1-22-2014

The Effects of Doubling Instruction Efforts on Middle School Students' Achievement: Evidence from a Multiyear Regression-Discontinuity Design

Timothy J. Bartik
W.E. Upjohn Institute for Employment Research, bartik@upjohn.org

Marta Lachowska
W.E. Upjohn Institute for Employment Research and Stockholm University, marta@upjohn.org

Upjohn Institute working paper ; 14-205

Citation

Bartik, Timothy J. and Marta Lachowska. 2014. "The Effects of Doubling Instruction Efforts on Middle School Students' Achievement: Evidence from a Multiyear Regression-Discontinuity Design." Upjohn Institute Working Paper 14-205. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
https://doi.org/10.17848/wp14-205

The Effects of Doubling Instruction Efforts on Middle School Students' Achievement: Evidence from a Multiyear Regression-Discontinuity Design

Authors
Timothy J. Bartik, *W.E. Upjohn Institute for Employment Research*
Marta Lachowska, *W.E. Upjohn Institute for Employment Research and Stockholm University*

Upjohn Author(s) ORCID Identifier
https://orcid.org/0000-0002-6238-8181

https://orcid.org/0000-0001-9158-2592

W.E. UPJOHN INSTITUTE
FOR EMPLOYMENT RESEARCH
Researching the causes and consequences of unemployment

The Effects of Doubling Instruction Efforts on Middle School Students' Achievement: Evidence from a Multiyear Regression-Discontinuity Design

Upjohn Institute Working Paper No. 14-205

Timothy J. Bartik
W.E. Upjohn Institute for Employment Research
email: Bartik@upjohn.org

Marta Lachowska
W.E. Upjohn Institute for Employment Research and *Stockholm University*
email: Marta@upjohn.org

January 22, 2014

ABSTRACT

We use a regression-discontinuity design to study the effects of double blocking sixth-grade students in reading and mathematics on their achievement across three years of middle school. To identify the effect of the intervention, we use sharp cutoffs in the test scores used to assign students to double blocking. We find large, positive, and persistent effects of double blocking in reading, but, unlike previous research, we find no statistically significant effects of double blocking in mathematics either in the short run or medium run.

JEL Codes: I21, C21

Keywords: Regression discontinuity, Double blocking, Middle school

Acknowledgments

We thank Wei-Jang Huang for excellent research assistance. All errors are our own.

1 Introduction

To boost student achievement, many American school districts have recently adopted so-called "double-dose programs." Such programs reallocate student instructional time toward a core subject area in which the student fails to meet standards (Cavanagh 2006). Over the past ten years, double dosing has become one of the most commonly used remedial policies in high schools (Durwood, Krone, and Mazzeo 2010).[1]

Double-dose programs are popular, but do they work? Existing research has mixed results. Some research finds benefits for some targeted students (Nomi and Allensworth 2009; Dougherty 2012). One study finds benefits persisting for years after the intervention (Cortes, Goodman, and Nomi 2013a and 2013b). In contrast, Fryer (2012) finds little systematic evidence of achievement gains.[2]

These different evaluation results may be in part due to different program designs. Double dose's effects may vary by subject, for example, whether the subject double dosed is reading (Dougherty 2012, Fryer 2012) or mathematics (Nomi and Allensworth 2009; Fryer 2012; Cortes, Goodman, and Nomi 2013b). Effects may depend on whether the intervention comprises only additional subject instructional time or a different curriculum or both. The grade level of double dosing may also make a difference. Evidence from other educational interventions suggests that one cannot typically assume that the effect of a similar program is the same across all grade levels,

[1] Durwood, Krone, and Mazzeo (2010) cite a 2009 survey, which suggests that about half of large urban school districts in the U.S. implement some form of double-dose programs.

[2] Research has also found that the change in the classroom composition due to double dosing had beneficial consequences for high-ability students, not directly targeted by the remedial policy (Nomi and Allensworth 2009). Such spillover effects are a separate issue from the direct effects on targeted students that are considered in our paper.

schools, or communities; see, for example, Fryer (2011) on the heterogeneous effects of randomized experiments in various settings.

A key challenge to evaluating educational interventions is unobserved student characteristics, which may prevent us from attributing any learning gains to the educational remedy. To address this challenge, the research on double dosing has commonly used regression-discontinuity designs (Nomi and Allensworth 2009; Dougherty 2012; Cortes, Goodman, and Nomi 2013b) or randomized control trials (Fryer 2012).

In this paper, we study a double-dose program implemented in the middle schools of an unnamed mid-sized Midwest School District (MSD). Starting in the school year 2010–2011, the MSD started increasing instructional time for middle-school students who did not meet required standards in a subject. This intervention is called "double blocking," because the student is assigned to a second block (a second period) in that subject. Students assigned to a second block are provided with a combination of direct instruction, independent practice time, and computerized exercises and quizzes.

Our paper estimates the effects of double blocking sixth-grade students on their academic performance in middle school using a regression-discontinuity design. The double-blocking policy assigns a double block of reading or mathematics to sixth-grade students who scored below a cutoff point on a fifth-grade test. The fifth-grade test for making double-blocking assignments is used by the state for school accountability purposes. Because the rule that assigns students into the double-blocking program is known and is based on a metric that is difficult to precisely manipulate by either the test takers or teachers, we can readily model the effects of the program using a regression-discontinuity design. We study the effects for the first cohort of double-blocked sixth graders on their test scores in the three years following the initial assignment to double blocking.

The double-blocking program studied in this paper differs from the double-dosing programs studied by previous research. First, double blocking consists of assigning students to a double period in the subject in which they are struggling, hence the structure of the curriculum might be different from a double dosing program. Second, the double-blocking program in the MSD offers additional instruction in reading as well as mathematics. Apart from Dougherty (2012) and Fryer (2012), most of the studies have focused on the benefits of double dosing in mathematics and have generally found learning gains in math. Less is known about the effects of double dosing on reading gains. Third, in the MSD, the double-blocking assignment rule in mathematics was conditional on passing the reading part of the test, whereas the assignment in the Chicago Public Schools studied by Nomi and Allensworth (2009) and Cortes, Goodman, and Nomi (2013b) was determined by scoring below the national median student score on the Iowa Tests of Basic Skills. Hence, the assignment rule in the MSD could target a somewhat different population. Fourth, as in Dougherty (2012), the MSD double blocked sixth graders, whereas the Nomi and Allensworth (2009) and Cortes, Goodman, and Nomi (2012b) studies of double dosing in Chicago Public Schools targeted 9^{th} graders. Fifth, the MSD is moderately sized with relatively small middle schools, and educational interventions may work differently in a smaller learning community than in a large school district such as the Chicago Public Schools (see Abdulkadiroğlu, Hu, and Pathak 2013).

Our paper is also relevant to the broader empirical literature that evaluates many educational policies using regression-discontinuity designs. This regression-discontinuity educational literature includes, in addition to the work on double dosing mentioned above, Angrist and Lavy (1999), van der Klaauw (2002), Jacob and Lefgren (2004), Goodman (2008), and

Matsudaira (2008). Our work helps provide another example of the potential of the regression-discontinuity approach in improving educational evaluations.

Our initial results suggest that assignment to and participation in double blocking in reading resulted in statistically positive and meaningful achievement gains that persisted beyond the first year of the intervention. In the first and second year following assignment to double blocking in reading, the state test scores of students assigned to double blocking increased by about 0.20 standard deviations (from now on, σ), which is a large gain. In the third year, the test scores increased by about 0.40σ. Not all students assigned to double-blocking actually participated, so the implied effect of double-blocking on actual participants is higher: an increase in about 0.30σ on the reading standardized test scores in the first and second year following the invention and 0.65σ in the third year following the intervention.

For students assigned to and participating in double blocking in math, the estimated coefficients are small and imprecise. Why was math double-blocking less successful? We speculate that the implementation of the mathematics intervention was not as systematic as the reading intervention.

The rest of the paper is organized as follows: Section 2 provides details of the double-blocking program and the characteristics of the school district. Section 3 describes the data and our empirical strategy. Section 4 discusses the results. The final section concludes.

2 The Midwest School District and Double Blocking

The Midwest School District (MSD) system is a mid-sized (5,000 to 20,000 students), mostly urban school district with a racially and economically diverse student body. There are multiple middle schools and multiple high schools in the MSD. Middle school in the MSD includes grades six through eight.

In the effort to raise student achievement levels, the MSD has implemented several educational interventions. As part of this effort, starting in the 2010–2011 school year, the MSD is "double blocking" students in middle school who on the 2009 state reading or math test scored below a cutoff. Currently about 40 percent of students in the Midwest state where the MDS is located score below this cutoff. "Double blocking" means that for the first semester of the 2010–2011 school year, students are assigned to a second block in either reading or math.

Students double blocked in reading follow a separate curriculum using Read 180 materials from the Scholastic Corporation. The classes using Read 180 follow a specific protocol; students are provided with a combination of direct instruction, computerized exercises and quizzes, and independent practice time. Hence, the "treatment" for students double blocked in reading has several dimensions and consists of additional instruction time and a curriculum different from that of students who were not double blocked in reading. Students double blocked in mathematics spent additional time in instruction but did not follow a separate curriculum from the students not double blocked in math.

Students may be removed from the double block at the end of trimesters if they score above a cutoff on a progress-tracking test or they get a grade of C or higher in the subject that they are double blocked in. Ex-ante it was estimated that over 90 percent of those double blocked will stay in the double-block program for at least two trimesters. The forecast turned out to be correct; ex-post some 98 percent of students assigned to a double block in reading had not exited the program by the end of the second trimester.

The protocol for assigning students into double blocking was based on performance on the state standardized test, which is taken by all students in the state from grade three to grade nine. The state standardized test yields a scaled score for each student. These scaled scores are designed to be comparable across the different tests given in various years, although not across grades. The scores are also designed to have a particular standard deviation, which we use to express the results in "standard-deviation" units. A cutoff is used to decide whether a student is proficient in the subject. The percentage of proficient students is used by the state for school accountability purposes, for example, in deciding whether a school may be forced under provisions of No Child Left Behind to undertake any of various reform measures.

2.1 Double Blocking in the MSD

The assignment to the MSD double-blocking program has two tiers: reading and mathematics. For middle-school students in the fall of 2010, the assignment rule for double blocking was the following:

- A student is double blocked in reading if they were not proficient in reading as judged by the fall 2009 reading state standardized test.
- A student is double blocked in math if they were proficient on the fall 2009 state reading test but not proficient on the fall 2009 state math test.

Because we know exactly how students were selected for the program, we can model the effects of the program.

If the program is effective, we expect to see an abrupt increase in postprogram outcomes for those who barely scored poorly enough on the fall 2009 tests to be double blocked, versus those who barely scored well enough on the fall 2009 tests not to be double blocked. This abrupt increase in test scores would cause a statistically noticeable "jump" in postprogram test scores relative to what would have been predicted from a smooth effect of the fall 2009 tests on postprogram scores. This estimation approach is referred to a regression-discontinuity (RD) design, which relies on a regression-prediction equation to reveal a smooth relationship between the variable determining selection into the program and the dependent variable, as well as on the discontinuity in program participation caused by rules for selection into the program.

RD designs rely on the “thought experiment” that in the absence of the program, student performance postprogram would be a smooth continuation of the fall 2009 test scores. The research literature suggests that regression-discontinuity evaluations obtain estimates similar to what would have been obtained by random assignment experimentation, which is widely accepted as the “gold standard” in empirical work; see Lee and Lemieux (2010).

3 DATA AND METHODS

3.1 Data

Our data come from the MSD administrative records. In our analysis, we focus on students who were rising sixth graders (that is, current fifth graders) at the time of the fall 2009 state test. We chose this focus, as it allows us to follow students for all of the three remaining years of middle school after they were tested in the fall of 2009.

Our regression sample consists of these students across the school years 2010–2011, 2011–2012, and 2012–2013. For these years we have records on student characteristics (gender, race, ethnicity, and free or reduced lunch status), which school they attend, an indicator for double-block eligibility and participation, and the state reading and scaled math scores. Our sample is an unbalanced panel—some students leave the school district and we do not observe them in our data set after they leave.

We use the 2009 state reading scaled scores to determine eligibility for double blocking in reading and math, based on the state test cutoff points. The assignment rule required that a rising sixth grader who scored less than a cutoff on the reading test be assigned to a double block of reading. If a rising sixth-grade student scored greater than that scaled score cutoff in reading but less than the cutoff on the math test, he or she was double blocked in math. As occurs with any intended policy rule, in some cases this clear-cut assignment was sidestepped. In our data, we have information on which students ended up not following the assignment protocol. We use this information to construct a participation indicator for double blocking.

For outcome variables, we use the 2010, 2012, and 2013 fall scaled reading and math state test scores. We chose this outcome variable, as the state test is considered a "high-stakes" test for schools under the No Child Left Behind Act, and the MSD certainly exerts great efforts to encourage students to perform well on the state test.

Table 1 shows summary statistics for the two estimation samples. The top panel shows descriptive statistics for the sixth graders who were assigned to double blocking in reading based on the cutoff of their fifth-grade 2009 scaled state test and those who were not. The bottom panel shows summary statistics for the sixth graders who were assigned to double blocking in math based on the cutoff of their fifth-grade 2009 scaled state test and those who were not. Note that in Table 1, the regression-discontinuity assignment variables for the two interventions, the 2009 scaled state tests in reading and math, have been centered on the cutoff for each intervention and then normalized by the statewide standard deviation for the test. We do this not to disclose the identity of the state based on the scaling of the scores.

For each intervention, we observe this cohort of sixth graders in the school years 2010–2011, 2011–2012, and 2012–2013. The population of middle-school students in the MSD is diverse, with two-thirds qualifying for a free or reduced lunch and about 50 percent being African-American.

Table 1 shows that based on the 2009 state scores, a large fraction of students qualified for double blocking in both reading and math, about 20 percent of students. We see that the fraction of students who participated in double blocking in both groups is smaller than the fraction that was intended to be assigned. This is because of two-sided noncompliance with the assignment protocol: some students who were meant to be double blocked managed to get exempt, while other students, who were not eligible, ended up participating.

We use the reading and math state test scores in the fall of 2010, 2011, and 2012 as outcome variables. This state test is the same test as the one in 2009 used to assign students to the interventions, and as we have done for the assignment variable, we center the test scores on the Midwest state's proficiency criteria in reading and math. In order to enable interpreting the impact of the interventions as effect sizes, the 2010, 2011, and 2012 state test scores have been divided by the statewide standard deviations.

Table 1 shows that the number of observations gets smaller the further away in time one gets from initial assignment to the program. This is because some students leave the MSD, and we do not observe them in the data after they move. The means in Table 1 do not change much over time, which suggests that this out-migration does not much affect the sample or estimates. However, we test further for possible differences in observable variables between the treatment and comparison group around the test-score cutoff for each school year (Table 2).

3.2 Methods

We measure the effect of double blocking on achievement by using a regression-discontinuity (RD) design. Specifically, we model postprogram student achievement outcomes as a smooth function of the fall 2009 test scores, a dummy variable for whether a student was below the double-blocking cutoff, and student characteristics.

We estimate a separate model for the effect of a double block in reading on reading and a separate model of the effect of a double block in math on math. Each model can be generalized by Equation (1):

(1) $Y = \alpha + \tau D + \mathrm{f}(\text{scaled score - cutoff}) + e.$

Y denotes one of the following outcome variables: the 2010, 2012, or 2013 fall scaled state test score in reading or math. D is a dummy variable that describes the assignment into the double-blocking program in reading or math and equals one if the scaled score is less than the cutoff and zero otherwise; f(.) is a flexible function determining the relationship between the 2009 scaled score (centered on the cutoff score) and the outcome variable Y; and e is a normally distributed error term.

The estimate of interest, τ, identifies the average effect of assignment to double blocking at the cutoff point for eligibility; Jacob et al. (2012) refer to this estimate as an intent-to-treat at the cutoff (ITTC). The model is estimated via an ordinary least squares (OLS) regression. We choose this parametric estimation approach because of the relatively small size (fewer than 800 observations) of our data set.

The function f(.) is specified as a polynomial in the centered 2009 state test score. Following Lee and Lemieux (2010), we allow the slopes of the polynomial to vary below the cutoff by interacting f(.) with the double-blocking assignment dummy, D. The order of the polynomial is determined using the Akaike and Bayesian information criteria. If the two criteria disagree, we choose the more parsimonious model. As the assignment variable is discrete, we follow Lee and Card's (2008) recommendation and cluster our standard errors on the value of the test score.

Because some students were exempt from double blocking, while other students who were ineligible participated in the program, we also estimate a treatment-on-the-treated (TT) effect by carrying out a two-stage least-squares estimation, in which we instrument the participation in double blocking with the assignment rule for the double-blocking program. Since both participation in double blocking and assignment to double blocking are binary indicator variables, this procedure will give an estimate of the local average treatment effect (LATE), that is, the effect of the double-blocking program on the students assigned to the program who actually participated compared to the students who were not assigned to double blocking and who did not participate (this effect is also sometimes called the causal average complier effect).

Similarly to a randomized trial in which the individual characteristics of the treated versus those in the control group are orthogonal to the treatment but are included in a regression to reduce residual variance, we can control for observable characteristics of the students. If the RD design is an appropriate method, controlling for observables should not change the estimate of τ; however, it may increase its precision. The available observables include gender, race and ethnicity, free and reduced lunch price status, and school fixed effects.

The research literature suggests that RD evaluations obtain estimates similar to what would have been obtained by random assignment experimentation, although typically with greater standard errors for a given sample size. Therefore, RD estimates are considered to have a high degree of internal validity. However, because RD is estimating effects for observations close to the threshold-determining program assignment, the estimates from an RD design often have limited external validity. Hence, the estimates are likely not to be generalized to populations other than students with relatively low achievement scores. Given that the current double-blocking policy is targeted at such students, this is not a significant limitation in evaluating the policy as currently designed.

4 RESULTS

RD design relies on the assumption that students cannot exactly manipulate the assignment variable so as to exactly place themselves to the left or to the right of the threshold. Hence, following the recommendations in Lee and Lemieux (2010), we begin with a graphical analysis to see if the density of the assignment variables, the 2009 scaled scores in reading and math, change at the double-blocking cutoffs.

Figure 1 shows the two distributions. When eyeballing the distributions, we do not see a clear change in the density. McCrary (2008) develops a formal test of whether the density differs around the cutoff. For the reading test, we cannot reject that the difference in the density is statistically significant (*t*-statistic equals 0.7). For the math test, the test statistic suggests a statically significant density difference at the cutoff (*t*-statistic equals 2.7).

Although this latter difference is disconcerting, the McCrary test assumes a continuous assignment variable, whereas our two assignment variables are inherently discrete. For example, the test picks a bin width equal to 2.18 for reading and 2.6 for math—smaller than the optimal bin width suggested below. However, as the results for math double blocking turn out to not be statistically different from zero, we focus the paper predominantly on the effects of double blocking in reading.

In Figure 2, we calculate an average value for each reading and math scaled score in 2010, 2011, and 2012 as a function of the assignment variable. We pick the optimal bin width using the “bin *F*-test” suggested in Lee and Lemieux (2010) and described in the notes to Figure 2. The graphs on the left side of Figure 2 trace the relationship between the assignment variable; the 2009 reading scaled score; and the 2010, 2011, and 2012 scaled reading scores. The graphs on the right do the same for the assignment variable to double blocking in math; the 2009 math scaled score; and the 2010, 2011, and 2012 scaled math scores. In each of the figures, we observe that the slope to the left of the cutoff (assigned to double blocking) is different from the slope to the right of the cutoff, but we do not observe a discrete “jump” in achievement level. This suggests that the simple graphical analysis may not be enough to discern an effect and needs to be augmented by a statistical model.

Before we proceed to the regression analysis of outcomes, in Table 2, we test for differences in observables around the cutoff for each school year of the sample and for different model specifications (linear, quadratic, and cubic polynomial in the assignment variable) using the reading double-block sample. If the RD design is sound, we should not detect a statistically significant discontinuity in the fraction of females, racial or ethnic composition, or free or reduced lunch status. If we do, it may raise doubt whether the student on either side of the cutoff are systematically different in unobserved ways and whether the estimated effect of the intervention is really due to the program. At the same time, we would expect that, for a large number of student characteristics, one out of ten characteristics to be different between those assigned and those not assigned at a 10 percent significance level.

Looking at Table 2, we observe that of the eight controls, students in School C are more likely to be assigned to RD than students in other schools and that in some specifications the students assigned to double block in reading are more likely to be African-American and male (at a 10 percent significance level). Although we would wish that these differences were not statistically different, they could arise because of the relatively small sample size we use. Our estimation results do not change much whether or not we include these controls.

In the Appendix A, we also conduct this analysis for the math double-block sample. There, we do not find that, conditional on the preferred polynomial assignment variable specification, the students are on average statistically different on either side of the cutoff.

4.1 The Estimated Effect of Assignment to a Double Block in Reading

Table 3 shows a set of intent-to-treat effects of double blocking on test scores in later years of middle school. Both the Akaike and Bayes information criteria (AIC and BIC) favor a quadratic specification for the reading test score in 2010 and 2011, but a cubic for 2012. These are our preferred models. However, in order to be transparent in presenting our results, we show the results from all specifications. We also show the results with and without controlling for student characteristics and school fixed effects.

In order to interpret the achievement gains and to make the scaled scores comparable across time, we report the "effect sizes" of assignment to double block in reading. This allows interpreting the estimated coefficient as a standard deviation change in the outcome variable. In our preferred quadratic specification in 2010, the intent-to-treat effect of double blocking in reading had a 0.17σ effect on reading scores. Controlling for student characteristics and school fixed effects, the reading gain is about 0.19σ, hence including these covariates does not change the point estimate of τ by a lot.

The estimated learning gain of assignment to a reading double block does not dissipate in the following year: the effect size ranges between 0.21 and 0.23σ in the quadratic specification and increases to about 0.41–0.43σ (using the cubic specification) in the third year following the initial assignment. In all regressions, controlling for student characteristics and school effects tends to increase the point estimates.

On the whole, the estimated effect sizes are typical of what one may expect of an educational intervention. Bloom, Hill, and Lipsey (2008) stress that when interpreting effect sizes, one needs to place them in the context of the educational intervention. The relevant context for this study is how the estimated effect size compares to the overall learning gain trajectory for students in middle school. Bloom, Hill, and Lipsey (2008) report that in middle school, the typical learning gain in reading ranges between 0.23 and 0.26σ, hence the estimated effects of double blocking in reading are large.

The estimated effect sizes do not dissipate over time, instead they trend up. We would expect the effect to be persistent, as many students initially assigned to double blocking in reading did not exit the program by the end of the first year.

4.2 The Estimated Effect of Participating in a Double Block in Reading

Table 4 presents estimated effects of participating (that is, the effect of treatment-on-the-treated) in the reading double block on reading test scores in the three years following initial assignment. We instrument the indicator for whether a student participated in the double block in reading with the assignment rule indicator and estimate this model using two-stage least squares (2SLS). We only report the estimates from both the quadratic and cubic specifications, since neither the AIC nor the BIC favor the linear specification.

Not unexpectedly, the treatment effect on the participants is greater than the effect of the initial assignment. The quadratic specifications suggest effect size increases of 0.28σ and 0.33σ in 2010 and 2011, respectively, and the cubic specification suggests a very big effect size, equal to 0.65σ.

4.3 The Estimated Effect of Double Blocking in Math

Tables 5 and 6 present the results from the intent-to-treat and treatment-on-the-treated RD analysis of double blocking in math. The AIC and BIC favor a quadratic specification for 2010 and 2011, whereas for 2012, AIC picks a cubic and BIC picks a linear specification. Since the two information criteria disagree, we choose the simpler model, that is, the linear specification for 2012. As with double blocking in reading, we also show the results with and without controlling for student characteristics and school fixed effects.

We do not find evidence of learning gains either when estimating the effect of assignment to a double block in mathematics or when estimating the effect of participation in the mathematics double block. Furthermore, although there is some statistical uncertainty, the results are precise enough to rule out large positive effects of double-blocking in mathematics. For example, the mathematics point estimates in Table 6 for treatment-on-the-treated are all precise enough to rule out effects as large as the 0.65σ found for reading. Even if we would find positive effects of double blocking in math, it is hard to argue that the evidence is convincing given the results from the McCrary test.

5 CONCLUSION

This paper uses a regression-discontinuity design to study whether double blocking sixth-grade students in reading and mathematics has an effect on reading and math in the three years following the initial intervention.

We find that assignment to and participation in double blocking in reading result in statistically positive and meaningful gains in reading that persist three years after the initial treatment assignment. In contrast to previous research, we do not find statistically significant gains in math for students assigned to a double block in that subject.

Based on our conversations with school administrators, one reason for the discrepancy in the efficacy of the reading and mathematics double-block programs could be the different intervention designs, with the mathematics intervention not being as systematic as the reading intervention. Unlike the students in the reading double block, students double blocked in math did not follow a separate curriculum but only received additional instructional time. As our paper focuses only on the effects of double blocking one cohort, we cannot preclude that later cohorts might have benefitted from a more systematic curriculum for double blocking in math.

Another important caveat is that we are studying the effects of the double-blocking program for the first cohort that was eligible for the program. Studying any educational or social program in its first year is perhaps unfair to the program, as one would think that program quality would improve with experience for programs. Certainly, policy makers should be hesitant to make final judgments on continuing a program based on its effectiveness in its first year. The school district implementing this double-blocking program is continuing to modify the program over time to improve its quality. A better long-term assessment of the potential for double blocking to improve achievement should be based on a program that has gone through its growing pains and reached maturity. As more data become available to us, we plan to study subsequent cohorts who participate in the modified double-blocking program.

REFERENCES

Abdulkadiroğlu, Atila, Weiwei Hu, and Parag A. Pathak. 2013. "Small High Schools and Student Achievement: Lottery-Based Evidence from New York City." NBER Working Paper No. 19576. Cambridge, MA: National Bureau of Economic Research.

Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114(2): 533–575.

Bloom, Howard S., Carolyn J. Hill, and Mark W. Lipsey. 2008. "Performance Trajectories and Performance Gaps as Achievement Effect–Size Benchmarks for Educational Interventions." MDRC Working Papers on Research Methodology. New York and Oakland, CA: MDRC.

Cavanagh, Sean. 2006. "Students Double-Dosing on Reading and Math." *Education Week* 25(40): 1–13.

Cortes, Kalena, Joshua Goodman, and Takako Nomi. 2013a. "A Double Dose of Algebra." *Education Next* 13(1): 70–76.

———. 2013b. "Intensive Math Instruction and Educational Attainment: Long-Run Impacts of Double-Dose Algebra." HKS Faculty Research Working Paper No. 13-09. Cambridge, MA: Harvard Kennedy School.

Dougherty, Shaun M. 2012. "Bridging the Discontinuity in Adolescent Literacy: Evidence of an Effective Middle Grades Intervention." Unpublished manuscript.

Durwood, Claire, Emily Krone, and Christopher Mazzeo. 2010. "Are Two Algebra Classes Better than One? The Effects of Double-Dose Instruction in Chicago." Consortium on Chicago School Research at the University of Chicago Urban Education Institute Policy Brief, 1–13.

Fryer, Roland G. 2011. “Financial Incentives and Student Achievement: Evidence from Randomized Trials.” *Quarterly Journal of Economics* 126(4): 1755–1798.

———. 2012. “Injecting Successful Charter School Strategies into Traditional Public Schools: Early Results from an Experiment in Houston.” NBER Working Paper No. 17494. Cambridge, MA: National Bureau of Economic Research.

Goodman, Joshua. 2008. “Who Merits Financial Aid?: Massachusetts’ Adams Scholarship.” *Journal of Public Economics* 92(10): 2121–2131.

Jacob, Brian, and Lars Lefgren. 2004. “Remedial Education and Student Achievement: A Regression Discontinuity Analysis.” *Review of Economics and Statistics* 86(1): 226–244.

Robin Jacob, Pei Zhu, Marie-Andreé Somers, and Howard Bloom. 2012. “A Practical Guide to Regression Discontinuity.” MDRC Working Papers on Research Methodology. New York and Oakland, CA: MDRC.

Lee, David S., and David Card. 2008. “Regression Discontinuity Inference with Specification Error.” *Journal of Econometrics* 142(2): 655–674.

Lee, David S., and Thomas Lemieux. 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature* 48(2): 281–355.

Matsudaira, Jordan. 2008. “Mandatory Summer School and Student Achievement: A Regression Discontinuity Analysis.” *Journal of Econometrics* 142(2): 829–885.

McCrary, Justin . 2008. “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test.” *Journal of Econometrics* 142(2): 698–714.

Nomi, Takako, and Elaine Allensworth. 2009. “‘Double-Dose’ Algebra as an Alternative Strategy to Remediation: Effects on Students’ Academic Outcomes.” *Journal of Research on Educational Effectiveness* 2(2): 111–148.

van der Klaauw, Wilbert. 2002. “Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression Discontinuity Approach.” *International Economic Review* 43(4): 1249–1287.

Results

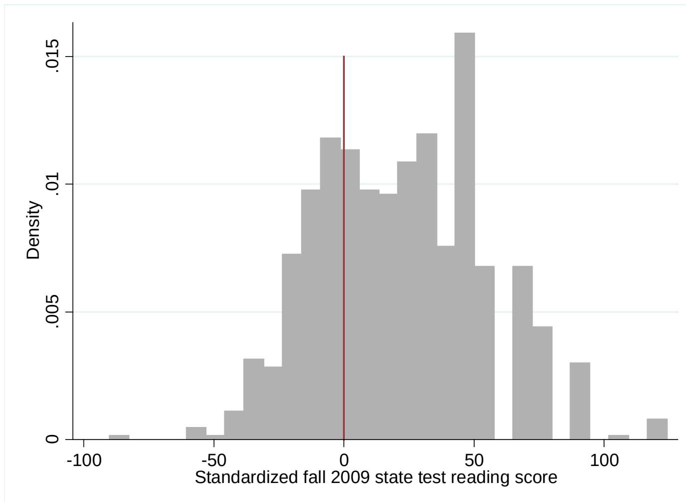
Table 1 Summary Statistics

School year	2010–2011		2011–2012		2012–2013	
Universe: 2009 rising 6th graders						
Statistic	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Controls						
Female	0.51	0.50	0.52	0.50	0.51	0.50
Free or reduced lunch	0.68	0.47	0.67	0.47	0.65	0.48
Race/ethnicity						
White	0.36	0.48	0.36	0.48	0.36	0.48
Black	0.50	0.50	0.50	0.50	0.51	0.50
Hispanic	0.10	0.30	0.11	0.31	0.10	0.31
Other race	0.04	0.19	0.03	0.18	0.03	0.18
School						
School A	0.21	0.41	0.21	0.41	0.22	0.41
School B	0.26	0.44	0.26	0.44	0.25	0.44
School C	0.27	0.45	0.27	0.45	0.28	0.45
School D	0.26	0.44	0.25	0.44	0.25	0.43
Regression discontinuity variables						
Assigned to double block in reading	0.26	0.44	0.25	0.43	0.23	0.42
Double blocked in reading	0.22	0.42	0.21	0.41	0.20	0.4
2009 state test reading score (centered)[a]	0.89	1.28	0.92	1.26	0.95	1.25
Outcome variables						
State test reading score (centered)[a]	0.87	1.19	-0.03	1.29	0.18	1.07
Observations	786		740		683	
Universe: 2009 rising 6th graders who passed the 2009 state test in reading						
Statistic	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Controls						
Female	0.54	0.50	0.55	0.50	0.54	0.50
Free or reduced lunch	0.60	0.49	0.59	0.49	0.58	0.49
Race/ethnicity						
White	0.44	0.50	0.43	0.50	0.43	0.50
Black	0.43	0.49	0.43	0.50	0.44	0.50
Hispanic	0.09	0.29	0.09	0.29	0.09	0.29
Other race	0.05	0.22	0.04	0.20	0.04	0.20
School						
School A	0.20	0.40	0.21	0.40	0.21	0.41
School B	0.25	0.43	0.25	0.43	0.25	0.43
School C	0.28	0.45	0.29	0.45	0.29	0.45
School D	0.27	0.44	0.26	0.44	0.26	0.44
Regression discontinuity variables						
Assigned to double block in math	0.23	0.42	0.23	0.42	0.24	0.43
Double blocked in math	0.18	0.38	0.17	0.38	0.17	0.38
2009 state test math score (centered)[a]	0.94	1.31	0.95	1.32	0.95	1.35
Outcome variables						
State test math score (centered)[a]	0.86	0.98	-0.31	1.00	-0.43	1.26
Observations	577		556		524	

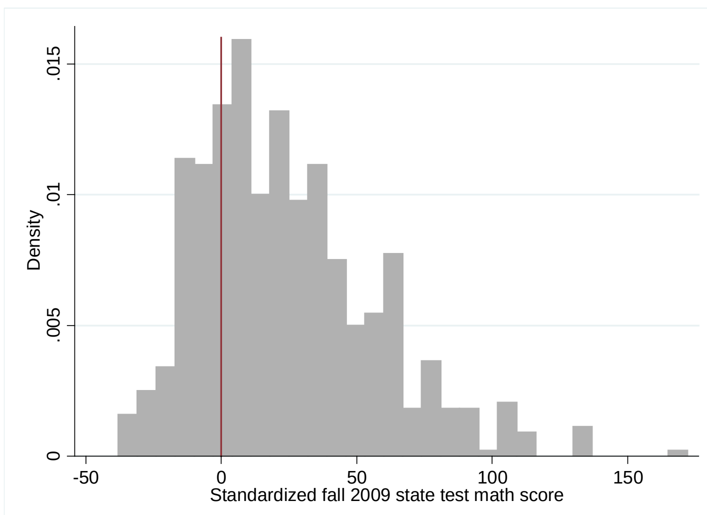
NOTE: Mean and standard deviation of each variable shown by the estimation sample: school year 2010–2011, school year 2011–2012, and school year 2012–2013.
[a]State test scores have been centered on the state requirement for proficiency in reading (or math) and divided by the statewide standard deviation.

Figure 1 Distribution of the 2009 State Test Scores for Rising 6th Graders, Centered at the Double-Block Cutoff for Reading and Math

Distribution of 2009 State Test Score in Reading

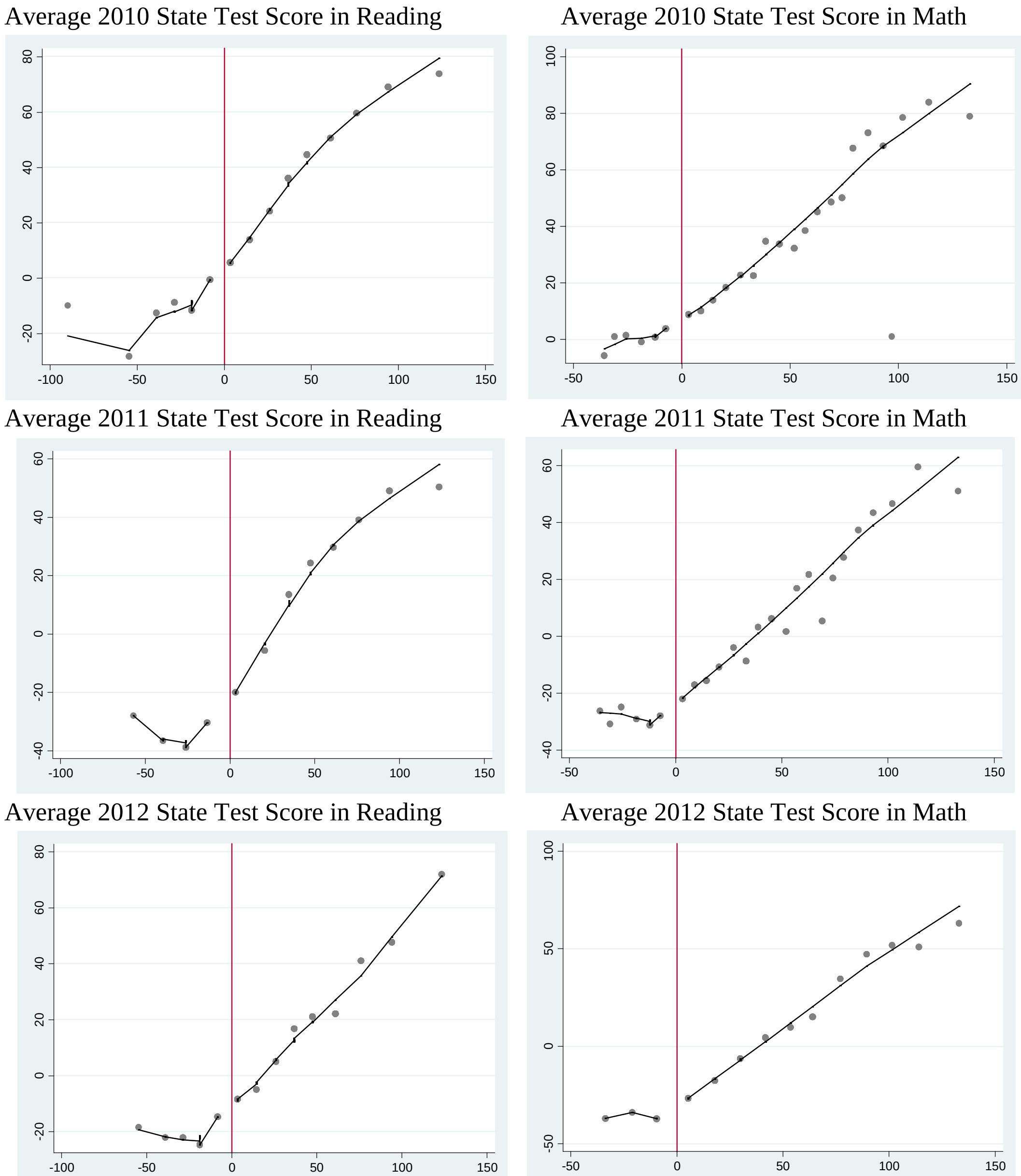


Distribution of 2009 State Test Score in Math



NOTE: Probability density of the 2009 state test scores around the double-block reading and math cutoffs. Vertical line shows the cutoff. The McCrary t-test statistic for the log-difference in the probability density around the cutoff of the 2009 state test score in reading equals 0.712. The McCrary t-test statistic for the log-difference in the probability density around the cutoff of the 2009 state test score in math equals 2.734.

Figure 2 Average State Test Scores in School Years 2010–11, 2011–12, and 2012–13 as a Function of Assignment to Double Block in Reading or Math



NOTE: Predicted average test scores on each side of the reading and math cutoff. Each figure shows a “lowess” (locally weighted scatterplot smoothing) regression on either side of the double-block cutoff. The bin width is chosen using the “bin size” test suggested by Lee and Lemieux (2010). For a given number of bins, this test adds twice as many bins until the additional bins are jointly insignificant at a 5 percent statistical level. For the raw 2010 state test scores in reading, the bin width is 11, for 2011 it is 14, and for 2012 it is 11. For the raw 2010 state test scores in math, the bin width is 6, for 2011 it is 6, and for 2012 it is 12.

Table 2 Differences in Means of Control Variables around the Cutoff for Assignment to Double Block in Reading

	2010–2011 sample							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Female	White	Black	Hispanic	FRL	School A	School B	School C
Variable	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Assigned to double block	-0.121*	0.074	-0.096*	-0.007	0.010	-0.047	-0.078	0.234***
	(0.066)	(0.074)	(0.050)	(0.070)	(0.087)	(0.064)	(0.063)	(0.081)
Observations	786	786	786	786	786	786	786	786
R-squared	0.012	0.188	0.166	0.013	0.182	0.010	0.019	0.022
Mean of the dependent variable	0.509	0.359	0.499	0.103	0.683	0.209	0.257	0.275

	2011–2012 sample							
	Female	White	Black	Hispanic	FRL	School A	School B	School C
Variable	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Assigned to double block	-0.083	0.103	-0.070	-0.027	0.027	-0.070	-0.002	0.223**
	(0.070)	(0.089)	(0.069)	(0.074)	(0.087)	(0.074)	(0.079)	(0.088)
Observations	740	740	740	740	740	740	740	740
R-squared	0.017	0.192	0.167	0.012	0.184	0.010	0.023	0.027
Mean of the dependent variable	0.516	0.362	0.499	0.105	0.666	0.212	0.261	0.273

	2012–2013 sample							
	Female	White	Black	Hispanic	FRL	School A	School B	School C
Variable	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Assigned to double block	-0.054	-0.113	0.221***	-0.086	0.005	-0.101	-0.013	0.408***
	(0.113)	(0.094)	(0.075)	(0.074)	(0.095)	(0.120)	(0.080)	(0.116)
Observations	683	683	683	683	683	683	683	683
R-squared	0.022	0.200	0.171	0.019	0.176	0.014	0.034	0.032
Mean of the dependent variable	0.511	0.359	0.505	0.104	0.649	0.220	0.253	0.275

NOTE: Standard errors clustered by 2009 state test score in reading are in parentheses (*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$). The regressions control for a quadratic polynomial in the assignment variable, the 2009 state test score in reading, which is allowed to vary differentially to the left of the cutoff. The order of the polynomial is chosen by the Akaike Information Criterion: quadratic in 2010–11 and 2011–12, and a cubic in 2012–13.

Table 3 Estimated Effect Sizes of Assignment to Double Block in Reading on Reading

Variables	(1) Linear No controls	(2) Linear Controls	(3) Quadratic No controls	(4) Quadratic Controls	(5) Cubic No controls	(6) Cubic Controls
	Reading state test score 2010					
Assigned to double block	-0.175*	-0.143	0.171*	0.195**	0.220*	0.266**
	(0.098)	(0.095)	(0.096)	(0.089)	(0.109)	(0.098)
Observations	786	786	786	786	786	786
R-squared	0.597	0.61	0.607	0.62	0.607	0.62
	Reading state test score 2011					
Assigned to double block	-0.225**	-0.172*	0.211*	0.231**	0.178	0.236*
	(0.106)	(0.100)	(0.113)	(0.107)	(0.133)	(0.127)
Observations	740	740	740	740	740	740
R-squared	0.591	0.623	0.605	0.636	0.605	0.636
	Reading state test score 2012					
Assigned to double block	-0.077	-0.049	0.161	0.151	0.411***	0.430***
	(0.116)	(0.104)	(0.113)	(0.111)	(0.126)	(0.121)
Observations	683	683	683	683	683	683
R-squared	0.515	0.551	0.518	0.553	0.521	0.557

NOTE: Standard errors clustered by 2009 state test score in reading are in parentheses (*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$). The regressions control for a flexible polynomial (the order of the polynomial is denoted in each column) in the 2009 state test score in reading, which is allowed to vary differentially to the left of the cutoff. Columns (2), (4), and (6) additionally control for gender, race/ethnicity, free or reduced lunch status, and school fixed effects. The preferred order of the polynomial chosen by the Akaike Information Criterion is a quadratic for years 2010–11 and 2011–12 and a cubic for 2012–13.

Table 4 Estimated Effect Sizes of Being Double Blocked in Reading (i.e., Treatment-on-the-Treated) on Reading

Variables	(1) Reading state test score 2010 Quadratic	(2) Reading state test score 2010 Cubic	(3) Reading state test score 2011 Quadratic	(4) Reading state test score 2011 Cubic	(5) Reading state test score 2012 Quadratic	(6) Reading state test score 2012 Cubic
Double blocked in reading	0.281**	0.362***	0.332**	0.355*	0.219	0.650***
	(0.120)	(0.120)	(0.153)	(0.196)	(0.161)	(0.172)
Observations	786	786	740	740	683	683
R-squared	0.617	0.615	0.631	0.63	0.551	0.542
First-stage F-statistic	276	231.3	302.6	175.2	314.7	140.6

NOTE: Standard errors clustered by 2009 state test score in reading are in parentheses (*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$). The regressions control for a flexible polynomial (the order of the polynomial is denoted in each column) in the 2009 state test score in reading, which is allowed to vary differentially to the left of the cutoff. Additional controls include gender, race/ethnicity, free or reduced lunch status, and school fixed effects.

Table 5 Estimated Effect Sizes of Assignment to Double Block in Math on Math

Variables	(1) Linear No controls	(2) Linear Controls	(3) Quadratic No controls	(4) Quadratic Controls	(5) Cubic No controls	(6) Cubic Controls
	Math state test score 2010					
Assigned to double block	-0.018	-0.017	0.063	0.091	0.017	0.092
	(0.060)	(0.079)	(0.078)	(0.095)	(0.091)	(0.145)
Observations	577	577	577	577	577	577
R-squared	0.625	0.645	0.626	0.645	0.631	0.652
	Math state test score 2011					
Assigned to double block	-0.181*	-0.143	0.058	0.134	-0.084	-0.059
	(0.092)	(0.092)	(0.097)	(0.096)	(0.141)	(0.135)
Observations	556	556	556	556	556	556
R-squared	0.63	0.662	0.632	0.665	0.639	0.672
	Math state test score 2012					
Assigned to double block	-0.288***	-0.207*	-0.088	0.015	-0.095	-0.094
	(0.103)	(0.123)	(0.118)	(0.163)	(0.147)	(0.184)
Observations	524	524	524	524	524	524
R-squared	0.594	0.632	0.596	0.634	0.6	0.637

NOTE: Standard errors clustered by 2009 state test score in math are in parentheses (*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$). The regressions control for a flexible polynomial (the order of the polynomial is denoted in each column) in the 2009 state test score in math, which is allowed to vary differentially to the left of the cutoff. Columns (2), (4), and (6) additionally control for gender, race/ethnicity, free or reduced lunch status, and school fixed effects. The preferred order of the polynomial chosen by the Akaike Information Criterion is a cubic for year 2010–11 and 2011–12 and a linear for 2012–13.

Table 6 Estimated Effect Sizes of Being Double Blocked in Math (i.e., Treatment-on-the-Treated) on Math

Variables	(1) Math state test score 2010 Cubic	(2) Math state test score 2011 Cubic	(3) Math state test score 2012 Linear
Double blocked	0.118	-0.079	-0.301
	(0.175)	(0.173)	(0.190)
Observations	577	556	524
R-squared	0.648	0.672	0.632
First-stage F-statistic	36.97	30.11	74.31

NOTE: Standard errors clustered by 2009 state test score in math are in parentheses (*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$). The regressions control for a flexible polynomial (the order of the polynomial is denoted in each column) in the 2009 state test score in math, which is allowed to vary differentially to the left of the cutoff. Additional controls include gender, race/ethnicity, free or reduced lunch status, and school fixed effects.

Appendix A

Table A1 Differences in Means of Control Variables around the Cutoff for Assignment to Double Block in Math

	2010–2011 sample							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Female	White	Black	Hispanic	FRL	School A	School B	School C
Variable	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Assigned to double block	-0.088	-0.034	0.167	0.048	-0.109	0.038	0.085	0.038
	(0.213)	(0.120)	(0.131)	(0.104)	(0.101)	(0.078)	(0.100)	(0.203)
Observations	577	577	577	577	577	577	577	577
R-squared	0.010	0.098	0.120	0.019	0.122	0.016	0.014	0.033
Mean of the dependent variable	0.536	0.435	0.426	0.0901	0.603	0.203	0.246	0.284
	2011–2012 sample							
	Female	White	Black	Hispanic	FRL	School A	School B	School C
Variable	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Assigned to double block	-0.097	-0.071	0.227	0.009	-0.130	0.043	0.166	0.007
	(0.195)	(0.153)	(0.150)	(0.091)	(0.084)	(0.068)	(0.108)	(0.197)
Observations	556	556	556	556	556	556	556	556
R-squared	0.013	0.104	0.122	0.012	0.126	0.016	0.017	0.030
Mean of the dependent variable	0.547	0.433	0.430	0.0935	0.586	0.205	0.252	0.288
	2012–2013 sample							
	Female	White	Black	Hispanic	FRL	School A	School B	School C
Variable	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Assigned to double block	-0.056	-0.019	0.026	-0.027	0.104	-0.132**	-0.119*	0.148
	-0.087	-0.069	-0.068	-0.057	-0.065	-0.052	-0.069	-0.092
Observations	524	524	524	524	524	524	524	524
R-squared	0.004	0.097	0.118	0.003	0.139	0.017	0.009	0.03
Mean of the dependent variable	0.542	0.427	0.439	0.0916	0.58	0.208	0.248	0.288

NOTE: Standard errors clustered by 2009 state test score in math are in parentheses (*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$). The regressions control for a polynomial in the assignment variable, the 2009 state score in math, which is allowed to vary differentially to the left of the cutoff. The order of the polynomial is chosen by the Akaike Information Criterion: a cubic in 2010–11 and 2011–12 and a linear in 2012–13.