. This bandwidth corresponds to 64 percent of our initial sample. We show that our results are robust to using different bandwidths as well.

The variable of interest in equation (4) is β_1 , which represents the LATE of being retained in 8th grade because one failed the high-stakes exam on adult criminal convictions. This parameter is likely to differ from retention effects at earlier ages (Fruehwirth et al. 2016) and from effect of retention that occur for non-test-based reasons. Nonetheless, the LATE estimates we identify pertain to an extremely relevant group of students, as they are the ones impacted by the test-based promotion policy. Our estimates therefore are more likely to generalize to other states and school districts that have similar policies for 8th grade promotion.

The assumptions underlying the identification of β_1 in equation (4) are very similar to those

²¹Our estimated standard errors are similar if we instead cluster at the level of the running variable (see Online Appendix Tables A-3 and A-4). We favor clustering at the school level because of recent work by Kolesar and Rothe (2018) that shows standard error estimation has poor properties when clustering on the running variable in regression discontinuity settings. Furthermore, students in the same school experience similar education inputs and tend to be similar on observed and unobserved dimensions, which makes the 8th grade school a sensible variable on which to cluster.

discussed above for γ_1 in the reduced form model. The additional assumption invoked here is that the entire effect of failing a July exam on adult crime comes through the effect of failing on retention. If there are independent effects of failing to meet the threshold, then β_1 will overstate the impact of grade retention. Such an independent effect could be driven by changes in a student's self-esteem or beliefs about the value of investing in education that respond to failing the exam. We argue it is unlikely that such independent effects exist. As we show below, there is only weak evidence of an effect on juvenile crime, which should be highly impacted by these forces. The non-cognitive responses to failing also take several years to show up, which suggests that any self-esteem effects are not immediate and are likely driven by retention rather than failing the July exam per se. Below, we show both ITT and IV estimates for completeness, since ultimately we cannot test this assumption. The IV estimates are our preferred results, however, and we focus on them in the results section.

The estimates obtained from the RD specification identify effects of retention only among those who are local to the pass-fail threshold. If there are heterogeneous treatment effects by student ability, such estimates may not be representative of the effect of retention for larger groups. To shed light on the generalizability of the estimates to students not local to the cutoff, we employ the method proposed by Angrist and Rokkanen (2015). This method is based on a conditional independence assumption, which states that conditional on a set of observed characteristics, there is no relationship between the running variable and the outcomes of interest on either side of the cutoff. Under this assumption, one can use these observed characteristics to estimate counterfactual outcomes among both the treated and untreated samples. This in turn permits estimation of treatment effects away from the cutoff. We show that the conditional independence assumption holds in our data, which allows us to estimate the causal effect of retention on adult criminal convictions for inframarginal treated and untreated students.

4.2 Validity Checks

The core assumption supporting an RD approach in this context is that students are unable to sort around the cutoff. Given that the running variable consists of test scores, absent direct

cheating behavior it is very unlikely such sorting occurs. In this section, we show evidence that student characteristics move smoothly through the cutoff and that there is no evidence of systematic sorting. Table 2 presents reduced form estimates that exclude observed characteristics as controls and use them as dependent variables. The first column shows that failing a July exam does not affect the likelihood of being in the data. Thus, students are not leaving the state because they fail the test.²² The next five columns show how a series of immutable or pre-exam characteristics vary as a function of the promotion cutoff: female, black, white, free/reduced priced lunch and 7th grade exam scores. All of the estimates are small in absolute value, and none is statistically significant at even the 10% level. Thus, there is no evidence in Table 2 of student sorting differentially across the promotion threshold.

Figure 1 displays the distribution of test scores as a function of the index value. As argued by McCrary (2008), heaping right over the threshold is evidence of sorting since the running variable is expected to be distributed smoothly through the cutoff. The test score distribution exhibits heaping at several relative scores, which is due to the underlying test score distributions. There is no evidence in the figure of excess score heaping right over the cutoff. Indeed, the mass right over the cutoff is lower than the mass right above, but this difference is small and clearly reflects the choppiness of the underlying test score distribution rather than a behavioral response to the cutoff. The test proposed by McCrary (2008) fails to reject the null hypothesis of a continuous distribution (p-value=0.89). The estimates in Table 2 and Figure 1, together with the lack of statistical evidence of heaping above the threshold, support our regression discontinuity approach by showing no evidence of student sorting relative to the promotion cutoffs.

5 Results

5.1 Main Results

First-stage estimates that show how July test failure relates to grade retention are shown in Table 3. All estimates include linear splines in the running variable and use a bandwidth of

²²Note that transferring to a private school does not generate attrition in our criminal conviction data. Thus, the only relevant margin of attrition is out-of-state migration. The LDOE codes reasons students leave the school prior to graduation that are based on U.S. Department of Education guidelines. Reasons for no longer attending include death, attending a private school, being homeschooled, dropping out, or leaving the state. Table 2 focuses on the set of students coded as leaving the state.

30 index points on either side of the cutoff. Column (1) includes only test year fixed effects and the running variable controls, while in column (2) we add student observed characteristics: race, gender, free/reduced price lunch status and immigrant status. In column (3), we include 8th grade school fixed effects in the model.

The estimates across columns are extremely similar and indicate that failing a July exam increases the likelihood of being retained by 68 percentage points. The estimates all are significant at the 1% level, and the resulting first-stage F-statistics are quite large. Figure 2 shows these results graphically. There is a large, significant jump in the likelihood of being retained at the cutoff of 68 percentage points. The estimate is not 100 because some students who pass the test are retained for other reasons, and some students who fail wind up being promoted after they appeal. This is a fuzzy regression discontinuity setting, and the July test score cutoff for promotion is a very strong instrument for grade retention.

Reduced form and second stage estimates are presented in Table 4, using an indicator for being convicted of a crime by age 25 as the dependent variable. The first three columns present reduced form effects (equation 1) while the last three show IV estimates of the effect of grade retention (equation 4). Each set of three results columns is laid out identically to those in Table 3. We examine whether individuals are convicted of any crime and whether they are convicted of a violent crime, a property crime or a drug-related crime. It is possible to be convicted of multiple types of crime, so these categories are not mutually exclusive. Note that dependent variables have been multiplied by 100 for ease of interpretation. The estimates in columns (2) and (5) that include test year fixed effects and observed student characteristics are our preferred estimates, and we focus on them throughout. Our estimates are robust to excluding these covariates and to including school fixed effects; the results change little across columns.

The impact of failing a July exam on any conviction in Panel A is positive but not significant at conventional levels. In column (2), the point estimate suggests that failing a July exam increases the likelihood of being convicted of any crime by 0.85 of a percentage point. This is a 7.36% increase relative to the baseline mean shown in Table 1. The resulting IV estimates also are positive, but similar to the reduced form results, they are not significantly different from zero. Still, they suggest that retention increases the likelihood of an adult criminal conviction

by 1.25 percentage points (10.85% relative to the mean).

The clearest evidence of an effect of grade retention on crime is obtained from examining violent crimes. Failing a July exam increases the likelihood of a violent crime conviction as an adult by 0.71 of a percentage point, which is 39.61% relative to the baseline likelihood of such a conviction. The IV estimate shows that grade retention leads to a 1.05 percentage point (58.44%) higher probability of an adult violent crime conviction. Almost all of these convictions are for assault or robbery; there are very few murder convictions in this sample. There is no evidence of an effect on the likelihood of being convicted of a property crime: estimates are negative, close to zero, and are not statistically significant at even the 10% level. However, we show in Section 5.3 that there is an effect of retention on property crime away from the cutoff.²³ Estimates for drug-related crimes are positive and non-trivial in size, though they are imprecisely estimated and are not statistically significant. The point estimates suggest that failing a July exam increases the likelihood of a drug conviction by 0.35 of a percentage point (6.78%) and grade retention increases the probability of a drug conviction by 0.51 of a percentage point (10.02%). Figure 3 presents these results graphically. The figure aligns well with the results in Table 4: there is an increase in violent crime at the cutoff,²⁴ but the other crime categories show less evidence of an effect.

Examining an indicator for whether any conviction occurs may miss variation coming from the number of convictions upon first conviction. People are sometimes convicted of multiple offenses at once even within a given crime category. For example, someone can be convicted of both battery and armed robbery, which are both violent crimes. One may reasonably interpret such convictions as representing more serious crimes than do convictions for one offense. In Table 5, we therefore present reduced form and IV estimates using the number of convictions (including zeros) when we observe an individual being convicted of a crime for the first time as an adult up to age 25.²⁵ This intensive margin arguably better measures the severity of the crime than does a conviction indicator, however the results are quantitatively and qualitatively

²³We additionally show in Section 5.4 that there are positive effects of retention on robbery convictions. Robberies are technically violent crimes, but they share many characteristics with property crime (such as burglary).

²⁴The discontinuity is difficult to see in Figure 3 because all of the panels are on the same scale. Online Appendix Figure A-5 shows the same figure with different scales across panels; the discontinuity in violent crime is evident.

²⁵We also could examine recidivism using the total number of times an individual was convicted of a crime up to age 25. However, examining re-offending with a sample that imposes an age cutoff of 25 years old will not allow us to measure the full effect of grade retention on re-offending. This follows from the fact that a subset of convicted individuals would still be in prison by the time they reach 25, and the likelihood of incarceration is highest among those who are most likely to re-offend. Examining recidivism in our sample thus may provide misleading estimates.

similar to those in Table 4. Column (2) of Table 5 shows that failing a July exam increases the number of crimes committed by 0.017, which is a 12.50% rise relative to the mean, although the estimate is not statistically significant. As displayed in column (5), being retained increases the number of adult convictions by 0.025, or 18.38%, which also is not significant at even the 10% level.

In Table 5, we again find a large and statistically significant positive effect of July exam failure on violent crime convictions, no effect for property crimes, and a positive but not significant effect for drug convictions. The violent crime effect of failing the July exam is 0.012 additional crimes, which is 60.00% of the mean. Being retained increases the number of violent crimes of which one is convicted by 0.017, or 85.00%. We also find suggestive evidence of a sizable increase in the number of drug convictions: retention increases the number of drug convictions at first conviction by 0.012 (20.00%), but the estimate is not significant at conventional levels.

Graphical representations of the regression discontinuity are shown in Figure 4 for the number of conviction outcomes. They match the estimates in Table 5 closely in showing an increase in the number of criminal convictions, violent crime convictions,²⁶ and drug convictions due to failing a July exam but no effect on property crime convictions.

The results thus far indicate that grade retention increases the likelihood of being convicted of a violent crime and increases the number of violent crime convictions upon first conviction. We also find suggestive evidence of a positive effect on drug convictions. It is important to emphasize the large private costs and externalities associated with these outcomes. The outcomes we examine are convictions, not arrests, which means that the increased criminal activity is both likely to be severe and leads to a permanent criminal record.²⁷

To put the increase in violent crime into perspective, we combine the estimates in Donohue (2009) of the weighted average cost of an assault and robbery with the discontinuity estimate from our preferred specification in column (5) of Table 4. Note that retention increases the probability of being convicted of a violent crime by around 58.44% (1.052/1.800). Taking

²⁶The discontinuity in the number of violent crime convictions is difficult to see in Figure 4 because all of the panels are on the same scale. Online Appendix Figure A-6 shows the same figure with different scales across panels; the discontinuity in violent crime is evident.

²⁷Online Appendix Table A-10 shows suggestive evidence that retention increases sentence length by 3 months (off of a mean of 79 months) among those convicted, however the estimate is imprecise. This is consistent with the severity of crimes committed increasing. The average age of first conviction also rises slightly, as shown in column (2) of the table, but the estimate is again imprecise and thus only suggestive of an effect.

the number of convicted adults from violent crimes in the control group as our benchmark, the percent increase implies that 104 more adults are convicted as a result of retention over our sample period. This increase corresponds to a \$2.6 million social cost using Donohue’s lower bound estimates of assault and robbery and a \$18.4 million cost using the upper bound estimates. Considering that about half of violent incidents are reported to the police and about half of reported incidents result in an arrest, the actual social cost of retention likely is even larger than what is reported here (FBI 2012).

5.2 Mechanisms

There are several potential mechanisms that could generate the impact of grade retention on adult crime. Holding students back a year may affect their accumulation of non-cognitive skills. Indeed, a core argument for “social promotion” is that it is damaging for students to be in grades with those who are younger (Black et al. 2011; Cook and Kang 2016), and qualitative evidence indicates that grade retention is a determinant of student stress (Anderson, Jimerson, and Whipple 2005). If retention reduces human capital accumulation and/or reduces students’ social fit in the school, it could lead to lower academic performance, more behavioral problems, and to eventual dropout from school for some students. Retaining students also can alter the educational environment, in particular the set of peers to which a student is exposed.²⁸ If retention reduces the quality of schooling inputs, student achievement may decline. Additionally, it is possible that being retained sends a signal to the student that they are not well equipped for high school, which could cause them to drop out. Reductions in non-cognitive skills can increase the likelihood young adults engage in criminal behavior as well.

Tables 6-8 provide evidence on some of these potential mechanisms. First, in Table 6, we examine how grade retention affects the composition of students in the high school the student attends. Being retained in 8th grade leads students to attend high schools in which the percentage of white students in 9th grade is 2.7 percentage points lower (significant at the 12% level) and the mean 9th grade test score is 6.2% of a standard deviation lower (significant at the 5% level). This finding indicates that retained students subsequently attend high schools

²⁸The retention policy itself is unlikely to generate peer effects outside of shifting students across schools. While lower-performing students are taken out of each grade, they are added to the grade below. Hence, overall peer composition in a cohort is not affected on average by holding low-performing students back each year.

that have lower-achieving and lower-SES peers.²⁹ The change in peer composition we document likely is facilitated by the large amount of school choice that exists in Louisiana, and resulting adverse peer effects could be a driver of the increases in crime we document.

The development of non-cognitive skills is particularly important to examine, since these skills are highly correlated with the propensity to engage in criminal activity (Carneiro et al. 2007; Jackson et al. 2020). We now turn to an analysis of behavioral outcomes in Table 7 that reflect such skills: absenteeism and disciplinary incidents (Lounsbury et al. 2004). First, we examine whether retention leads to a change in the likelihood of observing information on these outcomes for students. Failing the July exam leads to a small but statistically significant increase in the likelihood of a student having missing data on these behavioral outcomes.³⁰ If anything, this likely biases our estimates towards zero because those with behavioral problems are more likely to have missing data due to attrition. In columns (2) to (4) of Table 7, we estimate effects on students after 1, 2, and 3 years of grade retention to more closely examine the time pattern of any retention impacts. Students who are retained are absent 0.76 more days per year but are actually less likely to have a behavior incident in the first year following failure. Neither estimate is statistically significant, however. By the third year after retention, there is a statistically significant effect on absences of 1.97 days per year (11.14% relative to the mean). Students also are 1.12 percentage points more likely to have a disciplinary incident, although this estimate is not statistically different from zero at conventional levels. The year 3 estimate is statistically different from the year 1 estimate, which suggests grade retention induces a worsening pattern of behavior over time.

In addition to reductions in non-cognitive skill acquisition, grade retention can lead to less education attainment, as shown in Table 8. The first column of Table 8 indicates that students who barely missed the promotion cutoff are not more likely to leave the public school system before their status is determined as a graduating or dropping out student.³¹ Column (2) of

²⁹Cook (2018) shows evidence that attending more segregated schools in which the percentage of African American students is higher can reduce short- and long-run educational attainment.

³⁰Failing the July exam generates no change in the probability of a student having missing data on these behavioral outcomes after the first and second years of retention. In each case, the point estimates are close to zero. This suggests that the attrition effect in Table 7 reflects higher dropout rates as students age rather than students leaving the state.

³¹The LDOE analyzes enrollment data employing an algorithm coupled with U.S. Department of Education guidelines and flags an individual as being a drop out using these guidelines. Specifically, LDOE requires the use of three school years (the previous, current and the following) to identify the dropout status of a student. Therefore, a dropout flag is not complete until after the dropout correction period of the following year. Certain exit codes available in the administrative records exclude individuals from being coded as dropouts (e.g., death, out of state or private/home school) (Pastorek 2011). An individual is identified as a dropout only if he/she is flagged as one in the administrative records.

Table 8 shows that students who are retained in 8th grade due to failing a July exam are 7.17 percentage points more likely to drop out. This is a 15.66% increase relative to the baseline dropout likelihood of 45.8%. Online Appendix Table A-10 additionally shows that retention reduces the number of years of completed education post-July exam by 0.264 years, which is statistically significant at the 1% level.³² Hjalmarrsson et al. (2015) present the only estimates of which we are aware on the effect of educational attainment on violent crime convictions among young adults. They find no effect on the likelihood of committing any violent crime but a negative effect on the number of violent crime convictions for men (but not women). Using their results, we calculate that reduced educational attainment associated with grade retention predicts a 0.003 increase in the number of violent crime convictions.³³ Thus, reduced educational attainment explains only 18% of the total increase in the number of violent crimes and none of the increase in the likelihood of any violent crime conviction stemming from 8th grade retention.

Because it is unclear whether estimates from Scandinavia are generalizable to the US context, we supplement this evidence by estimating a regression of violent crime on a dropout indicator using the full set of controls (i.e., student demographics, school and year effects) for observations in the 30 index point bandwidth around the cutoff. The estimated correlation between violent crime and dropping out is 0.00017. Multiplying by the dropout estimate of 7.171 yields a predicted violent crime conviction effect of 0.001: increased prevalence of high school dropout explains only 12% of the total increase in violent crime stemming from 8th grade retention. These results are consistent with the calculations from Hjalmarrsson et al. (2015) and indicate that high school dropout can explain only a small portion of the crime effect we estimate.

The last two columns of Table 8 present estimates of grade retention on juvenile crime and juvenile felony convictions. Recall that our adult criminal conviction data exclude youth convictions, so there is no mechanical relationship between these two outcomes. The estimates in Table 8 are positive but not significant at even the 10% level. Furthermore, they are much

³²Online Appendix Table A-10 also shows that retention has a positive but not statistically significant effect on being classified as special education.

³³Hjalmarrsson et al. (2015) examine the effect of educational attainment on criminal convictions in Sweden. While a different setting from the US, they emphasize that their findings are similar to US estimates from Lochner and Moretti (2004) for outcomes that overlap across the two papers. Lochner and Moretti (2004) do not examine convictions, nor do they separately estimate models for young adults. We calculate the implied effect of the educational attainment reduction from grade retention by weighting the gender-specific estimates for 19-29 year olds by the gender proportion in our sample and then multiplying by the number of years of reduced educational attainment in Table A-10: $(0.52*0.004+0.48*-0.029)*-0.264=0.003$.

smaller than the IV estimates in Table 4. Understanding the implication of these estimates for our results requires estimating how juvenile crime conviction affects adult criminal convictions. Aizer and Doyle (2015) use random assignment of judges to identify the effect of youth *incarceration* on adult incarceration.³⁴ They find that youth incarceration increases the likelihood of adult incarceration by 23.4 percentage points and of adult incarceration for violent crime by 14.9 percentage points. Using the latter estimate, and making the extreme assumption that all convicted youths are incarcerated, the estimates in Table 8 suggest that the juvenile conviction effect would increase adult violent crime convictions by between 0.050 and 0.079 percentage points. This is almost certainly an upper bound of the juvenile crime effect, as many youths who are convicted of crimes are not incarcerated, and as Aizer and Doyle (2015) demonstrate the effect on adult criminal activity of a conviction is lower when it is not accompanied by incarceration. Furthermore, Eren and Mocan (forthcoming) show that juvenile incarceration in Louisiana has no impact on future violent crime. Hence, the juvenile crime effects we estimate can explain at most a small amount of the adult conviction effect.³⁵

Taken together, the results from Tables 6-8 indicate that grade retention leads to lower peer quality in high school, increases behavioral problems and absenteeism several years after being retained, and substantially increases the likelihood of dropout. There is little effect on juvenile crime, and any juvenile crime effect is too small to explain our adult conviction estimates. These results suggest that the increase in adult crime is not being driven by earlier engagement in the criminal justice system. We argue that this evidence is most consistent with students reducing investments in human capital over time and accumulating less non-cognitive skills, which makes it more likely they will engage in violent criminal behavior. Indeed, our results are consistent with the decline in violent crime from a summer jobs program (Heller 2014), which also was likely driven by changes in non-cognitive skill.³⁶ Those results relate to juvenile crime; our estimates indicate a similar pattern of effects for adult crime as well.

³⁴Sorensen, Bushway, and Gifford (2019) show that principal-driven student dismissals in middle school lead to more criminal convictions in early adulthood, but they do not examine how juvenile convictions affect subsequent adult convictions.

³⁵See Eren et al. (2017) for a more detailed analysis of the effect of the Louisiana accountability system on juvenile crime. They show that the total effect of summer school and potential grade retention is to lower juvenile crime, but the effect of retention per se is positive and not significantly different from zero.

³⁶Heller et al. (2017) show that several RCTs in Chicago designed to alter the decision-making of youth from disadvantaged backgrounds reduced crime overall and violent crime in particular. They do not find evidence of an effect on traditional non-cognitive skills but rather that the interventions caused students to “slow down” and make more measured decisions.

5.3 Heterogeneous Treatment Effects

While we do not find much evidence of an effect on juvenile crime, it still is possible that the impact of retention on adult crime has an age pattern. In Table 9, we examine age heterogeneity by redefining convictions as occurring only at ages 19+ (column (1)) and 20+ (column (2)). The effect of 8th grade retention on adult violent crime largely carries over to convictions that only occur among those who are between 20 and 25. This finding indicates that much of the adult crime effect occurs among young adults 20 and older. It also suggests that the different findings for adult and juvenile crime are not due to the fact that we observe adults over a longer period of time than we observe juveniles.

The regression discontinuity estimator identifies the effect of retention at the promotion cutoff. This local average treatment effect has clear policy relevance for those proximate to the threshold, but from a policy perspective it is important to understand whether our effects extend to other parts of the test score distribution. If effects are local to the cutoff, then it would be possible to mitigate or eliminate this adverse effect by moving the passing threshold. Conversely, if our results generalize to a larger group of lower-performing students, it means any reasonable test-based retention policy might increase criminal convictions later in life.³⁷ If students below the threshold who are retained also engage in more crime later in life, the size of the externality implied by our estimates is much larger than if the effect is localized to the test score cutoff. In order to examine the extent to which our regression discontinuity estimates generalize to a broader set of lower-performing student, we use the method proposed by Angrist and Rokkanen (2015) to estimate treatment effects away from a discontinuity.

The Angrist and Rokkanen (2015) approach relies on the ability to predict outcomes conditional on the running variable using observed characteristics. Under the “conditional independence” assumption that conditional on the observed characteristics the running variable is no longer related to outcomes, one can use estimated linear relationships between the observables and outcomes to predict counterfactual outcomes away from the discontinuity that can be used to estimate treatment effects in a linear reweighting framework (Kline 2011). Our prediction variables consist of indicators for gender, race, free/reduced lunch status and immigrant status,

³⁷We consider here the set of test-based promotion policies targeted towards lower-performing students. No promotion policy will set thresholds targeted at the upper end of the test score distribution, so the only feasible set of cutoffs available to policymakers is within the 30 point bandwidth we examine.

composite (ELA and math) seventh grade standardized test scores and school fixed effects. Angrist and Rokkanen (2015) suggest a simple test for the conditional independence assumption that entails estimating the relationship between outcomes and the running variable on each side of the cutoff, conditional on the set of observed characteristics. These tests are shown in Appendix Table A-9: the estimates are universally small, and none is statistically significantly different from zero at conventional levels. Thus, the tests strongly support the use of the linear reweighting method to estimate treatment effects away from the test score discontinuity.

Table 10 presents the results from implementing the Angrist and Rokkanen (2015) estimator. The first two columns show the effect of retention among students to the left of the test score cutoff who are retained as a result of the test, and columns (3) and (4) present estimates for those to the right of the cutoff who are not retained. Standard errors in the table are block bootstrapped at the school level using 500 replications. The table shows consistent evidence of positive effects of crime on the treated and untreated that are similar in magnitude to one another. For violent crime, retention increases the likelihood of an adult conviction by 1 percentage point and increases the number of convictions by over 0.01. That the effects are of similar size to each other and to the estimates in Tables 4 and 5 suggests that the effect of retention on violent crime is more general among low-performing students. Thus, any reasonable test-based cutoff is likely to generate an increase in violent crime later in life.

For property and drug crimes in Panels B and C of Table 10, we find effects that are larger than the RD estimates in Tables 4 and 5. These results indicate an effect on property crime conviction of about 1.4 percentage points and an increase in the number of property crime convictions of 0.017. For drug convictions, retention increases the likelihood of conviction by about 1 percentage point and increases the number of convictions by 0.014. Only the property crime estimates are consistently statistically significant, however. The effect of retention on property and drug crimes away from the cutoff is larger than our RD estimates shown in Tables 4 and 5. These findings help resolve the somewhat puzzling result that retention does not affect property crime. The results in Table 10 shows that this is the case only at the cutoff: retention affects property crime convictions among a larger set of treated students away from the cutoff.

Online Appendix Figures A-1 through A-4 show these results graphically and demonstrate

why the RD estimates are much smaller than the estimated effects away from the cutoff for property and drug crimes. The “extrapolated” counterfactual lines tend to have no slope, while the actual fitted lines slope upward or downward away from the cutoffs. As a result, the extrapolated effects grow away from the cutoffs. The local average treatment effects estimated with the RD model understate the impact of test-based promotion on drug and property crime: inframarginal retained students experience large increases in both types of crime convictions later in life. The same is not the case for violent crime, where neither the extrapolated nor the fitted lines have nonzero slopes. This causes the RD and extrapolated effects to be the same size.

The estimates in Table 10 have two important implications for the interpretation of our results. First, the RD estimates for property and drug crimes are understated. When extrapolated away from the cutoff, grade retention has large effects on all types of crime. Second, these results suggest that increasing the passing threshold would not mitigate these effects. In fact, it likely would increase the size of the crime effect as the untreated students would exhibit similarly-sized or larger treatment effects if they were retained. The results from Table 10 indicate that the effect of grade retention on adult crime is large and ubiquitous among the low-performing students we examine.

5.4 Robustness Checks

The results presented in Section 5.1 point to a large increase in violent criminal convictions among adults when they are held back because of failing the 8th grade promotion exam. In this section, we explore the sensitivity of these results to alternative modeling assumptions. In Online Appendix Tables A-1 and A-2, we alter the bandwidth and the modeling of the running variable using criminal conviction and the number of convictions as dependent variables, respectively. Column (1) shows estimates using local linear regression with the Imbens and Kalyanaraman (2012) optimal smoother. In column (2) we alter the baseline bandwidth of 30 index points to 15, while column (3) increases it to 45. Column (4) presents estimates that use a quadratic spline in the running variable with a bandwidth of 60, and finally column (5) employs a global quadratic spline. The estimates for violent crime change very little across

specifications in both tables. The drug crime estimates are somewhat sensitive to specification, with negative estimates in columns (1) and (2). However, these estimates are very imprecise, so they are not inconsistent with our baseline results.

We next examine how estimates are affected by the use of different samples, controls and clustering assumptions. Column (1) in Online Appendix Tables A-3 and A-4 show results that use all convictions up to age 28 for our analysis sample. We only observe outcomes at age 28 for the 1984 birth cohort, at age 27 for the 1984-1985 birth cohorts, at age 26 for the 1984-1986 birth cohorts, and at age 25 for the 1984-1987 birth cohorts. Thus, these estimates are based on an unbalanced panel of ages, but they allow us to examine effects past age 25. The estimates are extremely similar to baseline, which is expected given the lower likelihood of criminal engagement when individuals age into their mid 20s. The remaining results shown in Appendix Tables A-3 and A-4 exclude the regions most affected by Hurricane Katrina and Hurricane Rita in column (2),³⁸ add 7th and 8th grade March composite test score controls in columns (3) and (4), respectively, and cluster standard errors at the level of the running variable (i.e., the index) in column (5).³⁹ These changes to the baseline model have minimal effects on the results: the point estimates and standard errors change little across columns. Thus, our main results and conclusions are not sensitive to these core modeling assumptions.

Online Appendix Tables A-5 and A-6 explores potential heterogeneity by gender. The effects of grade retention on adult conviction are more pronounced for males and the point estimates are generally more precise. This result is sensible because the underlying likelihood that women are convicted of a crime is very low. We next analyze if the violent crime results are driven by robberies. This investigation is motivated by two concerns. First, robbery is the most prevalent crime in the violent crime category. Second, although robbery is a violent crime because it involves the use of a weapon, or the threat of the use of a weapon, robbery creates a

³⁸Hurricane Katrina hit the Gulf Coast in August of 2005 and caused massive flooding and damage in areas of Louisiana around New Orleans, much of which was due to flooding from breached levees in several New Orleans parishes. Hurricane Rita hit Louisiana less than a month after Katrina and caused additional flooding and damage in coastal areas and in parishes most impacted by Katrina.

³⁹Because we find a significant impact on violent crime but no impact on property crime or drug crime, it is reasonable to ask whether this finding is due to detecting a false positive. Thus, it is instructive to examine the sensitivity of the estimated standard errors to multiple comparisons. We used the benchmark models reported in Tables 4 and 5, as well as its alternative functional forms reported in Appendix Tables A-1 and A-2, for both the intensive and extensive margins. Adjusting the p-values following Simes (1986) does not alter the inference. As expected, the adjusted p-values of the estimated coefficients were larger, but they were never greater than 0.09 in any functional form specification for the intensive margin regressions, and it was 0.05 in its benchmark model. The extensive margin adjusted p-value was 0.11 in the benchmark model, but it was around an average of 0.07 in other functional forms displayed in Table A-1. For example, the adjusted p-values for the estimated effect of grade retention from a global quadratic spline are 0.02 and 0.03 for the extensive and intensive margins, respectively.

financial return to the perpetrator, and therefore it is different from other violent crimes such as assault. In fact, it is closer in many ways to a property crime than a violent crime. Estimates of the effect of grade retention on robberies and other violent crimes are similar to one another.

We also examine the effect of failing to achieve the March promotional standards on the likelihood of being convicted of an adult crime and the number of crimes of which individuals are convicted. The results, displayed in Online Appendix Table A-7, show that the impact of March discontinuity is very close to zero.⁴⁰

Thus far, we have used the minimum of the subject-specific distances from the respective cutoffs as our running variable. One may be concerned about the compositional changes of students across the subject-specific points of the running variable. To probe this concern, we use the math test score as the sole running variable.⁴¹ The estimated effects of grade retention on adult criminal convictions are very similar to baseline and are reported in Online Appendix Table A-8.

Finally, we conduct placebo tests in which we estimate our baseline model using false cutoffs at each test score over a range of 50 index points on either side of the true cutoff. We incrementally increase the cutoff from -50 to 50 by one point, and in Figures 5 and 6 we plot the distribution of point estimates. The vertical red line shows the effects from Tables 4 and 5, and we report the percentage of placebo estimates that are larger than the baseline effects on the x-axis of each figure. For any violent crime conviction (Figure 5) as well as for the number of violent crime and drug crime offenses (Figure 6), the RD estimates are in the tails of the distribution of point estimates. In Figure 5, the actual point estimate is larger than all but 1% of the placebo estimates for violent crime, and the point estimate for the number of violent offenses is larger than all but 3% of the placebo estimates in Figure 6. Thus, our violent crime

⁴⁰This finding, coupled with retention results, may imply a null effect of summer school assignment on adult crime conviction. Eren et al (2017) posed a related but different question, analyzed a different outcome, and employed a different sample than ours. Our sample differs from that of Eren et al. (2017) because we focus on a shorter interval of academic years to be able to analyze future (adult crime) of students. Nevertheless, we obtain the same results in our sample when replicating the analysis of Eren et al. (2017). Specifically, they find that failing to achieve the March cutoff had no statistically significant impact on overall juvenile convictions, but that it had a negative and significant impact on juvenile felony convictions (with a point estimate of -0.0026, se=0.0013). They obtained the same results when analyzing felony convictions one year after the test date (-0.0028, se=0.0010). We match these results exactly. In our sample, there is also no statistically significant impact of failing to achieve the March cutoff on juvenile convictions, but failing the test has a (weakly) significant impact on juvenile felony convictions (-0.0025, se=0.0015). We also obtained the same results when analyzing convictions one year after the test date (-0.0032, se=0.0012). Thus, despite using different academic years, we replicate the results of Eren et al. (2017), indicating that the findings reported in our paper are not an artifact of some structural change that may have taken place in Louisiana during that period. Moreover, Eren et al. (2017) showed that the effect of summer remediation programs are not persistent and fade out quickly.

⁴¹Specifically, we replace the running variable from equation (2) with the distance of math test score from its respective cutoff and use pass-fail threshold indicator from math test as instrument for grade retention.

results are driven by changes in criminal convictions at the retention cutoff and provide larger estimates than changes at the vast majority of placebo cutoffs.

6 Conclusion

We present the first estimates in the literature of the effect of grade retention on adult criminal convictions. The data we use come from a merge of administrative education records from the Louisiana Department of Education with administrative data on adult criminal convictions from the Louisiana Department of Public Safety and Corrections, Adult Services. To overcome the endogeneity of student retention, we exploit test score cutoffs imposed by the State of Louisiana for promotion to 9th grade whereby a student must repeat 8th grade if they fail the math or ELA state exam in July. We estimate both reduced form effects of failing the exam as well as the effect of 8th grade retention in an IV framework that imposes the additional assumption that the only reason test failure in 8th grade is related to crime is because of retention.

We first show that failing a July exam increases the likelihood of being retained by about 68 percentage points. Reduced form and IV estimates both show a large effect of test failure and subsequent retention on being convicted of a crime by age 25. Being retained increases the likelihood of being convicted of any crime by 1.25 percentage points, or 10.85% relative to the baseline mean, and increases the number of convictions by 0.025 (18.38%). The effects are largest for violent crime: retention increases the likelihood of a violent crime conviction by 1.05 percentage points (58.44%) and increases the number of violent crime convictions by 0.017 (85.00%). We find no evidence of an effect of retention on property crime and suggestive evidence of an increase in drug-related crime at the cutoff. However, we do find an effect on robbery convictions, which are a form of crime with a motive of financial gain.

The rich administrative data we use allow us to examine mechanisms that drive the effects of retention on adult criminal convictions. We find that retention leads students to subsequently attend high schools with lower-performing and more disadvantaged peers. Our estimates also indicate that retention causes higher levels of absence from school three years later, and there is suggestive evidence of increased disciplinary incidents after 3 years. Retention has a large effect on dropout, increasing the likelihood by 7.17 percentage points (or 15.66% relative to the

mean). We do not find a statistically significant impact on juvenile crime, and the estimate is too small to explain much of the adult crime effect. Thus, the main mechanism is unlikely to be increased juvenile criminal proclivity, which then leads to higher criminal activity as an adult. We argue that the evidence is most consistent with retention lowering the rate of human capital accumulation, most importantly non-cognitive skills, which results in higher levels of criminal engagement in early adulthood. Finally, using the method proposed by Angrist and Rokkanen (2015), we estimate effects away from the test passing threshold and find sizable increases in all types of crime due to grade retention among a broader set of low-performing students.

Our results indicate that retaining students in 8th grade due to failing a promotion exam induces higher criminal activity as an adult among students on the margin of passing as well as among low-scoring inframarginal students. This evidence is consistent with a sizable effect of education on crime, and it suggests that the promotion policy, in its current form, harms at least a subset of students and creates important negative externalities in the form of more violent felonies being committed in the future. Importantly, our results speak to the effect of 8th grade retention, and effects may differ for retention in earlier grades. Understanding whether this relationship exists for retention among younger students is an important question for subsequent research.

While our estimates may not generalize to younger students, that they generalize to a larger set of inframarginal 8th grade students highlights that the effects we report would be detected at any potential promotion cutoff set by the Louisiana Department of Education and is not driven by the particular threshold being set for promotion. The policy implications of these findings are important, as they suggest any test-based retention in 8th grade among low-performing students leads to more criminal engagement later in life. Test-based promotion policies are designed to ensure that students meet basic knowledge levels to advance in school, which may have benefits to both marginal and inframarginal students. Our results show that there are also private and social costs, which are substantial in magnitude. Considering only violent crimes, simple back-of-the-envelope calculations imply a social cost between \$2.6 and \$18.4 million over the three cohorts of students we examine. How to structure test-based promotion policies, or promotion policies more broadly, in a way that retains the benefits from imposing educational

standards while minimizing the costs to students and society is an important question for future work.

References

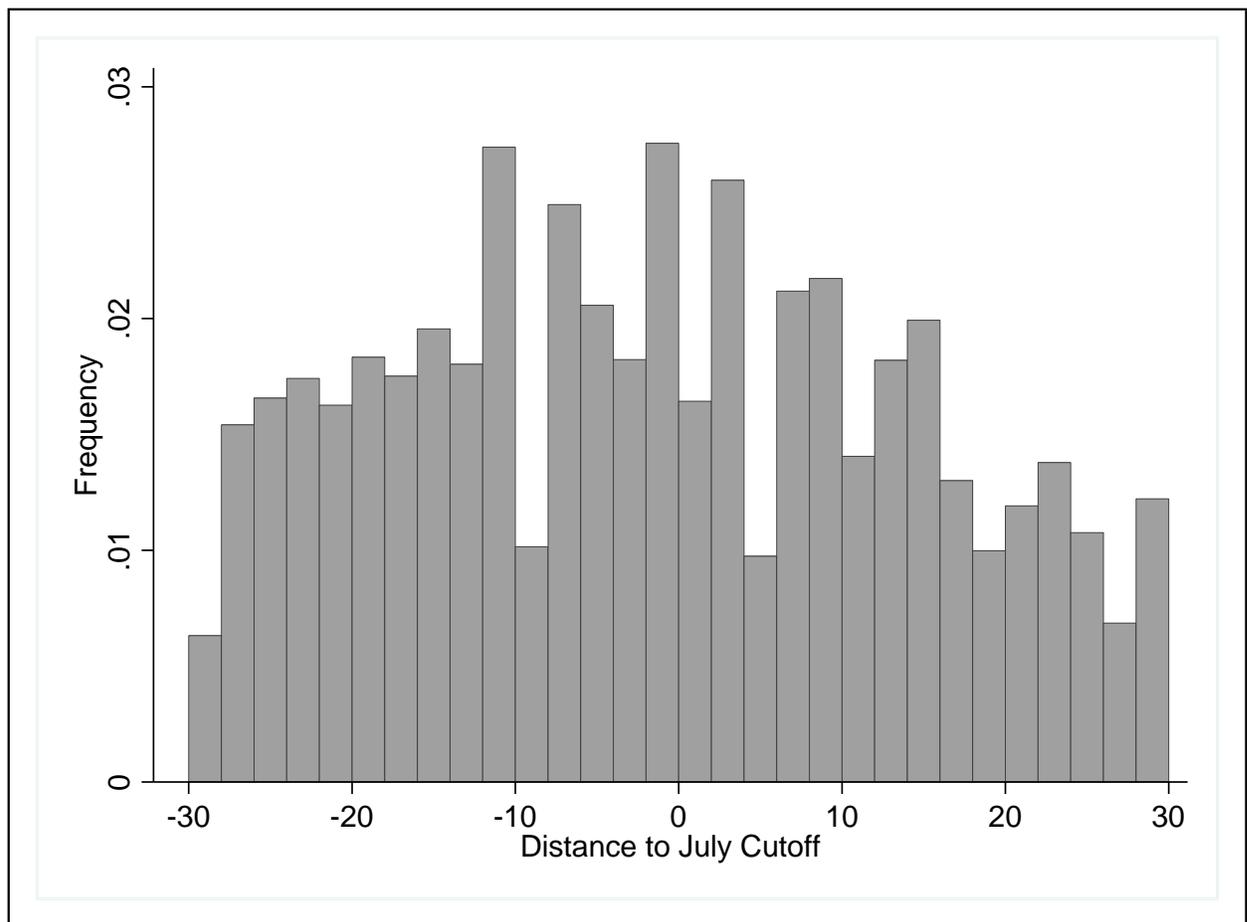
- [1] Agan, Amanda and Sonja Starr. 2017. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment." *Quarterly Journal of Economics* 133(1): 191-235.
- [2] Aizer, Anna and Joseph J. Doyle Jr. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *The Quarterly Journal of Economics* 130(2): 759-803.
- [3] Anderson, David A. 1999. "The Aggregate Burden of Crime." *The Journal of Law and Economics* 42(2): 611-642.
- [4] Anderson, Gabriella E., Shane R. Jimerson, and Angela D. Whipple. 2005. "Student Ratings of Stressful Experiences at Home and School." *Journal of Applied School Psychology* 21(1): 1-20.
- [5] Anderson, D. Mark. 2014. "In School and Out of Trouble? The Minimum Dropout Age and Juvenile Crime." *Review of Economics and Statistics* 96(2): 318-331.
- [6] Angrist, Joshua D. and Miikka Rokkanen. 2015. "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away From the Cutoff." *Journal of the American Statistical Association* 110(512): 1331-1344.
- [7] Becker, Gary S. and Casey B. Mulligan. 1997. "The Endogenous Determination of Time Preference." *Quarterly Journal of Economics* 112(3): 729-758.
- [8] Beuermann, Diether and C. Kirabo Jackson. 2019. "The Short and Long-Run Effects of Attending The Schools that Parents Prefer." Mimeo.
- [9] Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2011. "Too Young to Leave the Nest? The Effects of School Starting Age." *Review of Economics and Statistics* 93(2): 455-467.
- [10] Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69(5): 1127-1160.
- [11] Carneiro, Pedro, Claire Crawford, and Alissa Goodman. 2007. "The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes." CEE Discussion Papers 0092, Centre for the Economics of Education, LSE.
- [12] Charles, Kerwin Kofi and Ming Ching Luoh. 2010. "Male Incarceration, the Marriage Market, and Female Outcomes." *Review of Economics and Statistics* 92(3): 614-627.
- [13] Chetty, Raj, John Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *Quarterly Journal of Economics* 126(4): 1593-1660.
- [14] Chou, Shin-Yi, Jin-Tan Liu, Michael Grossman, and Ted Joyce. 2010. "Parental Education and Child Health: Evidence from a Natural Experiment in Taiwan." *American Economic Journal: Applied Economics* 2(1): 33-61.
- [15] Cohodes, Sarah, Daniel Grossman, Samuel Kleiner and Michael F. Lovenheim. 2016. "The Effect of Child Health Insurance Access on Schooling: Evidence from Public Insurance Expansions." *Journal of Human Resources* 51(3): 727-759.
- [16] Cook, Jason. 2018. "Race-Blind Admissions, School Segregation, and Student Outcomes: Evidence from Race-Blind Magnet School Lotteries." CESifo Working Paper No. 7335.
- [17] Cook, Philip J. and Songman Kang. 2016. "Birthdays, Schooling, and Crime: Regression-discontinuity Analysis of School Performance, Delinquency, Dropout, and Crime Initiation." *American Economic Journal: Applied Economics* 8(1): 33-57.
- [18] Cullen, Julie Berry, Brian A. Jacob and Steven Levitt. 2006. "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." *Econometrica* 74(5): 1191-1230.
- [19] Dee, Thomas S. 2004. "Are There Civic Returns to Education?" *Journal of Public Economics* 88(9-10): 1697-1720.

- [20] Deming, David J. 2011. "Better Schools, Less Crime?" *The Quarterly Journal of Economics* 126(4): 2063-2115.
- [21] Deming, David J., Sarah Cohodes, Jennifer Jennings, and Christopher Jencks. 2013. "School Accountability, Postsecondary Attainment and Earnings." NBER Working Paper No. 19444.
- [22] Donohue, John J. III. 2009. "Assessing the Relative Benefits of Incarceration: Overall Changes and the Benefits on the Margin." In S. Raphael and M. A. Stoll (Eds.), *Do Prisons Make Us Safer? The Benefits and Costs of the Prison Boom*, pp. 269-342. Russell Sage Foundation.
- [23] Eren, Ozkan, Briggs Depew and Stephen Barnes. 2017. "Test-based Promotion Policies, Dropping Out, and Juvenile Crime." *Journal of Public Economics* 153: 9-31.
- [24] Eren, Ozkan and Naci Mocan. Forthcoming. "Juvenile Punishment, High School Graduation and Adult Crime: Evidence from Idiosyncratic Judge Harshness." *Review of Economics and Statistics*.
- [25] Federal Bureau of Investigation. 2012. "Crime in the United States 2012," Uniform Crime Report-U.S. Department of Justice: Washington, DC.
- [26] Ferreira, Eduardo Ferraz Castelo Branco and Rodrigo Reis Soares. 2018. "Socially Optimal Crime and Punishment." Escola de Economia de Sao Paulo Dissertation Thesis.
- [27] Fruehwirth, Jane Cooley, Salvadore Navarro, and Yuya Takahashi. 2016. "How the Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatment Effects." *Journal of Labor Economics* 34(4): 972-1021.
- [28] Fudenberg, Drew and David K. Levine. 2006. "A Dual-self Model of Impulse Control." *American Economic Review* 96(5): 1449-1476.
- [29] Gelman, Andrew and Guido Imbens. 2017. "Why High-order Polynomials Should not be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics*.
- [30] Greene, Jay P. and Marcus A. Winters. 2007. "Revisiting Grade Retention: An Evaluation of Florida's Test-Based Promotion Policy." *Education Finance and Policy* 2(4): 319-340.
- [31] Grossman, Michael. 2006. "Education and Nonmarket Outcomes." In E. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 10, pp. 577-633. Elsevier.
- [32] Heller, Sara B. 2014. "Summer Jobs Reduce Violence Among Disadvantaged Youth." *Science* 346(6214): 1219-1223.
- [33] Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack 2017. "Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago." *Quarterly Journal of Economics* 132(1): 1-54.
- [34] Hirschi, Travis and Michael Gottfredson. 1983. "Age and the Explanation of Crime." *American Journal of Sociology* 89(3): 552-584.
- [35] Hjalmarsson, Randi, Helena Holmlund, and Matthew J. Lindquist. 2015. "The Effect of Education on Criminal Convictions and Incarceration: Causal Evidence from Microdata." *The Economic Journal* 125(587): 1290-1326.
- [36] Imbens, Guido and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79(3): 933-959.
- [37] Jackson, C. Kirabo. Forthcoming. "Can Introducing Single-Sex Education into Low-Performing Schools Improve Academics, Arrests, and Teen Motherhood?" *Journal of Human Resources*.
- [38] Jackson, C. Kirabo, Shanette C. Porter, John Q. Easton, Alyssa Blanchard, and Sebastin Kiguel. 2020. "School Effects on Socio-emotional Development, School-based Arrests, and Educational Attainment." NBER Working Paper No. w26759.
- [39] Jacob, Brian A. and Lars Lefgren. 2009. "The Effect of Grade Retention on High School Completion." *American Economic Journal: Applied Economics* 1(3): 33-58.

- [40] Johnson, Rucker C. and Steven Raphael. 2009. "The Effects of Male Incarceration Dynamics on Acquired Immune Deficiency Syndrome Infection Rates among African American Women and Men." *The Journal of Law and Economics* 52(2): 251-293.
- [41] Kline, Patrick. 2011. "Oaxaca-Blinder as a Reweighting Estimator." *American Economic Review* 101(3): 532-537.
- [42] Kolesr, Michal and Christoph Rothe. 2018. "Inference in Regression Discontinuity Designs with a Discrete Running Variable." *American Economic Review* 108(8): 2277-2304.
- [43] Lleras-Muney, Adriana. 2005. "The Relationship Between Education and Adult Mortality in the United States." *Review of Economic Studies* 72(1): 189-221.
- [44] Lochner, Lance. 2004. "Education, Work, and Crime: A Human Capital Approach." *International Economic Review* 45(3): 811-843.
- [45] Lochner, Lance. 2011. "Nonproduction Benefits of Education: Crime, Health, and Good Citizenship." in Eric A. Hanushek, Stephen J. Machin and Ludger Woessman (Eds.) *Handbook of the Economics of Education, Volume 4*. Elsevier: Amsterdam
- [46] Lochner, Lance and Enrico Moretti. 2004. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review* 94(1): 155-189.
- [47] Lounsbury, John W., Robert P. Steel, James M. Loveland, and Lucy W. Gibson. 2004. "An Investigation of Personality Traits in Relation to Adolescent School Absenteeism." *Journal of Youth and Adolescence* 33(5): 457-466.
- [48] Lovenheim, Michael F. and Alexander Willén. 2019. "The Long-run Effect of Teacher Collective Bargaining." *American Economic Journal: Economic Policy* 11(3): 292-324.
- [49] Machin, Stephen, Olivier Marie, and Suncica Vujic. 2011. "The Crime Reducing Effect of Education." *Economic Journal* 121(552): 463-484.
- [50] Manacorda, Marco. 2012. "The Cost of Grade Retention." *Review of Economics and Statistics* 94(2): 596-606.
- [51] Mariano, Louis T., Paco Martorell, and Tiffany Tsai. 2018. "The Effects of Grade Retention on High School Outcomes." Mimeo.
- [52] Marsh, Julie A., Daniel Gershwin, Sheila N. Kirby, and Nailing Xia. 2009. "Retaining Students in Grade-Lessons Learned Regarding Policy Design and Implementation." RAND Education. Technical Report.
- [53] Matsudaira, Jordan. 2008. "Mandatory Summer School and Student Achievement." *Journal of Econometrics* 142(2): 829-850.
- [54] McLaughlin, Michael, Carrie Pettue-Davis, Derek Brown, Chris Veeh, and Tanya Renn. 2016. *The Economic Burden of Incarceration in the US*. Washington University: Seattle.
- [55] Oreopoulos, Philip. 2007. "Do Dropouts Drop Out Too Soon? Wealth, Health and Happiness from Compulsory Schooling." *Journal of Public Economics* 91(11-12): 2213-2229.
- [56] Oreopoulos, Philip and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25(1): 159-184.
- [57] Ou, Suh-Ruu and Arthur J. Reynolds. 2010. "Grade Retention, Postsecondary Education, and Public Aid Receipt." *Education Evaluation and Policy Analysis* 32(1): 118-139.
- [58] Pastorek, P. 2010. "LEAP Remediation Programs." Louisiana Department of Education.
- [59] Schwerdt, Guido, Martin R. West and Marcus A. Winters. 2017. "The Effects of Test-based Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida." *Journal of Public Economics* 152: 154-169.
- [60] Sorensen, Lucy C., Shawn Bushway, and Elizabeth J. Gifford. 2019. "Getting Tough? The Effects of Discretionary Principal Discipline on Student Outcomes." Mimeo.

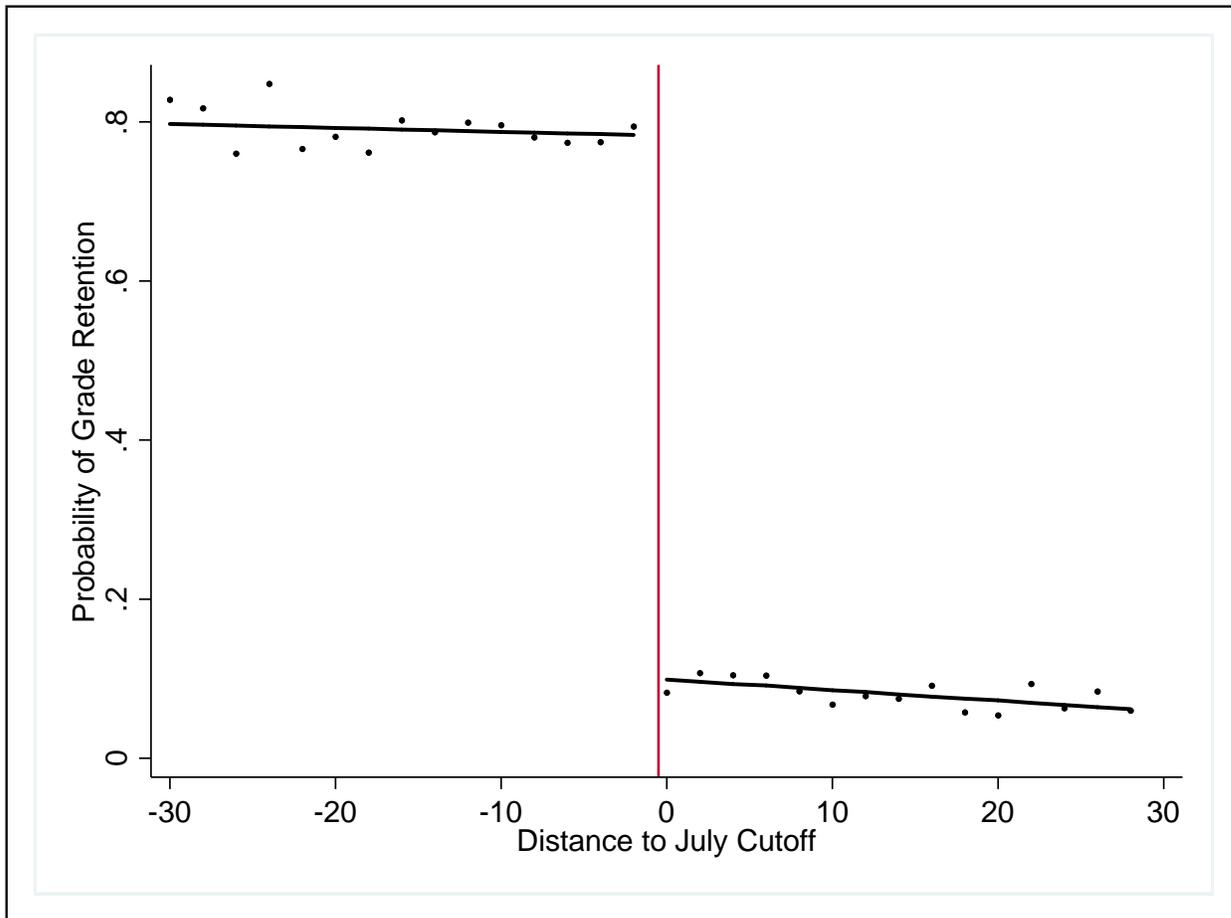
- [61] Stephens, Melvin Jr. and Dou-Yan Yang. 2014. "Compulsory Education and the Benefits of Schooling." *American Economic Review* 104(6):1777-1792.
- [62] Workman, Emily. 2014. *Third-Grade Reading Policies*. Education Commission of the States.
- [63] Zinth, Kyle. 2005. *Student Promotion/Retention Policies*. Education Commission of the States.

Figure 1: Distributions of Students around the July Promotional Cutoff



The figure shows the distribution of index scores relative to the July promotional cutoff (centered at zero), using the analysis sample with a bandwidth of 30 as described in the text.

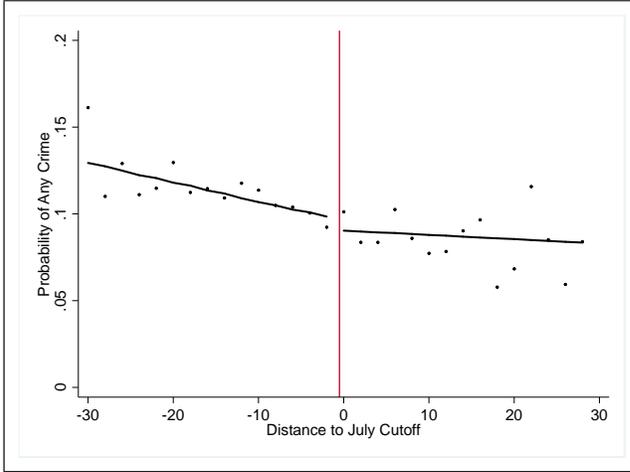
Figure 2: Probability of Grade Retention and Distance to the July Promotional Cutoff



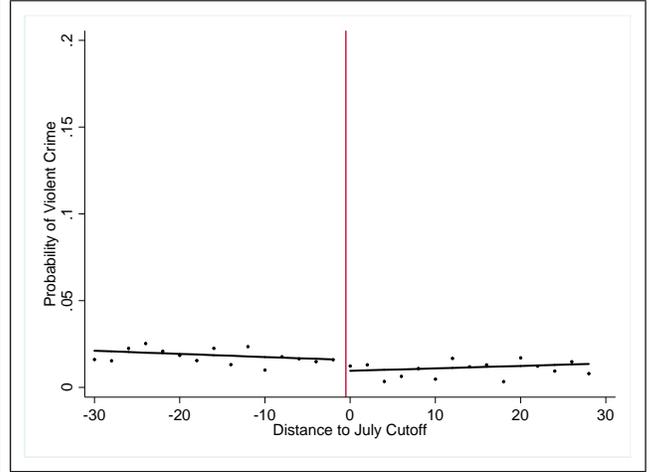
The vertical line denotes -0.5 points to the left of the July promotional cutoff (centered at 0). Each circle represents the unconditional mean of grade retention in two index point bins, based on the distance to July cutoff. The solid lines are fitted values of probability of grade retention from a linear spline over an index bandwidth of 30 points.

Figure 3: Any Criminal Convictions by Distance to the July Promotional Cutoff

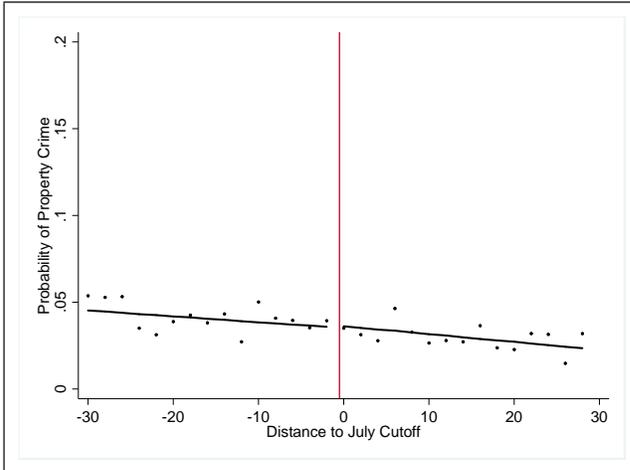
Panel A: Any Crime



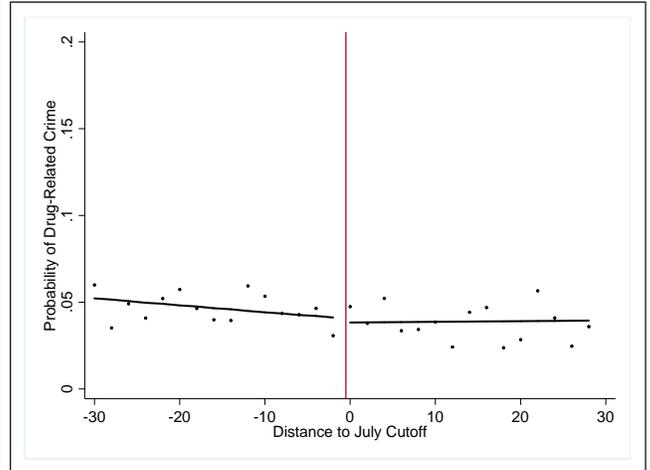
Panel B: Violent Crimes



Panel C: Property Crimes



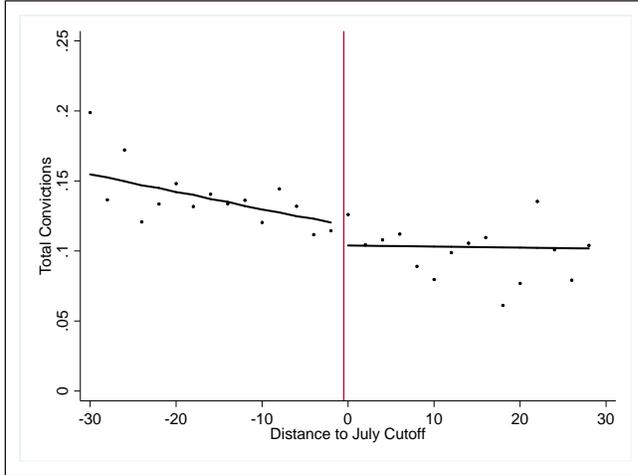
Panel D: Drug-Related Crimes



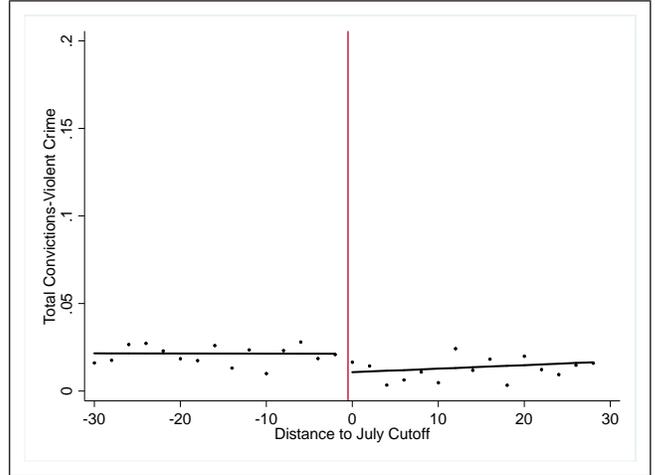
The vertical line denotes -0.5 points to the left of the July promotional cutoff (centered at 0). Each circle represents the unconditional mean of adult crime conviction and types in two index point bins, based on the distance to July cutoff. The solid lines are fitted values of probability of criminal conviction from a linear spline over an index bandwidth of 30 points.

Figure 4: The Number of Adult Crime Convictions by Distance to the July Test Cutoff

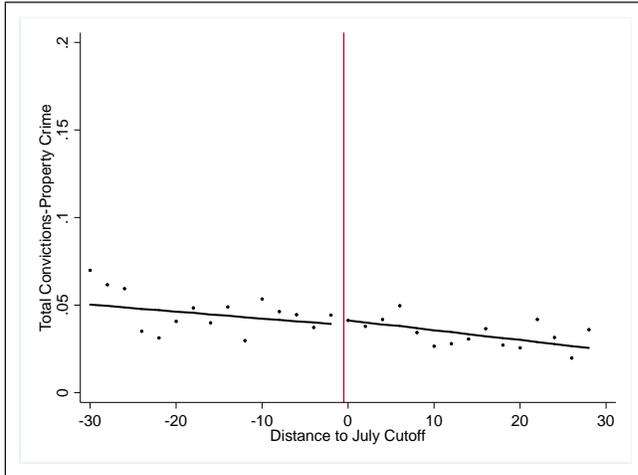
Panel A: Any Crime



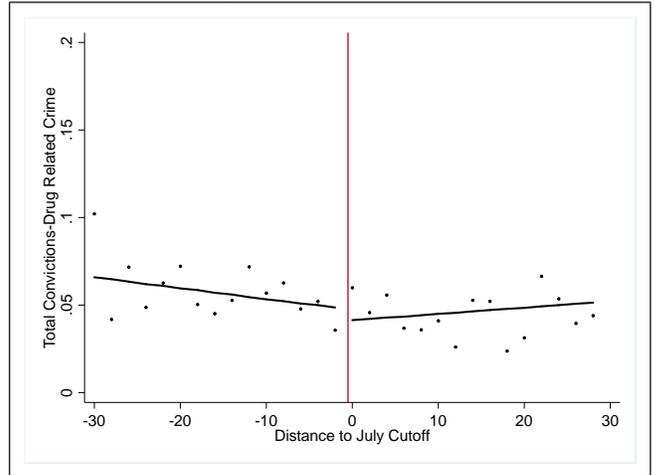
Panel B: Violent Crimes



Panel C: Property Crimes



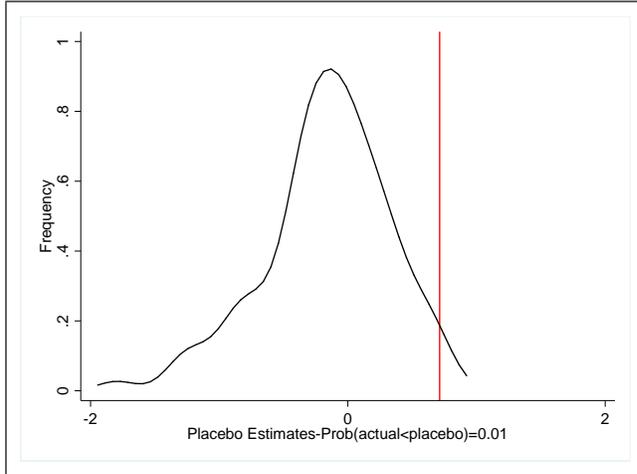
Panel D: Drug-Related Crimes



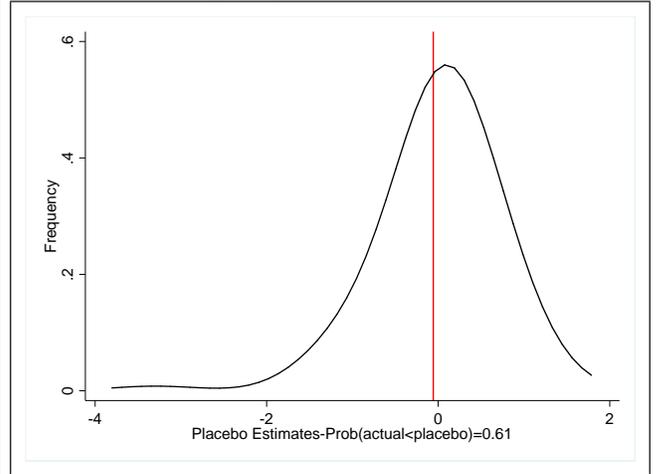
The vertical line denotes -0.5 points to the left of the July promotional cutoff (centered at 0). Each circle represents the unconditional mean of the number of adult crime convictions, overall and by type, in two index point bins based on the distance to July cutoff. The solid lines are fitted values of the number of criminal convictions from a linear spline over an index bandwidth of 30 points.

Figure 5: Placebo Estimates of the Effect of Failing a July Promotional Cutoff on the Probability of Adult Criminal Conviction

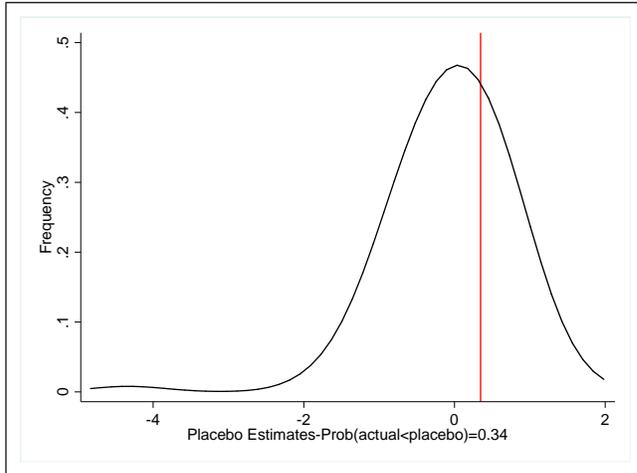
Panel A: Violent Crimes



Panel B: Property Crimes



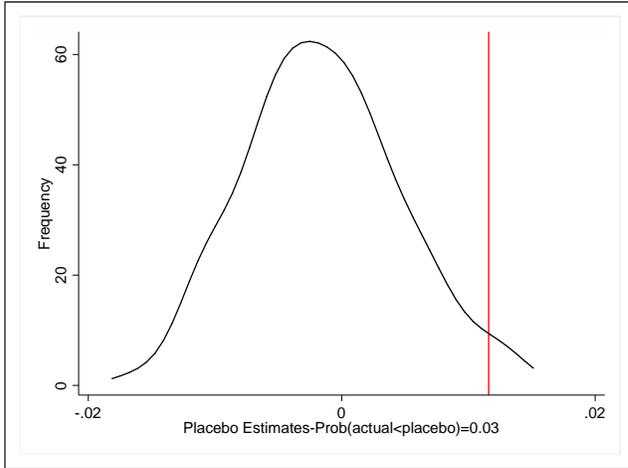
Panel C: Drug-Related Crimes



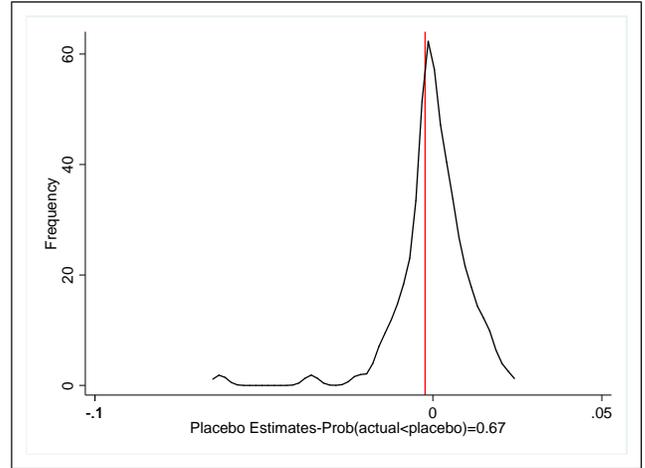
Each placebo estimate assigns a false promotion cutoff by incrementally adding one scale point to the subject-specific cutoffs (269 and 296 scale points in ELA and math LEAP tests, respectively) over a range of [-50, 50] scale points. A reduced form equation is then employed to estimate the effect of failing to meet the July cutoff on the probability of being convicted of different types of adult crime. All estimates are obtained from a linear spline using a bandwidth of 30 index points. The vertical lines denote the actual estimates. The fraction of placebo estimates larger than the actual estimate is also reported on the x-axis of each graph.

Figure 6: Placebo Estimates of the Effect of Failing a July Promotional Cutoff on the Number of Adult Criminal Convictions

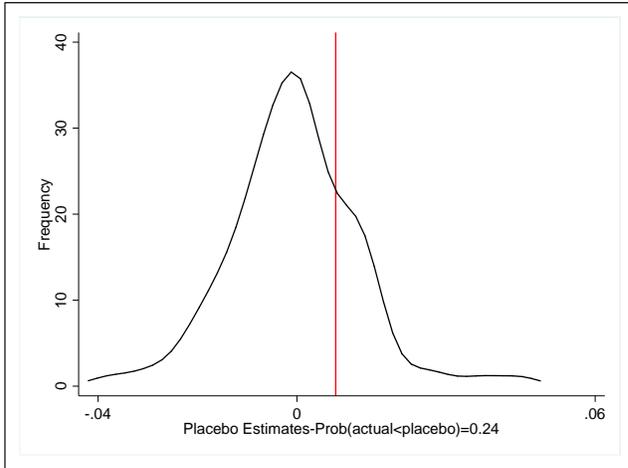
Panel A: Violent Crimes



Panel B: Property Crimes



Panel C: Drug-Related Crimes



Each placebo estimate assigns a false promotion cutoff by incrementally adding one scale point to the subject-specific cutoffs (269 and 296 scale points in ELA and math LEAP tests, respectively) over a range of [-50, 50] scale points. A reduced form equation is then employed to estimate the effect of failing to meet the July cutoff on the number of adult criminal convictions by type of crime. All estimates are obtained from a linear spline using a bandwidth of 30 index points. The vertical lines denote the actual estimates. The fraction of placebo estimates larger than the actual estimate is also reported on the x-axis of each graph.

Table 1: Summary Statistics

	Full Sample		Promoted		Retained	
	Mean (1)	SD (2)	Mean (3)	SD (4)	Mean (5)	SD (6)
Panel A: Student Characteristics						
Black	0.763	0.425	0.661	0.473	0.854	0.352
White	0.218	0.413	0.315	0.464	0.131	0.337
Female	0.520	0.499	0.504	0.500	0.535	0.499
Free/Reduced Lunch	0.454	0.498	0.473	0.499	0.438	0.496
Prior (7th Grade) ITBS Math z-score	-0.857	0.549	-0.727	0.569	-0.975	0.502
Prior (7th Grade) ITBS ELA z-score	-0.685	0.659	-0.586	0.695	-0.774	0.610
8th Grade LEAP Test Outcomes						
Failed Only ELA (July LEAP Exam)	0.039	0.194	0.027	0.162	0.050	0.21
Failed Only Math (July LEAP Exam)	0.452	0.498	0.189	0.391	0.689	0.463
Failed ELA and Math (July LEAP Exam)	0.173	0.379	0.133	0.340	0.209	0.407
Other Outcomes						
Drop Out of School	0.458	0.498	0.372	0.483	0.535	0.499
Committed a Juvenile Crime	0.042	0.199	0.039	0.193	0.044	0.206
Panel B: Adult Characteristics/Outcomes						
Adult Conviction	0.115	0.319	0.110	0.313	0.119	0.324
Adult Crime Type:						
Violent	0.018	0.131	0.016	0.127	0.019	0.136
Property	0.039	0.193	0.038	0.192	0.039	0.195
Drug Related	0.051	0.219	0.048	0.213	0.053	0.225
Other	0.010	0.102	0.011	0.103	0.010	0.100
Age of Adult Crime	20.83	2.81	20.77	2.90	20.88	2.72
Sample Size	22,929		10,868		12,061	

The tabulations reflect our research sample as described in the text, which consists of students enrolled in regular classes in grade 8 between the 1998-1999 and 2000-2001 academic years. The students in the sample took July ELA or math (or both) LEAP exams in eighth grade and did not move out of the state. Prior achievement scores are available beginning with the 1999-2000 academic year and are standardized with respect to the statewide mean and standard deviation by test year, separately for each subject. The full set of sample statistics is available upon request from the authors.

Table 2: Regression Discontinuity Validation Tests

	Dependent Variable:					
	Moved Out of State (Attrition)	Female	Black	White	Free/ Reduced Lunch	7 th Grade Composite Test Score
	(1)	(2)	(3)	(4)	(5)	(6)
Failed July Promotion Cutoff	-0.003 (0.007)	0.011 (0.014)	-0.011 (0.014)	0.016 (0.013)	-0.017 (0.014)	0.006 (0.013)
Sample Size	15,501	14,728	14,728	14,728	14,728	14,378

Authors' estimation of equation (1) as described in the text. The sample is restricted to students who scored within 30 points of one of the cutoffs on the July promotion exams. All specifications control for linear splines in index score as well as test year fixed effects. Column (1) includes students who moved out of state, while the other columns exclude these students. The 7th grade composite score is the average of the standardized test scores in ELA and math from this grade. These scores are available beginning with the 1999-2000 academic year and are standardized against the statewide mean and standard deviation, separately by test year and subject. Standard errors shown in parentheses and are clustered at the school level: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3: First Stage Estimates of the Effect of Failing to Meet a July Promotion Cutoff on Grade Retention

	Dependent Variable: Retained in 8 th Grade		
	(1)	(2)	(3)
	Failed July Promotion Cutoff	0.677*** (0.018)	0.678*** (0.018)
First-Stage F-Statistic	1,352.53	1,357.78	1,274.78
Sample Size	14,728	14,728	14,728
Controls:			
Test Year Fixed Effects	Yes	Yes	Yes
Covariates	No	Yes	Yes
School Fixed Effects	No	No	Yes

Authors' estimation of equation (3) as described in the text. The sample is restricted to students who scored within 30 points of one of the cutoffs on the July promotion exams. All specifications control for linear splines in index score. Covariates include indicators for gender, race, free/reduced lunch and immigrant status. Standard errors are shown in parentheses and are clustered at the school level: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 4: Reduced Form and IV Estimates of Test Failure and Grade Retention on the Likelihood of Adult Crime Conviction (in Percent)

	Reduced Form			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Convicted of Any Crime (x100)						
Failed July Promotion Cutoff	0.687 (0.875)	0.846 (0.864)	0.750 (0.886)			
Grade Retention				1.015 (1.290)	1.248 (1.270)	1.109 (1.283)
Panel B: Convicted of a Violent Crime (x100)						
Failed July Promotion Cutoff	0.668* (0.352)	0.713** (0.351)	0.648** (0.261)			
Grade Retention				0.987* (0.518)	1.052** (0.516)	0.959** (0.379)
Panel C: Convicted of a Property Crime (x100)						
Failed July Promotion Cutoff	-0.072 (0.569)	-0.057 (0.573)	-0.025 (0.587)			
Grade Retention				-0.107 (0.841)	-0.084 (0.845)	-0.037 (0.851)
Panel D: Convicted of a Drug-Related Crime (x100)						
Failed July Promotion Cutoff	0.254 (0.598)	0.346 (0.586)	0.254 (0.598)			
Grade Retention				0.375 (0.883)	0.511 (0.862)	0.376 (0.867)
Sample Size	14,728	14,728	14,728	14,728	14,728	14,728
Controls:						
Test Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	Yes	No	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	Yes

Authors' estimation of equations (1) and (4) as described in the text. The sample is restricted to students who scored within 30 points of one of the cutoffs on the July promotion exams. All specifications control for linear splines in index score. Covariates include indicators for gender, race, free/reduced lunch and immigrant status. Dependent variables take the value of one if the student is convicted as an adult at age 25 or younger. All outcome variables are multiplied by 100 to obtain percent values. Standard errors clustered at the school level are in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 5: Reduced Form and IV Estimates of Test Failure and Grade Retention on the Number of First-time Criminal Convictions by Age 25

	Reduced Form			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Any Crime (Mean=0.136)						
Failed July Promotion Cutoff	0.015 (0.012)	0.017 (0.012)	0.018 (0.012)			
Grade Retention				0.022 (0.018)	0.025 (0.018)	0.027 (0.018)
Panel B: Violent Crimes (Mean=0.020)						
Failed July Promotion Cutoff	0.011** (0.005)	0.012** (0.005)	0.011** (0.005)			
Grade Retention				0.016** (0.007)	0.017** (0.007)	0.017** (0.007)
Panel C: Property Crimes (Mean=0.043)						
Failed July Promotion Cutoff	-0.002 (0.007)	-0.002 (0.007)	-0.001 (0.007)			
Grade Retention				-0.004 (0.010)	-0.003 (0.010)	-0.002 (0.010)
Panel D: Drug Related Crimes (Mean=0.060)						
Failed July Promotion Cutoff	0.007 (0.008)	0.008 (0.008)	0.008 (0.008)			
Grade Retention				0.010 (0.011)	0.012 (0.011)	0.011 (0.011)
Sample Size	14,728	14,728	14,728	14,728	14,728	14,728
Controls:						
Test Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	Yes	No	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	Yes

Authors' estimation of equations (1) and (4) as described in the text. The sample is restricted to students who scored within 30 points of one of the cutoffs on the July promotion exams. All specifications control for linear splines in index score. Covariates include indicators for gender, race, free/reduced lunch and immigrant status. Dependent variables are the number of convictions up to age 25. Standard errors clustered at the school level are in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 6: Mechanism - The Effect of Grade Retention on High School Peer Composition

	Percent of Special Ed Students in High School (1)	Percent of White Students in High School (2)	Average 9 th Grade Math Standardized Test Score (3)
Grade Retention	-0.001 (0.004)	-0.027 (0.017)	-0.062** (0.027)
Sample Size	11,115	11,115	11,115

Authors' estimation of equation (4) as described in the text. The sample is restricted to students who scored within 30 points of one of the cutoffs on the July promotion exams. All specifications control for linear splines in index score as well as test year fixed effects. Covariates include indicators for gender, race, free/reduced lunch and immigrant status. Standard errors clustered at the school level are in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 7: Mechanism - Regression Discontinuity Estimates of Grade Retention on Absenteeism and Disciplinary Incidents

	Missing Outcome Data (1)	Elapsed Time Relative to July Exam		
		Year 1 (2)	Year 2 (3)	Year 3 (4)
Panel A: Missing Data				
Failed July Promotion Cutoff	0.022*** (0.011)			
Panel B: Total Days Absent from School				
Grade Retention		0.762 (0.709)	1.006 (0.923)	1.967*** (0.750)
Mean-Total Days Absent		[13.40]	[17.30]	[17.66]
Panel C: Any Disciplinary Incident (x100)				
Grade Retention		-2.338 (2.408)	0.915 (2.356)	1.124 (2.114)
Mean-Disciplinary Incident-%		[35.97]	[37.13]	[32.67]
Sample Size	15,501	12,963	12,963	12,963

Authors' estimation of equations (1) and (4) as described in the text. The sample is restricted to students who scored within 30 points of one of the cutoffs on the July promotion exams. All specifications control for linear splines in index score as well as test year fixed effects. Missing Outcome Data indicator takes the value of one if total absenteeism/disciplinary involvement information is missing in any of the three years following the July exam. Disciplinary incidents include any actions involving in-school suspension, out-of-school suspension and expulsion. Covariates include indicators for gender, race, free/reduced lunch and immigrant status. The outcome in Panel C is multiplied by 100 to obtain percent values. Standard errors clustered at the school level are in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 8: Mechanism - Regression Discontinuity Estimates of Grade Retention on Dropping Out of School and Juvenile Crime

	HS Attrition (1)	Drop Out of School (2)	Juvenile Conviction (3)	Juvenile Felony (4)
Failed July Promotion Cutoff	0.000 (0.011)			
Grade Retention		7.171*** (2.334)	0.528 (0.811)	0.335 (0.492)
Sample Size	15,501	12,963	14,728	14,728

Authors' estimation of equations (1) and (4) as described in the text. The sample is restricted to students who scored within 30 points of one of the cutoffs on the July promotion exams. All specifications control for linear splines in index score as well as test year fixed effects. Covariates include indicators for gender, race, free/reduced lunch and immigrant status. HS Attrition in Column (1) takes the value of one if student left the public school system before dropout status is determined. Standard errors clustered at the school level are in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 9: The Effect of Grade Retention on Adult Criminal Convictions, by Age

	Age of Adult Crime \geq 19 (1)	Age of Adult Crime \geq 20 (2)
Panel A: Any Violent Crime Conviction (x100)		
Grade Retention	0.835* (0.477)	0.673* (0.402)
Panel B: Any Property Crime Conviction (x100)		
Grade Retention	-0.327 (0.736)	-0.151 (0.684)
Panel C: Any Drug-Related Crime Conviction (x100)		
Grade Retention	0.306 (0.764)	0.269 (0.750)
Sample Size	14,728	14,728

Authors' estimation of equation (4) as described in the text. The sample is restricted to students who scored within 30 points of one of the cutoffs on the July promotion exams. All specifications control for linear splines in index score as well as test year fixed effects. Covariates include indicators for gender, race, free/reduced lunch and immigrant status. The dependent variable in Column (1) takes the value of one if the individual is 19 years old or older at the time of the offense, while in Column (2) the dependent variable is an indicator equal to 1 if the individual is 20 years old or older at the time of the offense. Both outcome variables are multiplied by 100 to obtain percent values. Standard errors clustered at the school level are in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 10: The Effect of Retention on Adult Criminal Convictions Away from the Cutoff

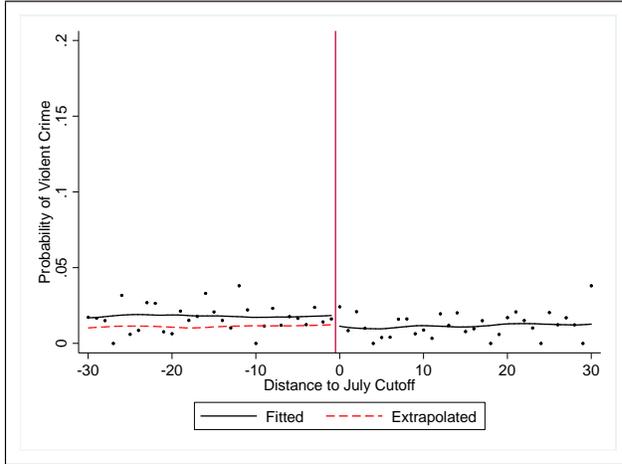
	Treatment Effect on the Treated (Extensive Margin) (1)	Treatment Effect on the Treated (Intensive Margin) (2)	Treatment Effect on the Untreated (Extensive Margin) (3)	Treatment Effect on the Untreated (Intensive Margin) (4)
Panel A: Violent Crime				
Grade Retention	1.055*** (0.263)	0.013*** (0.003)	0.920*** (0.283)	0.010*** (0.003)
Panel B: Property Crime				
Grade Retention	1.423*** (0.448)	0.017*** (0.005)	1.481*** (0.505)	0.017*** (0.006)
Panel C: Drug-Related Crime				
Grade Retention	1.010 (0.700)	0.016* (0.009)	0.904 (0.564)	0.014* (0.007)
Sample Size	14,378	14,378	14,378	14,378

Authors' estimation as described in the text. The sample is restricted to students who scored within 30 points of one of the cutoffs on the July promotion exams. Average treatment effect estimates are obtained using a linear reweighting estimator. Covariates include indicators for gender, race, free/reduced lunch, immigrant status, composite test scores and school and test year fixed effects. Outcome variables in Columns (1) and (3) are multiplied by 100 to obtain percent values, while total number of adult convictions up to age 25 are used in Columns (2) and (4). Standard errors computed using a nonparametric block bootstrap at the school level with 500 replications are in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

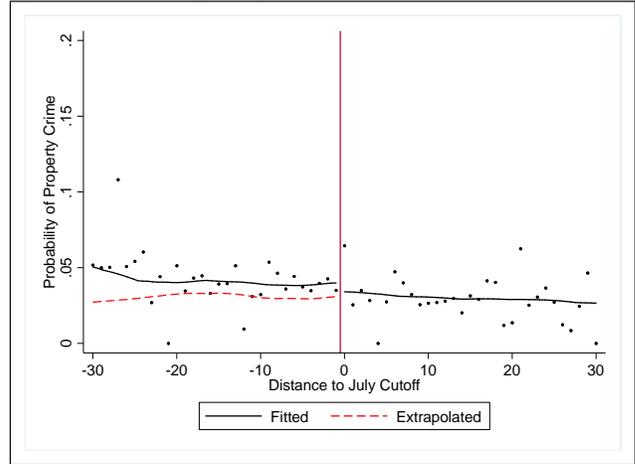
Online Appendix: Not for Publication

Figure A-1: Linear Reweighting Estimates of the Counterfactual-Treatment Effect on the Treated (Extensive Margin)

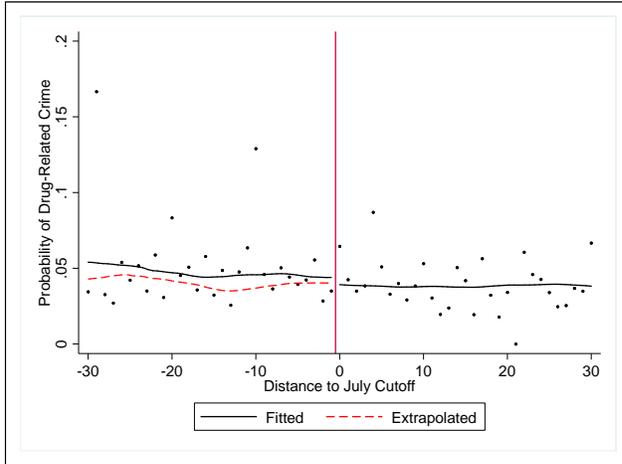
Panel A: Violent Crimes



Panel B: Property Crimes



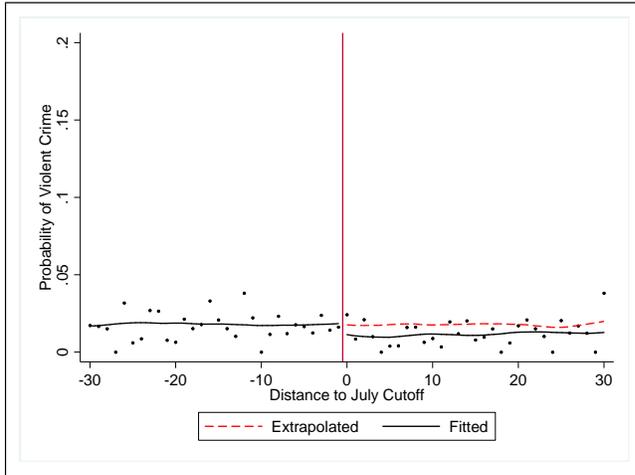
Panel C: Drug-Related Crimes



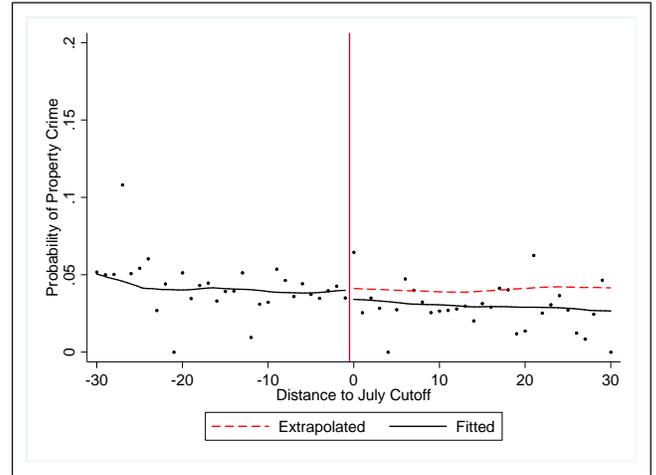
The Fitted line is the local linear estimate based on the actual data points shown. The Extrapolated line is the estimated counterfactual using the Angrist and Rokkanen (2015) method described in the text. The difference between the Fitted and the Extrapolated is the estimated effect away from the cutoff.

Figure A-2: Linear Reweighting Estimates of the Counterfactual-Treatment Effect on the Untreated (Extensive Margin)

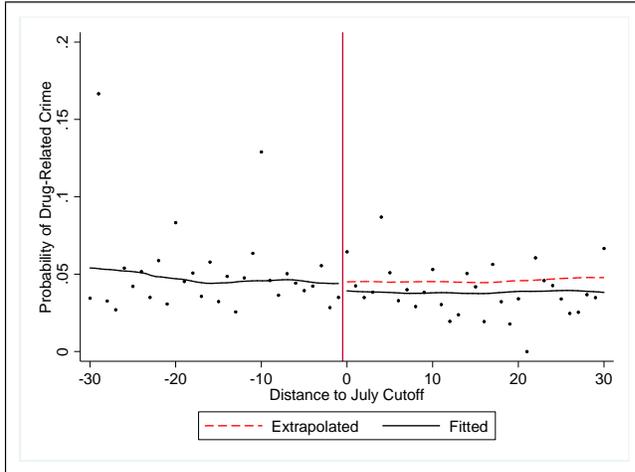
Panel A: Violent Crimes



Panel B: Property Crimes



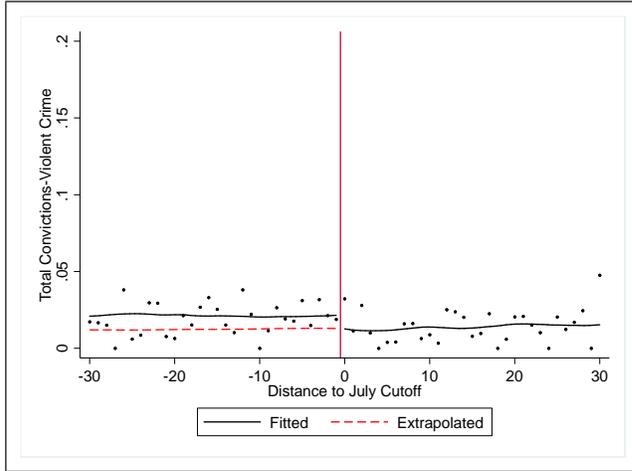
Panel C: Drug-Related Crimes



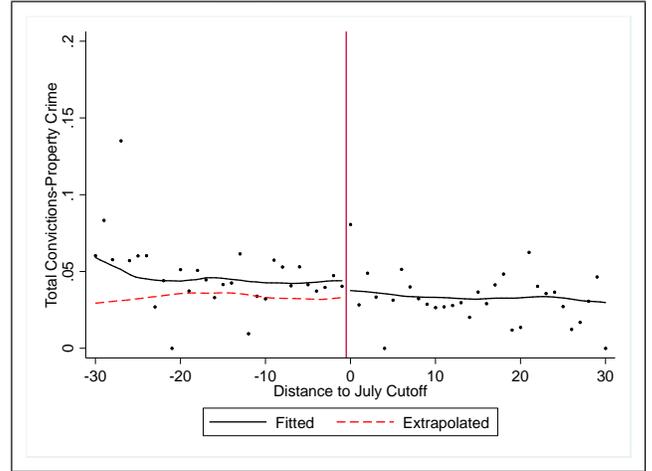
The Fitted line is the local linear estimate based on the actual data points shown. The Extrapolated line is the estimated counterfactual using the Angrist and Rokkanen (2015) method described in the text. The difference between the Fitted and the Extrapolated is the estimated effect away from the cutoff.

Figure A-3: Linear Reweighting Estimates of the Counterfactual-Treatment Effect on the Treated (Intensive Margin)

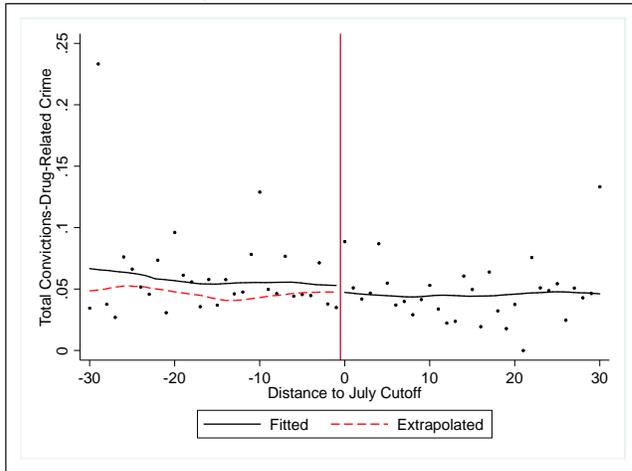
Panel A: Violent Crimes



Panel B: Property Crimes



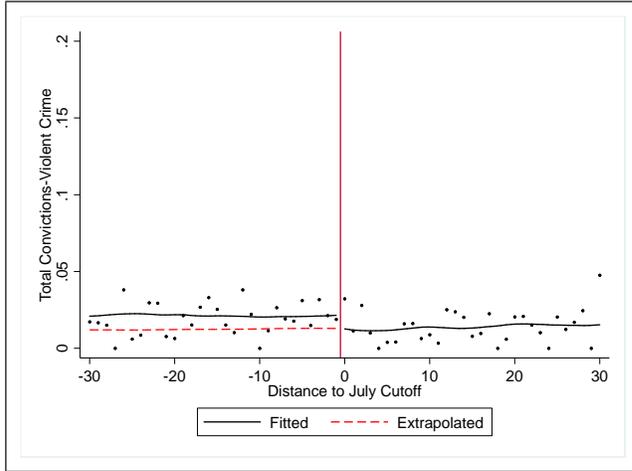
Panel C: Drug-Related Crimes



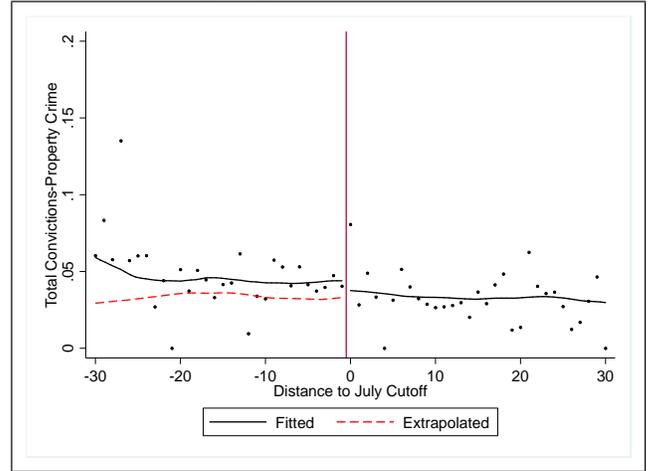
The Fitted line is the local linear estimate based on the actual data points shown. The Extrapolated line is the estimated counterfactual using the Angrist and Rokkanen (2015) method described in the text. The difference between the Fitted and the Extrapolated is the estimated effect away from the cutoff.

Figure A-4: Linear Reweighting Estimates of the Counterfactual-Treatment Effect on the Treated (Intensive Margin)

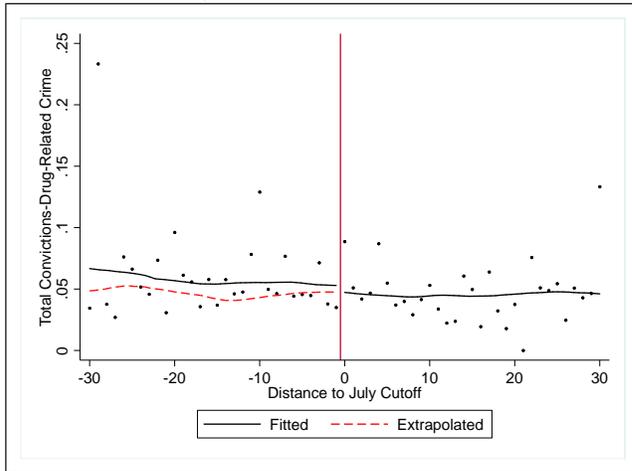
Panel A: Violent Crimes



Panel B: Property Crimes



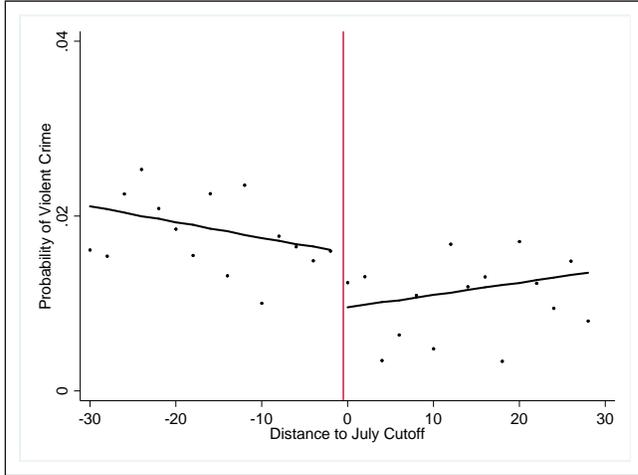
Panel C: Drug-Related Crimes



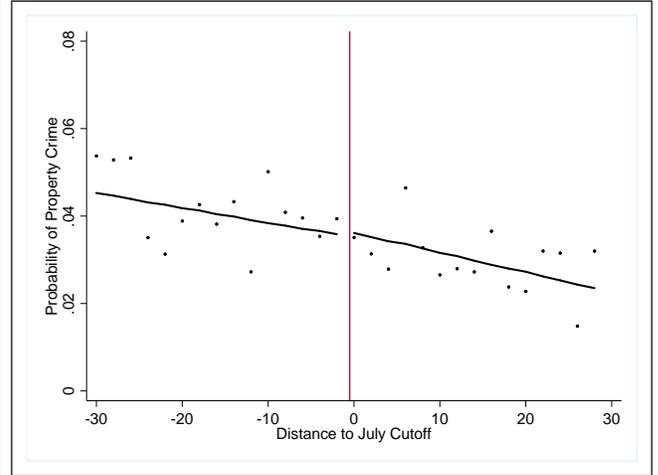
The Fitted line is the local linear estimate based on the actual data points shown. The Extrapolated line is the estimated counterfactual using the Angrist and Rokkanen (2015) method described in the text. The difference between the Fitted and the Extrapolated is the estimated effect away from the cutoff.

Figure A-5: Any Criminal Convictions by Distance to the July Promotional Cutoff - Different Scales

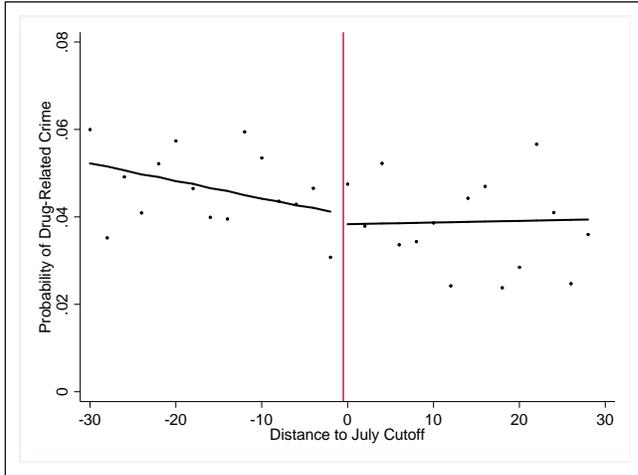
Panel A: Violent Crimes



Panel B: Property Crimes

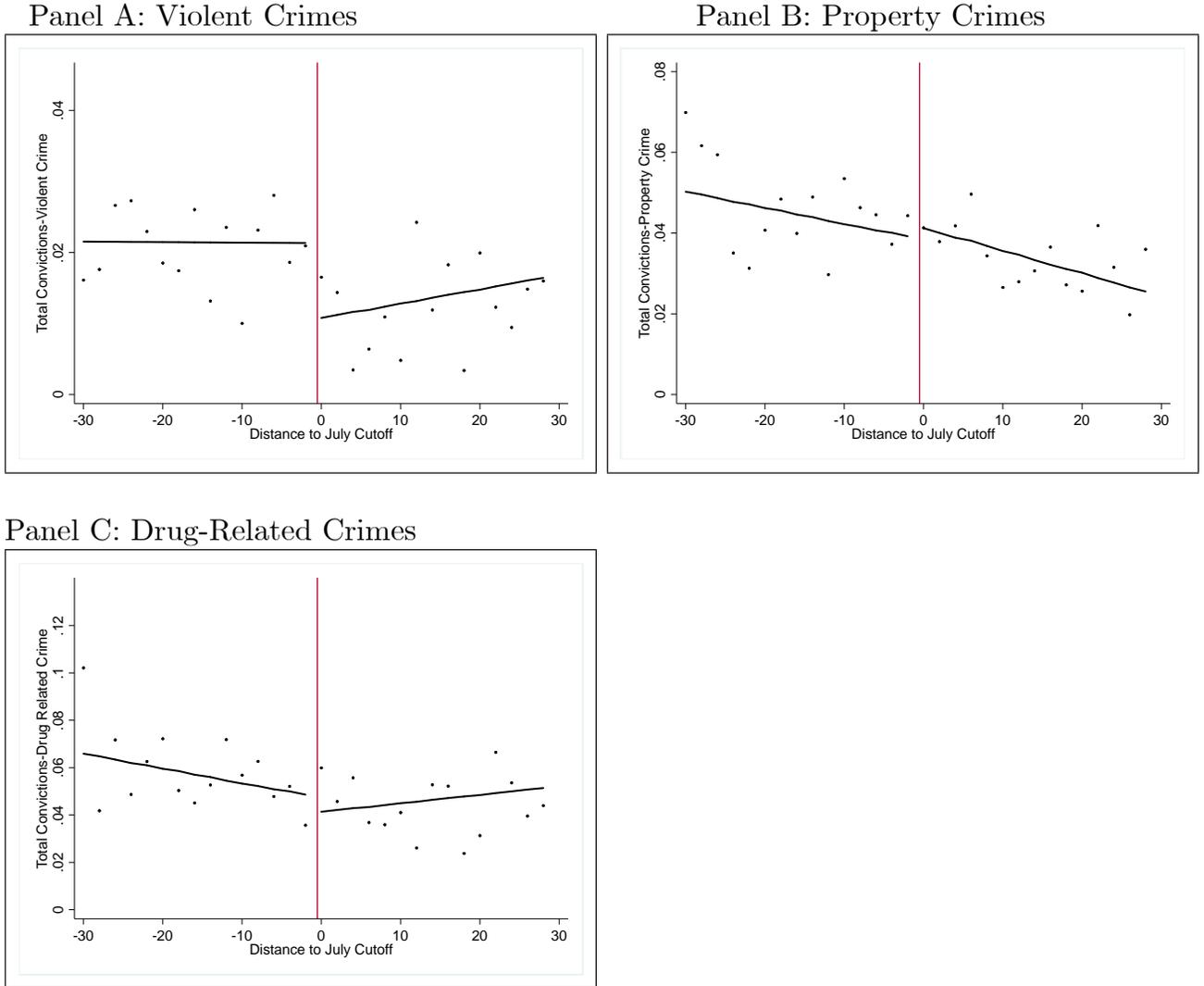


Panel C: Drug-Related Crimes



The vertical line denotes -0.5 points to the left of the July promotional cutoff (centered at 0). Each circle represents the unconditional mean of adult crime conviction and types in two index point bins, based on the distance to July cutoff. The solid lines are fitted values of probability of criminal conviction from a linear spline over an index bandwidth of 30 points.

Figure A-6: The Number of Adult Crime Convictions by Distance to the July Test Cutoff - Different Scales



The vertical line denotes -0.5 points to the left of the July promotional cutoff (centered at 0). Each circle represents the unconditional mean of the number of adult crime convictions, overall and by type, in two index point bins based on the distance to July cutoff. The solid lines are fitted values of the number of criminal convictions from a linear spline over an index bandwidth of 30 points.

Table A-1: Robustness Checks of Bandwidth and Running Variable Controls - Regression Discontinuity Estimates of Grade Retention on the Likelihood of Being Convicted of a Crime

	Local Linear (Index=IK) (1)	Linear Spline Index=[-15,15] (2)	Linear Spline Index=[-45,45] (3)	Quadratic Spline Index=[-60,60] (4)	Quadratic Spline (Global) (5)
Panel A: Convicted of a Violent Crime (x100)					
Grade Retention	1.033* (0.582) [11,460]	0.807* (0.430) [9,118]	1.047** (0.438) [18,151]	1.033* (0.607) [19,967]	1.150*** (0.419) [22,929]
Panel B: Convicted of a Property Crime (x100)					
Grade Retention	0.400 (1.025) [10,815]	0.276 (1.105) [9,118]	0.430 (0.690) [18,151]	-0.151 (0.987) [19,967]	0.330 (0.695) [22,929]
Panel C: Convicted of a Drug-Related Crime (x100)					
Grade Retention	-0.194 (1.085) [9,741]	-0.583 (1.125) [9,118]	0.831 (0.791) [18,151]	0.262 (1.023) [19,967]	0.480 (0.795) [22,929]

Authors' estimation as described in the text. All specifications in columns (2)-(5) control for linear or quadratic splines in index score using the listed bandwidth. Optimal bandwidths for local linear regression in the first column are obtained using the procedures in Imbens and Kalyanaraman (2012). All outcome variables are multiplied by 100 to obtain percent values. Covariates include indicators for gender, race, free/reduced lunch, immigration status, and test year fixed effects. Sample sizes underlying each specification are shown in brackets. Standard errors in parentheses are clustered at the school level: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A-2: Robustness Checks of Bandwidth and Running Variable Controls - Regression Discontinuity Estimates of Grade Retention on the Number of Adult Crime Convictions

	Local Linear (Index=IK) (1)	Linear Spline Index=[-15,15] (2)	Linear Spline Index=[-45,45] (3)	Quadratic Spline Index=[-60,60] (4)	Quadratic Spline (Global) (5)
Panel A: Violent Crime					
Grade Retention	0.015** (0.007) [14,493]	0.017* (0.009) [9,118]	0.013** (0.006) [18,151]	0.017** (0.008) [19,967]	0.015*** (0.006) [22,929]
Panel B: Property Crime					
Grade Retention	0.001 (0.013) [10,263]	-0.002 (0.014) [9,118]	0.004 (0.008) [18,151]	-0.008 (0.012) [19,967]	0.000 (0.009) [22,929]
Panel C: Drug-Related Crime					
Grade Retention	-0.009 (0.014) [9,345]	-0.009 (0.014) [9,118]	0.018* (0.010) [18,151]	0.007 (0.013) [19,967]	0.013 (0.010) [22,929]

Authors' estimation as described in the text. All specifications in columns (2)-(5) control for linear or quadratic splines in index score using the listed bandwidth. Optimal bandwidths for local linear regression in the first column are obtained using the procedures in Imbens and Kalyanaraman (2012). Covariates include indicators for gender, race, free/reduced lunch, immigration status, and test year fixed effects. Sample sizes underlying each specification are shown in brackets. Standard errors in parentheses are clustered at the school level: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A-3: Robustness Checks of Sample Composition and Clustering - Regression Discontinuity Estimates of Grade Retention on the Likelihood of Being Convicted of a Crime

	Conviction at Any Age Up to 28 (1)	Drop Hurricane Katrina/Rita Regions (2)	Add 7 th Grade Test Scores (3)	Add 8 th Grade March Test Scores (4)	Standard Errors Clustered at the Index Level (5)
Panel A: Convicted of a Violent Crime (x100)					
Grade Retention	1.178** (0.528)	1.397** (0.612)	1.080** (0.526)	1.038** (0.515)	1.052*** (0.404)
Panel B: Convicted of a Property Crime (x100)					
Grade Retention	-0.067 (0.876)	-0.499 (0.973)	0.058 (0.850)	-0.144 (0.845)	-0.084 (0.658)
Panel C: Convicted of a Drug-Related Crime (x100)					
Grade Retention	0.374 (0.937)	0.730 (0.947)	0.510 (0.860)	0.483 (0.861)	0.511 (0.783)
Sample Size	14,728	10,199	14,378	14,728	14,728

Authors' estimation as described in the text. All specifications control for linear splines in index score with a bandwidth of 30. The dependent variable takes the value of one if a student is convicted as an adult by age 25 in columns (2)-(4). In column (1), the dependent variable takes the value of one if a student is convicted of a crime up to age 28. The analysis sample in column (2) excludes parishes that are known to be most affected from Hurricanes Katrina and Rita. Column (3) adds 7th grade composite scores into the specifications and column (4) adds subject-specific 8th grade test scores from the March exam. Covariates include indicators for gender, race, free/reduced lunch, immigration status, and test year fixed effects. All outcome variables are multiplied by 100 to obtain percent values. Standard errors in parentheses are clustered at the school level in columns (1)-(3) and at the index (running variable) level in column (4): * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A-4: Robustness Checks of Sample Composition and Clustering - Regression Discontinuity Estimates of Grade Retention on the Number of Adult Crime Convictions

	Conviction at Any Age Up to 28 (1)	Drop Hurricane Katrina/Rita Regions (2)	Add 7 th Grade Test Scores (3)	Add 8 th Grade March Test Scores (4)	Standard Errors Clustered at the Index Level (5)
Panel A: Number of Violent Crime Convictions					
Grade Retention	0.018** (0.007)	0.020** (0.008)	0.018** (0.007)	0.017** (0.007)	0.017*** (0.005)
Panel B: Number of Property Crime Convictions					
Grade Retention	-0.003 (0.010)	-0.009 (0.011)	-0.000 (0.010)	-0.004 (0.010)	-0.003 (0.008)
Panel C: Number of Drug-Related Crime Convictions					
Grade Retention	0.010 (0.011)	0.013 (0.011)	0.012 (0.011)	0.011 (0.011)	0.012 (0.012)
Sample Size	14,728	10,199	14,378	14,728	14,728

Authors' estimation as described in the text. All specifications control for linear splines in index score with a bandwidth of 30. The dependent variable is the number of adult crime convictions by age 25 in columns (2)-(4). In column (1), the dependent variable is the number of adult convictions up to age 28. The analysis sample in column (2) excludes parishes that are known to be most affected from Hurricanes Katrina and Rita. Column (3) adds 7th grade composite scores into the specifications and column (4) adds subject-specific 8th grade test scores from the March exam. Covariates include indicators for gender, race, free/reduced lunch, immigration status, and test year fixed effects. Standard errors in parentheses are clustered at the school level in columns (1)-(3) and at the index (running variable) level in column (4): * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A-5: Regression Discontinuity Estimates of Grade Retention on Adult Crime Convictions Among Men

	Extensive Margin			Intensive Margin		
	ATT	ATUT		ATT	ATUT	
	(away from	(away from		(away from	(away from	
	cutoff)	cutoff)		cutoff)	cutoff)	
	(2)	(3)		(5)	(6)	
	R		R			
	D		D			
	(1)		(4)			
Panel A: Violent Crime						
Grade Retention	2.072*	2.486***	1.654**	0.035**	0.032***	0.019**
	(1.106)	(0.669)	(0.723)	(0.015)	(0.008)	(0.008)
Dep. Var. Mean	0.030			0.036		
Panel B: Property Crime						
Grade Retention	-1.121	2.476***	1.609	-0.018	0.032***	0.023*
	(1.167)	(0.916)	(1.138)	(0.021)	(0.010)	(0.013)
Dep. Var. Mean	0.060			0.067		
Panel C: Drug-Related Crime						
Grade Retention	0.046	1.617	2.014	0.014	0.031	0.036**
	(1.815)	(2.019)	(1.249)	(0.023)	(0.024)	(0.017)
Dep. Var. Mean	0.090			0.108		

Authors' estimation as described in the text. The sample is restricted to male students who scored within 30 points of one of the cutoffs on the July promotion exams; N=6,699. Results in columns (1) and (4) are IV estimates from equation (4) that use the July promotion cutoff as an instrument for retention. These specifications control for linear splines in index score with a bandwidth of 30. Results in other columns are extrapolated treatment effects away from the cutoff using the Angrist and Rokannen (2015) method. The ATT (average treatment effect on the treated) estimates show effects for those below the cutoff and the ATUT (average treatment effect among the untreated) estimates show effects for those above the cutoff. The dependent variable in columns (1)-(3) takes the value of one if a student is convicted as an adult by age 25. These estimates are multiplied by 100 to obtain percent values. The dependent variable in columns (4)-(6) is the number of adult crime convictions by age 25. Covariates include indicators for gender, race, free/reduced lunch, immigration status, and test year fixed effects. The estimates in columns (2), (3), (5), and (6) also include controls for composite test scores and school fixed effects. Standard errors are shown in parentheses. In columns (1) and (4) standard errors are clustered at the school level and in other columns they are computed using a nonparametric block bootstrap at the school level with 500 replications: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A-6: Regression Discontinuity Estimates of Grade Retention on Adult Crime Convictions Among Women

	Extensive Margin			Intensive Margin		
	RD	ATT (away from cutoff)	ATUT (away from cutoff)	RD	ATT (away from cutoff)	ATUT (away from cutoff)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Violent Crime						
Grade Retention	0.271 (0.421)	0.029 (0.269)	0.241 (0.233)	0.003 (0.005)	-0.000 (0.003)	0.0001 (0.003)
Dep. Var. Mean	0.005			0.006		
Panel B: Property Crime						
Grade Retention	0.727 (0.789)	0.850 (0.593)	1.282** (0.538)	0.008 (0.008)	0.008 (0.006)	0.012** (0.006)
Dep. Var. Mean	0.019			0.020		
Panel C: Drug-Related Crime						
Grade Retention	0.801 (0.725)	0.520 (0.396)	-0.002 (0.401)	0.009 (0.010)	0.002 (0.006)	-0.003 (0.005)
Dep. Var. Mean	0.013			0.016		

Authors' estimation as described in the text. The sample is restricted to female students who scored within 30 points of one of the cutoffs on the July promotion exams; N=8,029. Results in columns (1) and (4) are IV estimates from equation (4) that use the July promotion cutoff as an instrument for retention. These specifications control for linear splines in index score with a bandwidth of 30. Results in other columns are extrapolated treatment effects away from the cutoff using the Angrist and Rokannen (2015) method. The ATT (average treatment effect on the treated) estimates show effects for those below the cutoff and the ATUT (average treatment effect among the untreated) estimates show effects for those above the cutoff. The dependent variable in columns (1)-(3) takes the value of one if a student is convicted as an adult by age 25. These estimates are multiplied by 100 to obtain percent values. The dependent variable in columns (4)-(6) is the number of adult crime convictions by age 25. Covariates include indicators for gender, race, free/reduced lunch, immigration status, and test year fixed effects. The estimates in columns (2), (3), (5), and (6) also include controls for composite test scores and school fixed effects. Standard errors are shown in parentheses. In columns (1) and (4) standard errors are clustered at the school level and in other columns they are computed using a nonparametric block bootstrap at the school level with 500 replications: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A-7: Regression Discontinuity Estimates of The Effect of Failing the March Exam on Adult Criminal Convictions

	Dependent Variable:	
	I(Convicted of a Crime)x100 (1)	Number of Crimes Convicted (2)
Panel A: Violent Crime		
Failed March Promotion Cutoff	0.018 (0.174)	0.000 (0.002)
Dep. Var. Mean	0.011	0.012
Panel B: Property Crime		
Failed March Promotion Cutoff	-0.041 (0.264)	0.002 (0.003)
Dep. Var. Mean	0.026	0.029
Panel C: Drug-Related Crime (x100)		
Failed March Promotion Cutoff	-0.036 (0.317)	-0.000 (0.004)
Dep. Var. Mean	0.036	0.043
Sample Size	62,405	62,405

Authors' estimation of the effect of failing the March promotion cutoff as described in the text. All specifications control for linear splines in index score with a bandwidth of 30. The dependent variable takes the value of one if a student is convicted as an adult by age 25 in column 1 (1) and is the number of crimes of which the individual is convicted at first conviction by age 25 in column (2). The estimates in column (1) are multiplied by 100 to obtain percent values. Covariates include indicators for gender, race, free/reduced lunch, immigration status, and test year fixed effects. Standard errors in parentheses are clustered at the school level in columns (1)-(3) and at the index (running variable) level in column (4): * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A-8: The Effect of Grade Retention on Adult Criminal Convictions Using only the Math Score as the Running Variable

	Extensive	
	Margin (x100) (1)	Intensive Margin (2)
Panel A: Violent Crime		
Grade Retention	1.121** (0.426)	0.018** (0.006)
Panel B: Property Crime		
Grade Retention	-0.397 (0.888)	-0.007 (0.101)
Panel C: Drug-Related Crime		
Grade Retention	0.063 (0.879)	0.004 (0.013)
Sample Size	14,317	14,317

Authors' estimation as described in the text. All specifications control for linear splines in math score with a bandwidth of 30. The dependent variable takes the value of one if a student is convicted as an adult by age 25 in column (1) and is the number of crimes at first conviction of which one is convicted by age 25 in column (2). Covariates include indicators for gender, race, free/reduced lunch, immigration status, and test year fixed effects. Outcome variables are multiplied by 100 in column (1) to obtain percent values. Standard errors in parentheses are clustered at the school level: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A-9: Conditional Independence Tests

	Failed July Likelihood of Being Convicted (1)	Cutoff=1 Number of Crime Convictions (2)	Failed July Likelihood of Being Convicted (3)	Cutoff=0 Number of Crime Convictions (4)
Panel A: Violent Crimes				
Running Variable (Index Score)	-0.016 (0.020)	0.0001 (0.0003)	0.010 (0.018)	0.0001 (0.0002)
Panel B: Property Crimes				
Running Variable (Index Score)	-0.025 (0.028)	-0.0003 (0.0003)	-0.027 (0.027)	-0.0004 (0.0003)
Panel C: Drug Related Crimes				
Running Variable (Index Score)	-0.021 (0.028)	-0.0004 (0.0004)	-0.004 (0.035)	0.0003 (0.0005)
Sample Size	7,873	7,873	6,505	6,505

Authors' estimation as described in the text. All specifications control linearly for index score with a bandwidth of 30. Estimates use only observations to the left or right of the cutoff as indicated in column headings. Covariates include indicators for gender, race, free/reduced lunch, immigrant status, composite baseline test scores and school and test year fixed effects. Standard errors in parentheses are clustered at the school level: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A-10: Potential Channels - Regression Discontinuity Estimates of Grade Retention on Education and Crime Outcomes

	Classified as Special Ed in Any Year (Post-July Exam) (1)	Age at Conviction (2)	Sentence Length Imposed (# of Months) (3)	Highest Grade Attained (Post-July Exam) (4)
Grade Retention	0.233 (1.651)	0.262 (0.267)	3.042 (8.492)	-0.264*** (0.055)
Mean of Outcome	20.3%	20.83	79.09	3.02
Sample Size	12,745	10,199	14,728	14,728

Authors' estimation of equation (4) described in the text. The sample is restricted to students who scored within 30 points of one of the cutoffs on the July promotion exams. All specifications control for linear splines in index score as well as test year fixed effects. Covariates include indicators for gender, race, free/reduced lunch, immigration status, and test year fixed effects. Standard errors clustered at the school level are in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.