

BRIDGING THE DISCONTINUITY IN ADOLESCENT LITERACY? MIXED EVIDENCE FROM A MIDDLE GRADES INTERVENTION

Shaun M. Dougherty

Neag School of Education
University of Connecticut
Storrs, CT 06269-3093
shaun.dougherty@uconn.edu

Abstract

Strong literacy skills are crucial to ensuring an individual's future educational and economic success. Existing evidence suggests the transition from elementary to middle school is a decisive period for literacy development. In this paper I investigate the impact of extended learning time in literacy instruction on subsequent cognitive outcomes. I capitalize on the existence of a natural experiment born out of a district's use of an exogenously-determined cutoff in Iowa Test scores in fifth grade to assign students to an additional literacy course in middle school. My findings suggest that exposure to this intervention generates strong negative impacts for black students, and noisy positive impacts for white, Latino, and Asian students. My findings suggest that additional literacy instruction in middle school can have markedly different effects on students, and program differentiation or augmentation may be necessary to prevent harm for students of average literacy ability in fifth grade.

1. INTRODUCTION

The capacity of school districts to support the ongoing development of their students' literacy skills plays a critical role in enhancing their academic and labor-market outcomes. Though the average fourth graders' reading scores on the National Assessment of Educational Progress have been trending higher, Snow and Moje (2010) point out that score trends are flat among average eighth and twelfth graders (Lee, Grigg, and Donahue 2007, p. 3). These trends underscore the need for literacy support at the critical transition between elementary and secondary schooling (Chall and Jacobs 2003), not only for the most struggling readers but for students across the performance distribution. The transition from elementary to middle school is particularly critical for boys and students of color (Porche, Ross, and Snow 2004; Tatum 2008), and so extra care and attention is warranted in supporting both of these groups of students.

Schools and districts seeking to improve their adolescent literacy outcomes face resource constraints. Recent budget crises and mounting pressure from the requirements of No Child Left Behind (NCLB) necessitate that schools find ways to leverage existing resources and generate results in short timeframes. One widely used though undervalued method for improving student outcomes is increased learning time, especially in the tested areas of English-language arts and mathematics. To date there is mixed evidence as to whether increasing learning time overall or in specific subject areas can produce favorable impacts on student outcomes. For instance, Lavy (2010) finds a positive association between increased learning time and internationally benchmarked exam scores. In related work, several papers have used random variation in the length of a student school year to show that more time in school results in higher test score outcomes (Goodman 2014; Hansen 2011; Marcotte and Hemelt 2008; Sims 2008; Fitzpatrick, Grismmer, and Hastedt 2011).

Other evidence has focused on evaluating the provision of a "double dose" of instruction in subject areas tested for the purposes of NCLB, most notably reading and mathematics. Recent evidence from Chicago suggests increased exposure to algebra instruction and favorable-ability groupings can have positive short-term impacts on a student's academic performance, as well as positive longer-run impacts on high-school graduation and post-secondary enrollment (Nomi and Allensworth 2009; Cortes, Goodman, and Nomi 2012). Other recent work examining the effects of double dose strategies in mathematics has also found positive effects of extending the learning time in mathematics (Taylor 2012). Nevertheless, little is known about the effectiveness of expanded learning time in literacy skills as a way to boost literacy outcomes, despite evidence that such strategies have been and are used throughout the country (Cavanagh 2006; Mazzolini and Morley 2006; Paglin 2003; Wanzek

and Vaughn 2008). The paucity of good evidence on the effectiveness of literacy interventions at the key transition from elementary to middle grades is particularly notable in that what little evidence exists is not causal.

I fill this gap in the literature by providing causal evidence for the effectiveness of extended learning time in English-language arts instruction in middle school. I focus on an intervention where the additional course in literacy instruction uses research-based instructional strategies as a supplement to a typical English course. This program is designed for students who score near the national average on a fifth-grade measure of literacy and is taken in place of an introductory world-language course in middle school. I show this additional instruction leads to systematic improvement in adolescent reading comprehension for many students, but has a negative effect on the literacy performance of black students. Using a rich set of data from Hampton County Public Schools—one of the nation’s forty largest school districts, located in the southeast United States—I estimate the impact of a district-developed, classroom-based literacy intervention in middle school on both immediate and medium-term student test scores.¹ Specifically, I investigate whether, and by how much, participation in a supplementary reading class in middle school improved student test scores in reading and mathematics for black students as well as their white, Latino, and Asian counterparts.

Hampton County provided a desirable setting to evaluate the impact of a research-based and district-designed literacy intervention. In the district, student eligibility for the supplemental reading class was made using a cutoff rule based on a student’s fifth-grade test score, allowing me to use a regression-discontinuity design to obtain an unbiased estimate of the causal impact of the intervention on student outcomes for those students near the cutoff that determined eligibility. The student’s position relative to this cutoff provided an indicator of eligibility for the literacy intervention. Because not all students who were eligible to receive the intervention actually participated, I use a fuzzy regression discontinuity design where I used the indicator of eligibility to instrument for their “take-up” of the supplementary reading intervention. Thus, using a two-stage least squares estimation strategy with instrumental variables, I was able to identify the causal impact of eligibility for and enrollment in the program for students near, but on opposite sides of, the cutoff.

I find that aggregate effects of the program for students on the margin of eligibility appear to produce zero impact on test-score performance, but that there are distinct and opposite impacts by race. Specifically, black students on the margin of eligibility experience a large negative impact on test scores

1. Note that I use a pseudonym for the district to reduce potential negative impacts associated with my mixed results.

by participating in the supplemental reading program, whereas students in other racial groups (mostly white) experience smaller positive impacts. The effects of the additional literacy instruction on reading test scores in sixth grade are smaller and not statistically significant, and the impact on reading scores in seventh and eighth grades is large and significant. For black students these effects are negative, whereas for white, Latino, and Asian students the effects are positive, though with similar magnitude. The harm and benefit experienced by these students on the state assessments of literacy appear to extend to the Iowa Test of Basic Skills (ITBS) in eighth grade as well, indicating the measured learning impacts suggest generalizable impact and not artifacts of potential teaching to the test.

I have laid out the rest of the paper in four sections. In the next section, I consider the district's theory of action with respect to the extant literature on effective instructional strategies that promote adolescent literacy, and describe the school district setting and their implementation of the supplementary reading program itself. In section 3, I present my research design, including a description of my data collection and data-analytic strategy, followed by my results in section 4. In the final section of the paper, I discuss potential threats to both the internal and external validity of my findings and review the implications of my findings for practice and future research.

2. BACKGROUND AND CONTEXT

Background on the Intervention

For the last twenty years, the Hampton County Public School district (HCPSD) has adapted its approach to meeting the instructional needs of its students in literacy, as the policy environment has shifted around it. Initially, the district utilized a supplementary reading program as a means to improve the literacy skills of its students as they transition from primary to secondary schooling. The district-maintained reading lab was designed to provide instructional support in literacy for students in the late elementary and early secondary grades. This lab supported students outside of their regular course of instruction, but in the 1990s the district moved to embed literacy support within an established course of instruction. Some of this change was motivated by standards-based reforms that changed the way instructional targets, or standards, were defined (McLaughlin and Shepard 1995; Darling-Hammond 2004). The importance of the course was further underscored when the policy landscape was modified again in 2001 by the passage of NCLB and the implementation of high-stakes, standards-based testing that began in the 2002–03 school year. In response to these changes, the HCPSD revised its instructional strategy to meet the needs of its students and to ensure that its schools satisfy, among other things, the adequate yearly progress provision of NCLB.

Each of the district's nineteen middle schools serves students in grades 6 through 8. In all district middle schools, students must earn a passing grade in a language-arts course to fulfill their annual English requirement. Language-arts courses address all of the state standards' domains: reading, writing, literary conventions, listening, speaking, and viewing. To address these domains, the language-arts classes use a literature anthology, a grammar text, and selected novels assigned specifically by grade level. The supplementary reading course was designed to complement a student's language arts curriculum by focusing only on the reading standards, and the standards for writing in response to reading, and to improve the development of a student's literacy skills to levels consistent with grade-level expectations. Teachers address the reading standards in the supplementary reading classes using grade-level-appropriate nonfiction texts and novels. Though similar instructional strategies are pursued in both classes, the literacy course takes a more narrow focus on reading strategies, and the language-arts course takes a broader focus.

Theory of Change and Recent Literature

The theory of change used by the HCPSD is that enrolling students who have demonstrated a need for additional literacy support in a course that was designed to utilize research-proven strategies is likely to improve literacy outcomes for those students. Specifically, this district drew on research such as Dole et al. (1991), and designed the supplementary reading class to explicitly dwell on seven "basic" reading strategies: activating background knowledge, questioning the text, drawing inferences, determining importance, creating mental images, repairing understanding when meaning breaks down, and synthesizing information. In addition, the district also encouraged the use of writing activities to support each of these seven reading strategies.

Though the research from Dole et al. is more than twenty years old, more recent research continues to substantiate the use of these strategies, particularly with adolescents. A meta-analysis on the effectiveness of reading interventions for struggling readers in grades 6 through 12 revealed that many of the same strategies suggested by Dole et al. were used across the thirteen studies that could be included in that meta-analysis (Edmonds et al. 2009). This meta-analysis found a large effect size of 0.89 standard deviation (SD) for reading comprehension outcomes. Evidence from another recent meta-analysis on writing to read further supports the strategies used by the HCPSD. Graham and Hebert (2012) found that writing to read strategies improve student reading comprehension by about 0.37 SD. In yet another teacher-delivered intervention, Vaughn et al. (2011) performed an experimental evaluation of collaborative strategic reading with middle school students, where

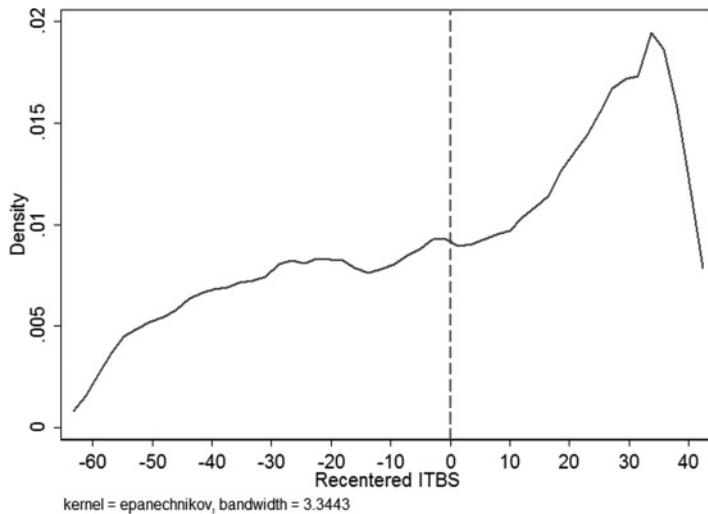


Figure 1. Kernel Density Plot of the Recentered ITBS Score in Fifth Grade

English-language arts teachers provided a multicomponent reading comprehension instruction twice a week for eighteen weeks, and found modest positive effects on reading comprehension. All of this more recent evidence suggests the research used to design the supplementary reading class continues to be valid and relevant.

Though HCPSD uses instructional strategies to improve the performance of all struggling readers, the supplementary reading intervention that I evaluate here is not a remedial intervention as it focuses on students who performed at the 60th percentile nationally in fifth grade on the ITBS. Despite the fact that these students achieve average performance on a national scale, the students who are the target of the intervention in Hampton are well below the mean performance in the district (see figure 1). The purpose of the intervention is to ensure progress at grade level, and is part of a district effort to ensure that students are equipped with the literacy skills that will allow them to complete at least their high school education. The supplementary reading course in Hampton is taken in place of a world language course. So, rather than beginning an exploratory language program in grade 6, a student who participates in the literacy intervention is likely to delay study of a world language until he exits the supplementary reading program. This substitution away from world language participation is simply a delay in the start of exposure. Though earlier exposure to other languages has been shown to make for better language acquisition, waiting until as late as high school to begin studying a language does not preclude a student from taking four years of high-school level language study, which is a common requirement to be admitted to more competitive colleges.

For this reason the substitution enforced by this literacy intervention poses little risk of negative longer-term consequences, while carrying the upside risk of improved reading skills in English.

Though Hampton does offer a remedial intervention, it is limited to those students who score in the lowest NCLB-defined performance category on the reading assessment in grade 5. The assignment mechanism for this intervention also lends itself to being evaluated using a regression discontinuity (RD) design, although there are so few students in this end of the distribution that any potential effects cannot be identified with sufficient statistical power to make it feasible. Though many schools and districts would like to know what interventions can be effective in serving their neediest students, it is similarly important to understand what can be done to keep students on track toward graduating from high school and being prepared to enroll in post-secondary education. Students on the margin of the 60th percentile on the ITBS nationally are certainly not guaranteed to graduate from high school, nor are they certain to be college-bound, particularly in light of the fact that only 30 percent of the U.S. population holds a bachelor's degree. It stands to reason that falling further behind during middle school may have a strong negative impact on a student's longer-term outcomes, and so understanding the impact of this intervention remains policy-relevant.

Assignment to the Supplementary Reading Program

Students in the HCPSD were assigned to receive supplementary reading instruction in middle school based on how they scored on the ITBS in reading during their fifth-grade year. Students who scored at or below the nationally defined 60th percentile on ITBS in reading were assigned, by rule, to complete the supplementary reading program in middle school. The HCPSD policy was designed to enroll students in the supplementary reading course for all three (grades 6, 7, and 8) years of middle school, with the goal of preparing students to meet proficiency requirements on the criterion-referenced eighth-grade state test in reading (used in making decisions about grade promotion), and on the norm-referenced eighth-grade administration of the ITBS in reading. Students not identified to participate in the reading intervention could elect to take a reading course or enroll in an exploratory world-language course.

3. RESEARCH DESIGN

Site, Data Set, and Sample

The HCPSD is a large suburban school district in the southeastern United States. My data are drawn from a comprehensive administrative data set covering all students enrolled in the district during the school years of 1999–2000

through 2009–10. This data set contains test scores and enrollment data for students in middle school and follows them longitudinally within the district. The data include course enrollment data; mandated state accountability test scores in reading, literature, and mathematics; and ITBS scores from grades 5 and 8. Hampton County resembles the changing demographic structure of many suburban settings, with substantial racial and socioeconomic variation. The student population is 33 percent white, 42 percent African American, 13 percent Latino, 9 percent Asian, and 4 percent identify as either Native American or multiracial. Forty-four percent of students receive free- or reduced-price lunch, 8 percent are English language learners, and 18 percent have an Individualized Education Program.

The district is composed of schools classified as traditional, charter, converted charter, and alternative schools. Only traditional and conversion charter schools are subject to district policies, and alternative and other charters are exempt. I restrict my analysis to students who go through one of the nineteen traditional or conversion charter middle schools that serve Hampton students in grades 6 through 8 and were subject to the policy. My sample includes all students from the five cohorts who took the fifth-grade ITBS reading test in the school years 2002–03 through 2006–07.

Measures

My academic outcomes of interest are state test scores in reading and mathematics in grades 6 through 8 (*READ6*, *READ7*, *READ8*, *MATH6*, *MATH7*, *MATH8*). For each of these outcomes I wish to estimate the effect of participating in the supplementary reading intervention, which I measure as the ratio of total semesters that a student is enrolled in supplementary reading to the total number of semesters that a student has been in middle school (*SUPREAD*). This variable has a minimum at 0 for students who enroll in no semesters of the supplementary reading course, and a maximum of 1 for those who participate for all possible semesters to that particular grade in middle school (see Angrist and Imbens 1995 for a discussion about the benefits of using a continuous measure for first-stage exposure). For instance, seventh-grade students have experienced a maximum of four semesters (assuming they were not retained in grade) and could have enrolled in the supplementary reading program for between zero and four semesters. Because student eligibility for the reading intervention is conditional on their fifth-grade ITBS percentile score, I also include this measure (*ITBS5*) as the forcing variable, as well as a binary indicator (*ELIG*) equal to one if a student scored at or below the 60th percentile on the fifth-grade ITBS, and is therefore eligible to receive the supplementary reading instruction. To improve the precision of my estimates

I include a vector of student covariates, \mathbf{X}_{ijs} . This vector includes indicators for sex, race, free and reduced-lunch status, special education status, and English language learner status. Despite designing the literacy intervention as a district-level policy, the application of the policy may vary based on the individual behavior of school administrators. For instance, individual schools may be more or less stringent in their requirement that students who are eligible for supplementary reading take up the treatment. Likewise, adherence to a long-standing policy may experience drift over time. To control for potential differences in the implementation of the literacy intervention across schools and cohorts of students, I additionally include school-by-year fixed effects (γ_{js}).

Statistical Model

Within my regression-discontinuity research design, I use a two-stage least squares approach to estimate the causal effect of participating in supplementary reading while in middle school. Because take-up of the supplementary reading instruction is potentially endogenous, I use the random offer of eligibility in the program, generated by a student's position relative to the 60th-percentile cutoff, to isolate the exogenous variation in participation. In my first stage, I fit the following statistical model:

$$\begin{aligned} SUPREAD_{ijs} = & \beta_0 + \beta_1 CITBS5_{ijs} + \beta_2 ELIG_{ijs} + \beta_3 CITBS5 \times ELIG_{ijs} \\ & + \mathbf{\Gamma}' \mathbf{X}_{ijs} + \gamma_{js} + \varepsilon_{ij}. \end{aligned} \quad (1)$$

I model the proportion of total semesters of middle school that a student is enrolled in the supplementary reading course ($SUPREAD_{ijgs}$), for student i in school j and cohort s . I estimate this participation variable as a function of students' fifth-grade ITBS score re-centered at the 60th percentile cutoff score $CITBS5_{ijs}$, the exogenous instrument, $ELIG_{ijs}$, the vector of student-level covariates (\mathbf{X}_{ijs}), and fixed effects for school and cohort. As is typical in the RD design literature, to allow the relationship between supplementary reading participation and fifth-grade ITBS score to vary on either side of the exogenous cutoff, I also include the interaction term $CITBS5 \times ELIG_{ijs}$ (Murnane and Willett 2011). To model potential heterogeneity of the effect of the program by race, I fit separate models by whether a student is African American or not.²

2. To achieve greater statistical power I create a second instrumental variable, $BLACK \times ELIG$, and a variable that allows take-up of the supplemental reading program to vary by race, $BLACK \times SUPREAD$. Using these two additional variables I can then use my two-stage least squares approach to fit two first-stage models (one each for the outcomes $SUPREAD$, and $BLACK \times SUPREAD$) with my two instruments $ELIG$ and $BLACK \times ELIG$. My instrumental variable results are not sensitive to the approach that I use.

Following the suggestion of Card and Lee (2008) I model the error structure to account for the clustering of students by values of the discrete forcing variables.

In the second stage of my estimation, I use the following statistical model:

$$Y_{ijs} = \theta_0 + \theta_1 CITBS5_{ijs} + \theta_2 SUPREAD_{ijs} + \theta_3 CITBS5 \times ELIG_{ijs} + \Psi' X_{ij} + \gamma_{js} + \mu_{ij}. \quad (2)$$

In this model I estimate Y_{ijgs} , a generic placeholder for my several outcomes of interest, as a function of the recentered fifth-grade ITBS score, student exposure to the supplementary reading course, as well as a vector of student covariates and fixed effects for school and cohort. As in my first stage, I also allow the slope of relationship between ITBS score and the outcome to vary on either side of the cutoff. Importantly, because the take-up of supplementary reading is endogenous, I use the fitted values of $SUPREAD_{ijgs}$ from my first-stage model to isolate the variation in this treatment that is exogenous, to estimate the causal effect of an additional semester of supplemental reading on the student outcome, Y_{ijgs} . As in the first stage, I also cluster standard errors at levels of the forcing variable.

The coefficient that answers my research question is θ_2 , which represents the causal effect of experiencing an additional semester of the literacy intervention for a student who fell just shy of the required passing score on the fifth-grade ITBS compared to students who scored just above this score threshold on the fifth-grade test.

Following the model of Imbens and Kalyanaraman (2012), I model the relationship between fifth-grade ITBS score and the outcome in each stage as “locally linear.” I chose an optimal bandwidth using the Imbens and Kalyanaraman method (IK) and fit models across analytic windows of varying width. The results of this process yield different optimal bandwidths for each outcome. To verify the robustness of my results, I fit my models using the IK and additional choices of bandwidth. Using wider bandwidths allows me to increase my statistical power and precision. In the subsequent Threats to Validity section, and throughout my analyses, I perform a number of tests to verify that my results are not sensitive to my choice of bandwidth or the functional form of my forcing (running) variable, $ITBS5$.

Verifying Assumptions for Regression-Discontinuity

All regression discontinuities have the potential to be undermined by failures of important assumptions, most notably discontinuities in other variables, or discontinuities in the forcing variables at unexpected locations. It is necessary to establish that these findings are not driven by discontinuities in control

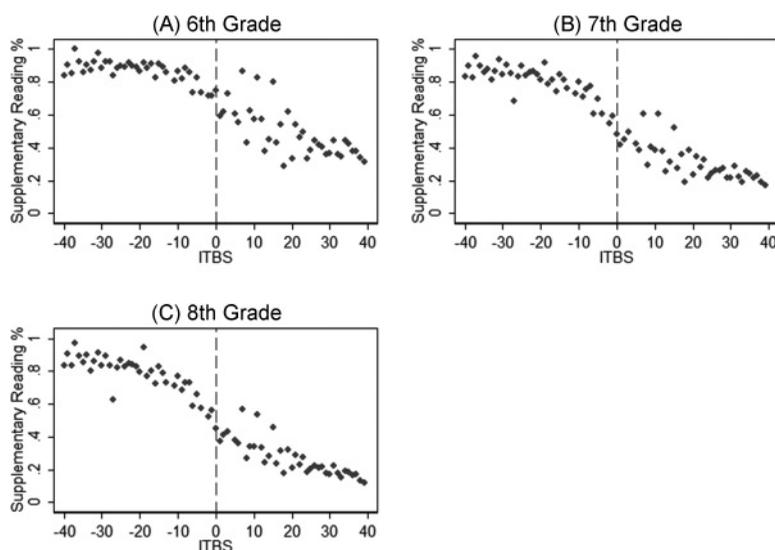


Figure 2. Evidence of Discontinuity in Treatment Exposure in All Three Grades for All Students

variables, that the instrument is working the way it was intended, and the only discontinuity in student exposure to the treatment is at the point of the cutoff designated by the school district.

To verify the soundness of my regression discontinuity design, I use several checks on my model. Following the example of McCrary (2008), I first investigated whether any evidence existed to suggest manipulation of the forcing variable. Manipulation of a student's position relative to the district-defined cutoff is highly implausible. For instance, students cannot manipulate their position relative to the cutoff because the percentile rank is generated from a nationally normed sample. And district administrators may likewise not manipulate the eligibility of students with respect to the exogenously chosen cutoff, which lessens the potential threat to the RD design. Despite the absence of a real threat to the validity of my forcing variable, in figure 1 I present the empirical distribution of the forcing variable (fifth-grade ITBS score) to illustrate that it is smooth across the whole distribution, and in particular around the discontinuity used for assigning students to the supplementary reading (denoted by the vertical dashed line). The empirical distribution that I present in figure 1 does not show evidence of particularly high densities of individuals on either side of the cutoff, which might suggest evidence of manipulation.

To further attest to the validity of the RD approach, I display in figure 2 evidence of a discontinuity in exposure to treatment at the exogenously determined cutoff in ITBS score. Each of the three panels depicts evidence of

a modest discontinuity in the number of semesters of the reading support that students on either side of the eligibility cutoff received by the specified grade in middle school. For instance, the gap shown between the trends at the cutoff score in panel A suggests that exposure to treatment by the end of sixth grade differed by about 20 percentage points, or about half of a semester. Though the discontinuity is not large, it is pronounced enough to warrant the use of a fuzzy regression-discontinuity design. I demonstrate in my first-stage analysis that, despite this small discontinuity, the instrument is quite strong by conventional measures of instrument strength (Stock, Wright, and Yogo 2002) in both the seventh and eighth grades.

As a final check on the appropriateness of my RD approach, I examined the distributions of covariates that I used as control variables to ensure that no other discontinuities existed that might have generated my results. To examine potential discontinuities, I fit the model:

$$W_{ij} = \alpha_0 + \alpha_1 CITBS_{ij} + \alpha_2 ELIG_{ij} + \alpha_3 CITBS \times ELIG_{ij} + \delta_{ij}. \quad (3)$$

I fit this model across multiple bandwidths to confirm that, near the cutoff, there are no discontinuities in the covariates and that students who are eligible for assignment to supplementary reading are equal in expectation to those who are not eligible, based on their fifth-grade ITBS score. Evidence that this assumption is upheld is demonstrated by my failing to reject the null hypothesis that α_2 is equal to 0 for each of the covariates. I display the results of this specification check in table 1. These results suggest that, except for the IK bandwidth, which is just two percentile points on either side of the cutoff, there are just three tests that reject the null hypothesis of no difference in the covariates in the treatment or control groups. The difference in the share of Latino students on either side of the cutoff appears that it is likely aberrant, although there is some evidence the share of students with disabilities may differ slightly on either side of the cutoff. In my preferred models, subsequently, I include controls for these student covariates as a way to adjust for these small differences in student characteristics.

4. RESULTS

I find evidence that the reading intervention positively impacts the performance of white, Latino, and Asian students on state tests of reading and mathematics performance, as well as the eighth-grade ITBS. However, I find larger and negative impacts on literacy measures for black students. The effects I estimate are noisy in sixth grade, but quite precise in both seventh and eighth grades.

Table 1. Estimated Differences in Covariate Balance at the Cutoff for Eligibility for Supplementary Reading

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Female	Black	Asian	Latino	Other Race	ELL	FRPL	SWD	5th Grade Math	5th Grade Reading
IK bandwidth	-0.101 (0.094)	0.046 (0.055)	0.129*** (0.024)	-0.123*** (0.000)	0.053*** (0.007)	0.090*** (0.000)	0.153** (0.053)	0.018** (0.006)	-0.102 (0.083)	0 (0.001)
N	804	804	804	804	804	440	804	804	867	802
Bandwidth = 5	-0.022 (0.077)	0.088 (0.059)	0.026 (0.027)	-0.043** (0.017)	0.014 (0.013)	0.016 (0.020)	0.073 (0.056)	-0.023 (0.022)	-0.092 (0.071)	-0.035 (0.020)
N	1,160	1,160	1,160	1,160	1,160	1,160	1,160	1,160	1,160	1,158
Bandwidth = 10	-0.011 (0.029)	0.029 (0.039)	0 (0.024)	-0.013 (0.021)	-0.002 (0.010)	0.006 (0.013)	0.02 (0.035)	-0.055** (0.021)	-0.042 (0.042)	-0.016 (0.011)
N	2,576	2,576	2,576	2,576	2,576	2,576	2,576	2,576	2,576	2,574
Bandwidth = 20	-0.001 (0.024)	0.015 (0.028)	-0.003 (0.020)	-0.004 (0.018)	-0.006 (0.007)	-0.003 (0.011)	0.012 (0.035)	-0.041** (0.019)	-0.014 (0.034)	0.017 (0.019)
N	5,399	5,399	5,399	5,399	5,399	5,399	5,399	5,399	5,399	5,396
Mean at threshold	0.584	0.403	0.013	0.143	0.013	0.104	0.416	0.143	-0.052	-0.056

Notes: Heteroskedasticity robust standard errors clustered by ITBS score are in parentheses. Each row shows estimates for the difference in the group characteristic listed for students just below the cutoff for supplemental reading eligibility when compared with those just above the cutoff. The coefficients shown are generated by local linear regression using an edge kernel with the specified bandwidth.
 FRPL: free and reduced price lunch; SWD: students with disabilities; IK: Imbens and Kalyaraman (2012) method.
 ***Statistically significant at the 0.1% level; ** statistically significant at the 1% level; * statistically significant at the 5% level.

Patterns of Participation

Despite the rule-based determination of eligibility for the supplemental reading program, the ability of students and families to select in and out of the program makes for patterns of participation that require closer examination. In table 2 I present the patterns of participation by grade and race for students who were within 10 percentile points on either side of the cutoff used to determine eligibility. Two features of this table are particularly noteworthy. First, black and white students participate in the supplemental reading program at similar rates, regardless of their formal eligibility. Second, students who are by rule not eligible for the reading program are more likely to elect to take the reading class in sixth grade than they are in grades 7 or 8. For instance, 75 percent of all students spend both of their semesters in sixth grade (100 percent of their enrollment) in supplemental reading if they are eligible and within ten points of the cutoff, while nearly 50 percent of those who are above the cutoff and within ten percentile points also choose to participate for the whole of their sixth-grade year. By seventh grade, only about half of those initially eligible for the reading course still participate fully (which is two-thirds of the sixth-grade participants), while those technically ineligible for the program participate at much lower rates. The same pattern holds in eighth grade, suggesting the supplementary reading course is widely used in sixth grade but used more selectively and more in line with the policy in subsequent years. As I detail in the following discussion, the result of these patterns of participation is that there is not much difference in the overall rates of participation in sixth grade as a function of formal eligibility.

Patterns of participation in the supplemental reading program were jointly determined by families as well as district officials. Students and families had flexibility to override the policy and enroll or not enroll in supplemental reading as they saw fit. Likewise, the district intentionally reevaluated the need for students to persist in the supplemental reading class based on their middle school performance, though there was not a formulaic way in which this was operationalized. To understand whether attrition from the treatment from sixth to seventh grade was related to a student's end-of-year reading test results, I fit a model that used treatment exposure in seventh and eighth grades, respectively, as the outcomes, as a function of prior participation in supplemental reading, end-of-year reading test scores, demographic characteristics, and fixed effects for school and cohort. I find that a positive 1 SD difference on the sixth-grade state reading examination is associated with a 5 percentage point reduction in the total amount of time a student spends in supplemental reading in seventh grade, with a slightly larger reduction in the probability for black students. Analogously, I find that a positive 1 SD difference in seventh-grade reading test performance is associated with a 1 percentage point difference in the total

Table 2. Share of Total Enrolled Semesters Spent in Supplemental Reading, Separate by Race, Across Grades in Middle School for Those Eligible and Not Eligible as Defined by the Cutoff Rule

Panel B: Black Students												
Panel A: White, Latino, and Asian Students						Panel B: Black Students						
6th Grade		7th Grade		8th Grade		6th Grade		7th Grade		8th Grade		
Enrollment Share (%)	Eligible (%)	Not (%)	Eligible (%)	Not (%)	Eligible (%)	Not (%)	Eligible (%)	Not (%)	Eligible (%)	Not (%)	Eligible (%)	Not (%)
0	24.0	43.4	22.2	42.1	20.3	40.6	19.5	32.4	16.7	30.2	15.3	28.8
17	—	—	—	—	0.7	1.0	—	—	—	—	2.5	3.7
25	—	—	0.8	0.8	—	—	—	—	2.7	3.4	—	—
33	—	—	—	—	23.5	31.7	—	—	—	—	22.5	27.3
50	1.5	0.8	26.2	37.5	1.7	1.6	4.2	5.4	29.5	33.9	3.2	3.9
67	—	—	—	—	5.7	7.0	—	—	—	—	8.1	8.3
75	—	—	3.4	1.7	—	—	—	—	5.3	8.5	—	—
83	—	—	—	—	3.0	1.5	—	—	—	—	4.7	7.1
100	74.5	55.8	47.4	17.9	45.0	16.6	76.3	62.2	45.8	23.9	43.6	21.0
Total <i>n</i>	757	881	757	881	757	881	528	410	528	410	528	410

Notes: Descriptive statistics are for students in the analytic sample and within 9 percentile points of the 60th percentile on the fifth-grade ITBS used to determine eligibility for supplementary reading in middle school. All percentages are total semesters enrolled in the intervention as compared with total semesters enrolled in middle school by the end of the stated grade.

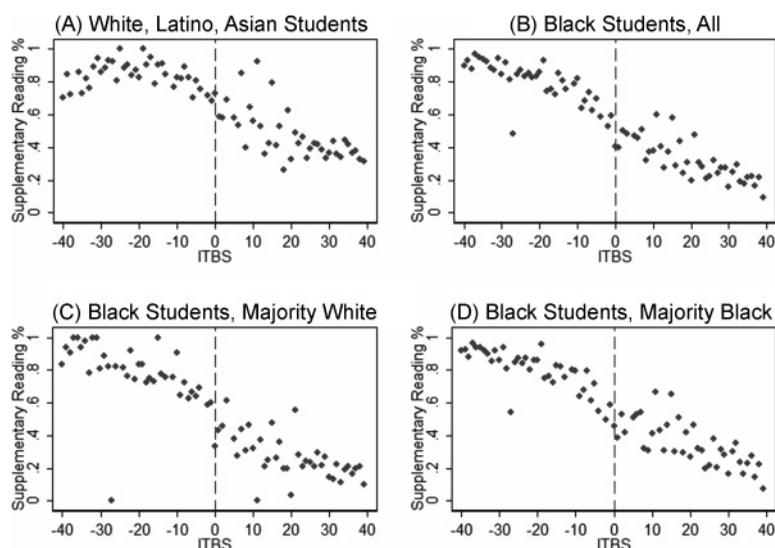


Figure 3. Evidence of Heterogeneity in the First Stage Exposure to the Treatment in Seventh Grade, by Race and Racial Makeup of Middle School

participation in supplemental reading in eighth grade. I interpret these findings to suggest that attrition from the treatment was to some degree a function of test performance but other factors must account for this selective attrition from treatment, particularly in the first year. Though this selection out of treatment is potentially endogenous, using a continuous measure, rather than a binary measure, of treatment exposure gives a better estimate of policy impact and may be less susceptible to violations of the exclusion restriction (Angrist and Imbens 1995).³

Reduced-Form Results

In table 3 I present my reduced-form estimates of the effect of supplementary reading exposure on subsequent student test scores fit using my entire sample of students. To fit the reduced-form models, I regressed the outcome on the forcing variable, fifth-grade ITBS scores—the measure of exposure to supplementary reading and fixed-effects for cohort and school. I include the reduced-form estimates for my outcomes of interest, standardized reading and mathematics scores in sixth through eighth grades, as well as the scaled score on the eighth-grade ITBS. In the rows of table 3 I report my reduced-form estimates across several choices of bandwidth. Most specifications do not include demographic control variables, though I add them for the bandwidth

3. I find similar results in terms of sign and statistical significance when using a binary measure of exposure to treatment, although I do not present those results here given that they are less policy-relevant.

Table 3. Reduced Form Estimates

	(1) ELA Score Grade 6	(2) ELA Score Grade 7	(3) ELA Score Grade 8	(4) ITBS Score Grade 8	(5) Math Score Grade 6	(6) Math Score Grade 7	(7) Math Score Grade 8
IK bandwidth	-0.075 (0.080)	-0.029 (0.083)	-0.231*** (0.038)	-0.607 (0.690)	-0.093 (0.058)	-0.177*** (0.007)	-0.062*** (0.006)
N	804	867	804	1,979	867	867	867
Bandwidth = 5	0.069 (0.069)	0.099 (0.072)	-0.160*** (0.040)	-0.137 (1.200)	-0.004 (0.056)	-0.04 (0.031)	-0.009 (0.024)
N	1,160	1,160	1,160	1,160	1,160	1,160	1,160
Bandwidth = 10	0.011 (0.045)	0.028 (0.044)	-0.062 (0.039)	-0.289 (0.671)	0.006 (0.032)	0.015 (0.042)	-0.005 (0.022)
N	2,576	2,576	2,576	2,576	2,576	2,576	2,576
Bandwidth = 10, controls	0.037 (0.047)	0.047 (0.056)	-0.042 (0.037)	-0.003 (1.092)	0.054** (0.021)	0.057 (0.038)	0.051* (0.029)
N	2,576	2,576	2,576	2,576	2,576	2,576	2,576
Bandwidth = 20	0.039 (0.033)	0.035 (0.030)	-0.018 (0.042)	0.054 (0.439)	0.017 (0.024)	0.034 (0.032)	0.006 (0.020)
N	5,399	5,399	5,399	5,399	5,399	5,399	5,399

Notes: Heteroskedasticity robust standard errors clustered by ITBS score are in parentheses. Each row shows reduced-form estimates of the impact of eligibility for supplementary reading on the outcome listed in a particular column. The coefficients shown are generated by local linear regression using an edge kernel with the listed bandwidth. No additional controls are included except as indicated. The models with controls condition on gender, race, LEP and disability status, prior math scores, cohort, and middle school fixed effects.

ELA: English language arts; IK: Imbens and Kalyaramanan (2012) method.

***Statistically significant at the 0.1% level; **statistically significant at the 1% level; *statistically significant at the 5% level.

of 10 which is my preferred specification. My reduced-form estimates suggest that there are small positive effects of being just eligible to participate in supplemental reading in mathematics test scores compared with those who just miss being eligible (figures 4 and 5). Specifically, students who are just eligible score, on average, about 0.05 of a standard deviation higher on subsequent tests of their mathematics skills.

The demographic makeup of Hampton County schools is such that black and white students constitute similarly large portions (42 and 33 percent, respectively) of the total student population. To explore the possible differences in the impact of eligibility for the reading course by race, I introduce table 4. In table 4 I display my estimates of the impact of eligibility on two groups of students: white, Latino, and Asian students, and then separately for black students. I also go one step further to estimate whether the effects on black students differ based on whether they are in schools with more than 50 percent

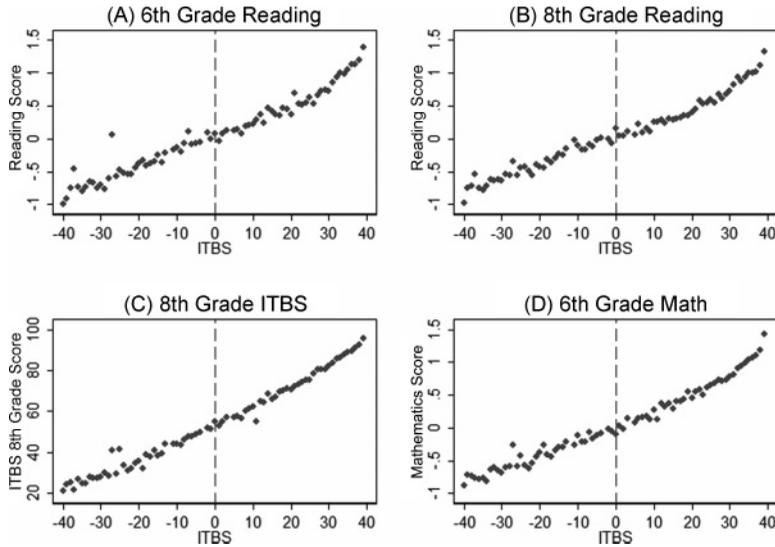


Figure 4. Evidence of Reduced-Form Impact of Eligibility on Four Selected Outcomes

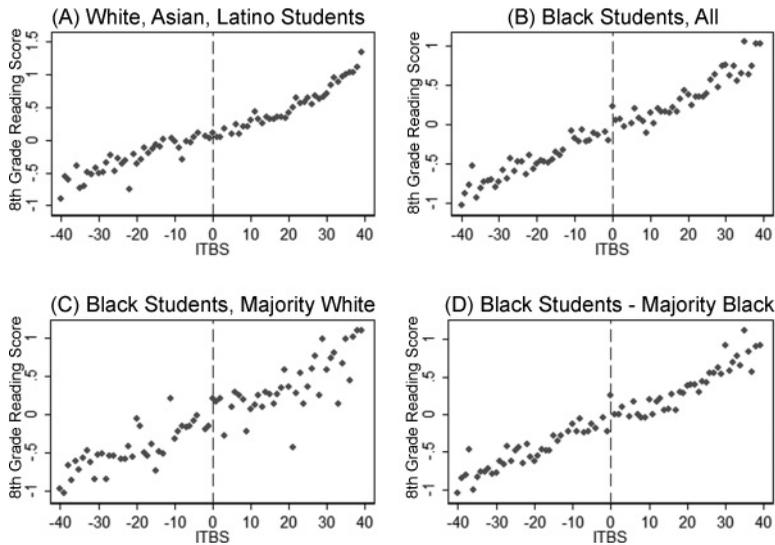


Figure 5. Reduced-Form Impact on Eighth-Grade Reading Scores by Race

enrollment of black students (a coarse measure of racial segregation). Exploring these disaggregated results exposes important and policy-relevant heterogeneity in the impact of this literacy intervention.

In table 4 each row corresponds to one of the subgroups for which I have estimated the reduced-form effects. My estimates suggest that scoring just below the cutoff for eligibility for the supplementary reading program has a negative impact on all measures of subsequent reading performance for black

Table 4. Reduced Form Heterogeneity by Race and School Racial Composition

	(1) ELA Score Grade 6	(2) ELA Score Grade 7	(3) ELA Score Grade 8	(4) ITBS Score Grade 8	(5) Math Score Grade 6	(6) Math Score Grade 7	(7) Math Score Grade 8
All students	0.037 (0.047)	0.047 (0.056)	-0.042 (0.037)	-0.003 (1.092)	0.054** (0.021)	0.057 (0.038)	0.051* (0.029)
N	2,576	2,576	2,576	2,576	2,576	2,576	2,576
Black students	-0.058 (0.043)	-0.143** (0.065)	-0.204*** (0.031)	-2.729*** (0.726)	-0.068 (0.048)	-0.031 (0.030)	-0.048 (0.077)
N	938	938	938	938	938	938	938
White, Latino, & Asian students	0.100* (0.055)	0.154** (0.071)	0.06 (0.041)	1.675 (1.308)	0.114*** (0.033)	0.115** (0.048)	0.102*** (0.030)
N	1,638	1,638	1,638	1,638	1,638	1,638	1,638
Black students, majority white	-0.14 (0.111)	-0.214** (0.100)	-0.179** (0.064)	-3.100* (1.530)	0.014 (0.062)	0.052 (0.094)	0.185** (0.064)
N	283	283	283	283	283	283	283
Black students, majority black	-0.013 (0.042)	-0.1 (0.060)	-0.185*** (0.030)	-2.427* (1.159)	-0.117* (0.055)	-0.052 (0.031)	-0.14 (0.088)
N	655	655	655	655	655	655	655

Notes: Heteroskedasticity robust standard errors clustered by ITBS score are in parentheses. Each row shows reduced-form estimates of the impact of eligibility for supplementary reading on the outcome listed in a particular column. The coefficients shown are generated by local linear regression using an edge kernel and a bandwidth of 10. All models include controls for gender, race, LEP and disability status, prior math scores, cohort, and middle school fixed effects.

ELA: English language arts.

***Statistically significant at the 0.1% level; **statistically significant at the 1% level; *statistically significant at the 5% level.

students relative to students who just missed being eligible. For instance, I interpret the statistically significant coefficient of -0.143 in row 4 of column 2 to mean that, on average, black students who are just eligible to participate in supplemental reading score about 0.14 SD lower on the seventh-grade reading assessment than similar students who just missed being eligible. This negative effect is comparable to moving from the 60th percentile of the national distribution of reading performance to the 54th percentile. The intent-to-treat (ITT) impact on the eighth-grade ITBS score is equivalent to moving only from the 60th to the 58th percentile. Conversely, on average, white, Latino, and Asian students who were just eligible for supplemental reading appear to perform better relative to those who just missed being eligible on subsequent measures of reading and mathematics, though not all estimates are statistically different from zero. Interestingly, the negative impact that I estimate for black students who are just eligible for supplemental reading suggests these students

are negatively affected whether they are in majority-black or majority-white schools, though the negative impacts appear to be larger and more precisely estimated for students in schools that enroll a majority of white students. Importantly, in schools that enroll a majority of black students there are not many students who score above the 60th percentile on the ITBS. As a result, these estimates are particularly imprecise for black students in schools that are majority black since there is less variation in the eligibility indicator, despite the larger sample size relative to majority white schools.

My reduced-form estimates constitute the ITT estimates of the offer of eligibility for supplementary reading. If take-up of the supplementary reading program was perfectly predicted by eligibility for the program these estimates would be the estimates of greatest policy interest, since they apply to the whole distribution of reading ability. However, because take-up of the treatment, conditional on eligibility, is not perfect, I contrast my ITT estimates with the instrumental variable estimates from my subsequent regression-discontinuity analysis, and emphasize in the Discussion the implications this has for research and practice. These instrumental variable (IV) estimates constitute the treatment-on-the-treated (TOT) effects of supplementary reading, for students who are just eligible for supplemental reading and who participate, relative to those who just miss being eligible and who do not participate. These TOT effects are of arguably larger importance in answering the question of whether those who experienced the treatment actually benefited from it.

First-Stage Results

Both theory and the results of my reduced-form analyses suggest the impact of this literacy intervention likely differs by student race, particularly given the large white and black enrollments in Hampton County. In estimating my first-stage results I model the effect of the eligibility for supplemental reading on take-up for all students as well as the same racial and enrollment subgroups that I articulate earlier. In figure 2 I show the impact of eligibility on take-up for all students in each of their three years of enrollment in middle schools. In panel A I illustrate differences in take-up for all students in sixth grade, and in panels B and C I illustrate the discontinuity in take-up as a function of eligibility in grades 7 and 8, respectively. In all three grades the difference in take-up as a function of the offer of eligibility for students on the margin of receiving that offer is not very large.

I corroborate this graphical evidence with fitted models (following equation 1) in table 5. The columns in table 5 represent the three years of middle-school enrollment and the rows include estimates of the effect of eligibility on enrollment for different choices of bandwidth. Except where noted explicitly,

Table 5. First Stage Estimates by Grade of Enrollment

	(1) Grade 6	(2) Grade 7	(3) Grade 8
IK bandwidth	0.005 (0.043)	0.190*** (0.031)	0.183*** (0.033)
<i>F</i>	0.013	36.404	30.53
<i>N</i>	804	804	804
Bandwidth = 5	0.101 (0.064)	0.174*** (0.034)	0.179*** (0.036)
<i>F</i>	2.549	25.575	24.187
<i>N</i>	1160	1160	1160
Bandwidth = 10	0.046 (0.034)	0.101*** (0.032)	0.116*** (0.030)
<i>F</i>	1.838	10.133	14.88
<i>N</i>	2576	2576	2576
Bandwidth = 10, controls	0.044 (0.026)	0.101*** (0.023)	0.116*** (0.020)
<i>F</i>	2.927	19.651	32.791
<i>N</i>	2576	2576	2576
Bandwidth = 20	0.058** (0.026)	0.105*** (0.023)	0.117*** (0.022)
<i>F</i>	4.974	20.929	29.338
<i>N</i>	5399	5399	5399

Notes: Heteroskedasticity robust standard errors clustered by ITBS score are in parentheses. Each row shows first stage estimates of the impact of eligibility for supplementary reading on the probability of enrollment in the reading course. The coefficients shown are generated by local linear regression using an edge kernel with the listed bandwidth. No additional controls are included except in the final row, which conditions on gender, race, LEP and disability status, prior math scores, cohort, and middle school fixed effects. Below each coefficient is the *F*-statistic associated with the excluded instrument.

***Statistically significant at the 0.1% level; **statistically significant at the 1% level; *statistically significant at the 5% level.

IK: Imbens and Kalyaramanan (2012) method.

these estimates are from models specified without control variables. My results suggest that my instrument is weak in sixth grade, but relatively strong in both seventh and eighth grades. In my preferred specification of a bandwidth of 10 percentile points, the *F* statistic on my instrument in sixth grade is just less than 3, compared with a value of 10, which is commonly regarded as the minimum threshold for a single instrument to be sufficiently strong (Stock, Wright, and Yogo 2002). In seventh and eighth grades, however, my instrument is quite strong, with *F* statistics that exceed the threshold of 10. In

fact, in seventh and eighth grades, students who are just eligible to participate in supplemental reading in middle school spend between 10 and 15 percent more of their total semesters of middle school in supplemental reading than their peers who just miss being eligible. These results are robust to choice of bandwidth.

To explore heterogeneity in take-up by race I first turn to figure 3 where I illustrate differences in take-up in seventh grade only. In panel A of figure 3, I show the first-stage graphically for white, Latino, and Asian students; panel B demonstrates take-up for black students regardless of school enrollment; and panels C and D examine take-up by black students in majority-white and majority-black schools, respectively. These four panels illustrate that take-up of the intervention is stronger for white students than for black students, but that this is largely an artifact of differences between take-up among black students enrolled in majority-black versus majority-white schools. In fact, black students on the margin of eligibility in majority-white schools actually take up treatment at higher rates than their peers who just miss being eligible.

As with my aggregate first-stage analysis, I also fit equation 1 for my racial subgroups. In table 6 I present my estimates of the impact of being just eligible for supplemental reading on subsequent take-up by grade in middle school. As is the case in my aggregate analyses, my first-stage instrument is weak in sixth grade but considerably stronger in grades 7 and 8. Importantly, my first-stage estimates for black students are weaker in part because the instrument appears to function differently in schools with different racial composition. Specifically, schools with a majority of black students are generally less responsive to the instrument, regardless of grade, whereas black students enrolled in schools with a majority of white students are much more sensitive. Consequently, in my estimates of the treatment on the treated, I focus on the effects of participating in supplemental reading by race, but do not continue to subdivide the analyses by racial composition of the school.

TOT Estimates of Supplementary Reading

I find that spending a greater percentage of middle-school time in supplemental reading has differential effects by grade and by racial subgroup. Specifically, although I find no overall effect on students, I find persistently negative effects of exposure to reading for black students, and only imprecise and suggestive evidence of positive effects for white, Latino, and Asian students. As expected, the weak instrument in sixth grade yields only very noisy estimates of the effect of participating in supplementary reading on sixth-grade test scores, although the strength of the instrument in seventh and eighth grades does allow for much better precision.

Table 6. First Stage Estimates by Grade, Including Heterogeneity by Race and Racial Composition

	(1) Grade 6	(2) Grade 7	(3) Grade 8
All students	0.044 (0.026)	0.101*** (0.023)	0.116*** (0.020)
<i>F</i>	2.927	19.651	32.791
<i>N</i>	2576	2576	2576
White, Latino, & Asian students	0.057 (0.040)	0.122*** (0.035)	0.132*** (0.033)
<i>F</i>	2.099	12.158	15.964
<i>N</i>	1638	1638	1638
Black students	0.018 (0.030)	0.063* (0.034)	0.088* (0.042)
<i>F</i>	0.362	3.5	4.443
<i>N</i>	938	938	938
Black students, majority white	0.216*** (0.069)	0.237*** (0.041)	0.238*** (0.037)
<i>F</i>	9.735	33.208	41.1
<i>N</i>	283	283	283
Black students, majority black	-0.041 (0.038)	-0.001 (0.040)	0.035 (0.051)
<i>F</i>	1.141	0	0.478
<i>N</i>	655	655	655

Notes: Heteroskedasticity robust standard errors clustered by ITBS score are in parentheses. Each row shows first stage estimates of the impact of eligibility for supplementary reading on the probability of enrollment in the reading course. The coefficients shown are generated by local linear regression using an edge kernel with the listed bandwidth. Additional controls are included and condition on gender, race, LEP and disability status, prior math scores, cohort, and middle school fixed effects. Below each coefficient is the *F*-statistic associated with the excluded instrument. *** Statistically significant at the 0.1% level; ** statistically significant at the 1% level; * statistically significant at the 5% level.

In column 2 of table 7 the coefficient of -2.277 in the fourth row suggests that for black seventh graders, participating in supplemental reading for 100 percent of their semesters (or four by the end of seventh grade) decreases their seventh-grade state reading test scores by over 2 SDs. The magnitude of this IV estimate may be deceptively large for two reasons. First, although I have a strong instrument, the discontinuity in participation in supplementary reading for students on the margin of eligibility in seventh grade is only about 10 percentage points. This means that the students who are just eligible to participate in supplemental reading only spend about 10 percentage points more time in the course by the end of seventh grade than their peers who

Table 7. Instrumental Variables Heterogeneity by Race and Gender

	(1) ELA Score Grade 6	(2) ELA Score Grade 7	(3) ELA Score Grade 8	(4) ITBS Score Grade 8	(5) Math Score Grade 6	(6) Math Score Grade 7	(7) Math Score Grade 8
All students	0.833 (1.444)	0.471 (0.598)	-0.365 (0.254)	-0.024 (9.071)	1.234 (0.997)	0.567 (0.473)	0.44 (0.282)
N	2,576	2,576	2,576	2,576	2,576	2,576	2,576
Black students	-3.183 (5.282)	-2.277* (1.227)	-2.331** (1.090)	-31.164 (19.271)	-3.731 (7.399)	-0.491 (0.520)	-0.553 (1.056)
N	938	938	938	938	938	938	938
White, Latino, Asian students	1.736 (2.019)	1.265 (0.796)	0.451 (0.365)	12.679 (12.125)	1.988 (1.291)	0.947 (0.623)	0.776** (0.332)
N	1,638	1,638	1,638	1,638	1,638	1,638	1,638

Notes: Heteroskedasticity robust standard errors clustered by ITBS score are in parentheses. Each row shows instrumental variables estimates of the impact of advanced coursework on end of grade math scores. The coefficients shown are generated by local linear regression using an edge kernel and a bandwidth of 10. All models include controls for gender, race, LEP and disability status, prior math scores, cohort, and middle school fixed effects.

***Statistically significant at the 0.1% level; **statistically significant at the 1% level; *statistically significant at the 5% level.

ELA: English language arts.

just missed being eligible. Thinking about this in reference to table 2 means that whereas those who are not technically eligible spent about 50 percent of their time enrolled in supplemental reading, those who were just eligible spent about 60 percent of their enrolled time in the class. The second reason my IV estimates are so large is a function of how I defined treatment. The IV estimate represents the effect of a one-unit change in enrollment on subsequent test scores, but my enrollment variable is measured as a percentage from 0 to 1. As a result, to estimate the effect of a difference in participation of 10 percentage points I have to scale my IV estimate down by dividing by ten. The results are still impressive. A black student who spends 10 percentage points more time in supplemental reading is likely to score 0.23 SD lower on his seventh-grade reading test than a similar student who just missed being eligible and did not participate.

My estimates of the negative impact of participating in supplemental reading for black students extend to eighth-grade reading test outcomes as well, though they appear not to affect mathematics outcomes. The negative effect of participating in supplemental reading on the eighth-grade state reading test is similar to that in seventh grade, while the ITBS reading score for black students is marginally significant ($p = 0.105$) and suggests students who participate fully in the intervention through eighth grade score three-fifths of a

standard deviation lower on the ITBS in eighth grade than similar students who do not participate at all. This is tantamount to moving from the 60th national percentile to the 37th percentile. My estimates for white, Latino, and Asian students (as a single group) are suggestive and border on marginal significance. Like my estimates of the effects of participation on black students, my point estimates are consistently signed and of large enough magnitude to be noteworthy if I could achieve better statistical precision.

5. DISCUSSION

Threats to Validity

There are several potential threats to the validity of my findings, some of which are methodological, and others that are related to program implementation. The chief methodological threats to the validity of my findings are that they may be sensitive to my choice of bandwidth, and a linear specification of the relationship between the forcing variable and my outcomes may not be appropriate. My analyses could also be threatened by attrition from the treatment group over time. I consider each of these threats below, beginning with the methodological threats.

In table 8, I display the results of fitting reduced-form models in my preferred bandwidth of 10 percentile points using several nonlinear specifications of the forcing variable. The rows are organized according to the maximum degree of the polynomial specification of the forcing variable, and the columns are organized by outcome by racial subgroup. For instance, the first three columns pertain to the multiple specifications for three outcomes for black students, and columns 4–6 contain the analogous estimates for white, Latino, and Asian students. My point estimates suggest that my results are robust to nonlinear specification regardless of racial group or outcome. Though my point estimates fluctuate somewhat, the sign and significance of the estimates is generally preserved.

A substantive threat to the validity of my findings concerns the attrition of students from the pool of students initially tested in fifth grade across the effective treatment and control groups. Among the 6,219 students who fall within my analytic window and who were tested on the ITBS in fifth grade, 4,758 remain in my final analytic sample of students who are tested in each of their sixth-, seventh-, and eighth-grade years in middle school. Of the roughly 1,500 students who leave the sample, these represent 25 percent of those not eligible for supplementary reading and 21 percent of those who were eligible based on their fifth-grade test score. In this instance, the rate of attrition is somewhat higher among the control group. Although differential attrition may be a potential source of bias, the fact that the covariate balance on either side

Table 8. Testing Robustness of Reduced Form Results to Nonlinear Specifications of the Forcing Variable

	(1)	(2)	(3)	(4)	(5)	(6)
	Black Students			White, Latino, Asian Students		
	ELA Score Grade 6	ELA Score Grade 7	ELA Score Grade 8	ELA Score Grade 6	ELA Score Grade 7	ELA Score Grade 8
Linear	-0.143** (0.065)	-0.204*** (0.031)	-2.729*** (0.726)	0.154** (0.071)	0.06 (0.041)	1.675 (1.308)
N	938	938	938	1,638	1,638	1,638
Quadratic	-0.190* (0.107)	-0.252*** (0.035)	-2.362** (0.916)	0.261*** (0.072)	-0.057 (0.035)	1.424 (1.854)
N	938	938	938	1,638	1,638	1,638
Cubic	-0.252 (0.233)	-0.363*** (0.062)	-0.617 (1.735)	0.381*** (0.126)	-0.233*** (0.039)	3.443 (3.227)
N	938	938	938	1,638	1,638	1,638
Quartic	-0.916** (0.326)	-0.358** (0.136)	-6.424 (3.945)	0.420** (0.168)	-0.17 (0.122)	-1.923 (4.881)
N	938	938	938	1,638	1,638	1,638

Notes: Heteroskedasticity robust standard errors clustered by ITBS score are in parentheses. Each row shows reduced-form estimates of the impact of eligibility for supplementary reading on the outcome listed in a particular column. The coefficients shown are generated by local linear regression using an edge kernel with a bandwidth of ten. Controls are included for gender, race, LEP and disability status, prior math scores, cohort, and middle school fixed effects. The forcing variable is specified to include up to the polynomial listed in the associated row.

*** Statistically significant at the 0.1% level; ** statistically significant at the 1% level; * statistically significant at the 5% level.

ELA: English language arts.

of the eligibility cutoff is strong suggests this differential attrition likely does not have a meaningful impact on my analyses.

I analyze my sample attrition in a manner consistent with the evidence standards for regression-discontinuity designs established by the What Works Clearinghouse (WWC) (Schochet et al. 2010). My analyses focus on both analytical samples; those cohorts of students who are observed between fifth and eighth grades and those observed in fifth and sixth grades. For the purposes of analyzing attrition, I define the students who scored within ten points of the cutoff score on the ITBS-Reading in fifth grade as the focal group.

Interpreting Findings

My findings have several implications, both clear and suggestive, for how Hampton County and other districts could consider using their supplementary reading programs in the future. The clearest finding in my study is that the impact of the program is strongly negative for black students, and suggestively positive for white, Latino, and Asian students. Of particular interest is

that the reduced-form effects suggest the mere labeling of students as falling below a normative level of performance can impact their subsequent cognitive outcomes. In the following I consider possible interpretations and policy implications for my findings.

For black students, the negative effect of eligibility and participation appears to get larger (but is certainly at least stable) over time, whereas the positive effects for white students appear to fade out from seventh to eighth grades (my points estimates are roughly half the size and no longer significant). Though some of this fadeout may be a statistical artifact of increasing variance in knowledge across years (Cascio and Staiger 2012), there are likely substantive reasons that this fadeout seems apparent. One plausible hypothesis is that, because state reading and mathematics test scores are used for grade promotion in eighth grade, but not sixth or seventh grade, students in the treatment group may work harder to generate higher scores under these high-stakes conditions. The incentives should not differ for black students and all others, and later in this paper I explore other hypotheses for why these differences may persist in eighth grade. One important message from my results is that the divergence between black students and their white, Latino, and Asian counterparts is consistent across the three years of assessments. The recent work by Cascio and Staiger (2012) suggests that even if the magnitude of differences in achievement by race appears identical these differences are likely larger at higher grade levels.

One reason I might observe negative policy impacts of a literacy intervention is related to receiving a negative label as the result of falling below the specified cutpoint. The potential negative effect of receiving a negative label or signal from a test score is certainly consistent with work by Papay, Murnane, and Willett (2010), who found that just failing a low-stakes state standardized test in eighth grade was associated with a lower probability of investing in post-secondary education relative to students who had just passed that test. In this case all students who score below the ITBS cutoff in fifth grade receive this negative label, but only black students appear to be negatively affected. This fact is also consistent with Papay, Murnane, and Willett's (2010) findings in that they found the receipt of the negative label affected lower-income students to a larger degree than higher-income students. These findings are consistent with earlier work that students' perceptions of their own ability can impact their later academic performance (Shen and Pedulla 2000).

The negative shock to students' self-perception by being just eligible may not be sufficient to explain the differential effects of supplemental reading eligibility and participation by racial groups. In fact, the differential and negative impact on black students may be the result of stereotype threat (Aronson and Steele 2005). Many studies have established and replicated that being primed

negatively with information consistent with negative group stereotypes can lead to lower performance. In this setting, it may be that scoring below the 60th percentile in fifth grade leads to the invitation to participate in supplemental reading in middle school, which may itself be a form of negative priming.

Another plausible hypothesis for my observed effects is racial-homogenous groupings in supplementary reading classes could also produce differential impact by race. We might expect that if supplemental reading classes are grouped homogeneously by race, or if black students are overrepresented among the students in these classes, the same sort of negative priming may be possible. Using the data available to me (I am not able to look at specific classroom characteristics), within a school the racial composition of students who enroll in supplemental reading appears to be consistent with the overall racial composition of that school. This fact, coupled with the knowledge from table 2 that black students participate in supplemental reading at rates similar to those among white, Latino, and Asian students, suggests it is unlikely the negative impact of the supplemental reading course is coming through racially segregated classes.

The negative impact on black students is of potential concern, not only related to the immediate outcomes, but also with respect to longer-term outcomes. Work related to creating early-warning indicator systems to reduce school dropout have found that poor performance on middle-grades test scores may not increase the risk of school dropout (Balfanz and Boccanfuso 2007; Balfanz, Herzog, and Mac Iver 2007). However, this same work has found that failing a mathematics or English language arts course in sixth grade is a strong predictor of failing to complete high school within five years of starting. I argue that the negative signal received by being just eligible for supplementary reading in Hampton County may be comparable to the negative signal that students receive by failing a course. This assertion is consistent with the findings of Papay, Murnane, and Willett (2010), and it seems reasonable that students could interpret a negative signal that relays a particular message—falling below some established bar of performance—in a way that would not be conveyed through a continuous test score. The strong negative impact of supplemental reading eligibility and participation for black students on their eighth-grade ITBS reading scores, coupled with the fact that the eighth-grade state reading test is linked to grade retention, suggests there could be additional negative impacts for black students.

Any concern these apparent effects might raise with regard to whether they are limited only to the state-required accountability tests is allayed by the impacts evident on the ITBS reading test in eighth grade. Although the possibility of coaching to tests or “score inflation” are phenomena noted in

other scholarly work (Hamilton and Koretz 2002; Jennings and Bearak 2010; Koretz 2003, 2005), the consistency of my findings (both positive and negative effects) across test types is evidence against a hypothesis that the results are not indicative of learning (or learning loss) in general. In addition, the fact that the positive impact for white, Latino, and Asian students extends to mathematics appears to be consistent with the findings of Cortes, Nomi, and Goodman (2013), who find the effects of additional mathematics instruction are larger for stronger readers. In this case, I hypothesize that the supplemental reading program improved reading skills in a manner that generalized to performance on the state test in mathematics, perhaps through improved skills with respect to open-ended or constructed-response questions.

As with all studies that utilize a regression-discontinuity design, a limitation on the interpretability of my findings is posed by their external validity. By construction, the effects I estimated in my study apply only to those students who were just below the cutoff and eligible for the treatment, in comparison with those who were just above and not eligible. This is a limitation of all studies that use a regression-discontinuity design, but is noteworthy particularly because many policy makers would like to know how literacy interventions can impact the performance of students who are very low performing. In this instance, the margin of analysis is the 60th percentile nationally, although this corresponds to the 40th percentile of performance in the HCPSD. These students may not be the lowest performers, but they are on a margin whose proficiency on tests used for NCLB accountability may be affected and who may be in danger of not graduating from high school. This is particularly true since the state in question uses the state tests in reading and mathematics in eighth grade to make grade promotion decisions. Not being able to generalize away from the cutoff is less concerning given this margin of interest since the district has other programs for addressing the needs of lower performers, and has arguably less need to intervene with higher performers.

Policy Considerations and Alternatives

The results of my study do not make it clear whether alternatives to teaching supplementary reading courses are likely to yield more favorable results. For instance, if the labeling or stereotype threat hypothesis for the differential effects by race is true, it is not clear that using another proven program would yield better results. In addition, a review of the evidence provided through WWC revealed that of the ten experimental or quasi-experimental programs reviewed promoting effective adolescent literacy interventions, six of them are copyrighted or registered trademarks, and one is available through a major educational publishing company (WWC 2012; see also Rouse and Krueger

2004). The availability of off-the-shelf, proven products may be enticing, particularly for schools operating within a tier of the school-improvement cycle, although it is not clear whether the net cost of changing approaches is likely to yield different results. One area worth pursuing further (though it was not possible in my data), is to explore whether students are grouped in supplemental coursework by ability, and whether groups of lower ability are staffed with teachers who have achieved lower value added in the past. Though exploring such mechanisms cannot provide causal evidence, it may provide effective insights for local policy makers and is certainly prudent to explore before undertaking expensive (in resources, time, and political capital) curricular or programmatic reforms.

Student self-perception and feelings of efficacy are quite malleable in middle school (Gillet, Vallerand, and Lafreniere 2012) and it is imperative that schools work to ensure their students don't suffer serious setbacks in this area based on school policies. One option that Hampton County and other districts might consider is making supplemental reading a required course for all students in sixth grade. If this transition from elementary to middle school is particularly crucial (as research suggests), delaying the onset of this reading intervention might diminish the harm of being labeled concurrent with making this transition. Though my hypothesis is speculative, the assertion is testable, and may prove a reasonable experiment on the way to reforming current practice. Sufficient evidence exists to suggest that students' perceptions and experiences with their school characteristics and environment impact their subsequent engagement and academic performance (Wang and Holcombe 2010). As a result, it would be optimal to provide supports for their feelings of efficacy during middle school, particularly if any performance-related identification could negatively impact these feelings.

Districts considering policies similar to the one used in Hampton County should take seriously the limitations on performing an impact evaluation on a discontinuity-based policy that allows for such agency among students and families. Although it may be good political and educational practice to allow for exceptions from the rule, permitting unrestrained exceptions may create asymmetries in the profiles of the compliers. For instance, I find that the instrument for participation in supplemental reading (being eligible) is a strong instrument for black students in majority-white schools, but not so in majority-black schools. Though strong inferences cannot be made from this observation, it is worth exploring (and certainly considering for future policies) whether populations of students or families that may already feel marginalized may be more or less likely to exercise agency in response to school-based policy.

6. CONCLUSION

In an era of high-stakes testing and school accountability, schools care as much now as at any time in the past about improving the literacy skills of their students. For the students, there is nothing more important to their long-term success than their ability to participate in their lives, and the economy, as fully literate individuals and as high-school graduates. My findings suggest that a research-based supplementary reading course in middle school can have differential effects by race on short-term measures of students' reading comprehension, and measures of mathematics knowledge. These findings are concerning in that it appears that black students and their white, Latino, and Asian counterparts received access to similar interventions and yet, even controlling for school of enrollment and graduation cohort, these disparities persisted. Even if the impact of the course is limited to those students who were just below the cutoff used for assigning students to the course, there is still reason to be concerned and to pursue additional understanding of what experiences and processes might be driving these differential results.

My work with Hampton County also underscores the potential value of adopting assignment rules when deciding who to assign to support courses, however it also highlights some associated challenges, particularly around rule compliance (Schlotter, Schwerdt, and Woessmann 2011). Such rules allow for the estimation of causal estimates and can reduce the continuation of ineffective programs. These rules may also help districts modify, develop, or switch to interventions that have been proven to be effective.

Choosing cutoff scores to assign students to academic interventions is not without risk, however, and the determination of whether and where to apply these rules warrants careful consideration. Though there are clear merits to the ability to assess the effectiveness of interventions deployed in a way that allows quasi-experimental evaluation, the potential for rationing of inputs could have deleterious effects. Cutoff scores must be chosen, and interventions designed, in such a way as to be consistent with the needs of the population it is intended to impact. In the HCPSD, all students scoring below the 60th percentile were eligible for supplementary reading, but this evaluation only addresses the impact of those who were eligible but near the cutoff. Other means of evaluation, and perhaps interventions tailored to learners who scored in lower percentiles, are necessary if we are to achieve equitable outcomes in education. It bears further note that simply because a program is impactful on one margin, it need not necessarily maintain its impact when extended to students on other margins and of different abilities.

School and district officials may find it valuable to use these results to impact their own decisions about policy and practice. As schools and districts

make decisions about how to allocate funding for literacy programs, they may find it advantageous to develop and deploy extended instruction time in literacy in ways that resemble the structure used in the HCPSD, paying careful attention to how to avoid the adverse impacts experienced by black students in the HCPSD.

This delivery method is appealing for several reasons, though these results underscore some potential caveats. First, there is a well-defined literature about what practices are effective in literacy instruction, and the HCPSD provides a concrete example of how this may be done. Second, deploying the intervention using district employees allows for flexibility in scheduling teachers and classes within and across schools and school years. Third, the extended learning time approach is likely to have strong face validity among stakeholders in the community. Most groups will find it hard to argue with the idea of using research-proven instruction to supplement traditional curricula as a way to bolster literacy skills. Nevertheless, such policies are all subject to the potential for unintended harm. If students feel stigmatized by being identified as weak readers, or if they receive systematically different services or experiences, this approach to improving educational outcomes can be extremely flawed and, at a minimum, would require additional features. It is difficult to make an ultimate determination of whether the costs to some outweigh the benefits to others when deciding whether to adopt or eliminate a program like the one in Hampton County. District officials would do well to understand the particular needs of their students, and continuously monitor and adjust to the impacts on their students as they become evident.

Further research into literacy interventions like the one in the HCPSD is certainly warranted. Because our ultimate concern is with long-term outcomes that we believe are associated with measures of adolescent literacy, future research should collect data across more years so that we may learn whether there are longer-term impacts on SAT scores, high school graduation, or decisions to apply to or attend college. Establishing the effectiveness of similar supplementary literacy coursework should be pursued in other research contexts as well. While the HCPSD context is representative of many large changing suburban districts, there may be factors associated with the HCPSD that could limit the generalizability of these findings.

I would like to thank Marty West, Larry Katz, Stephen Lipscomb, Matthew Kraft, and John Willet for their insights and feedback. I would also like to thank my anonymous reviewers who helped advance the analysis and findings of this paper. All mistakes or omissions are my own. I would also like to acknowledge the Dean's Summer Research Fellowship at the Harvard Graduate School of Education for material support of this research.

REFERENCES

- Angrist, Joshua, and Guido Imbens. 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90(430):431–442.
- Aronson, Joshua, and Claude Steele. 2005. Stereotypes and the fragility of academic competence, motivation, and self-concept. In *Handbook of competence and motivation*, edited by Andrew J. Elliot and Carol S. Dweck, pp. 436–456. New York: Guilford Press.
- Balfanz, Richard, and C. Boccanfuso. 2007. *Falling off the path to graduation: Middle grade indicators in [an unidentified northeastern city]*. Baltimore, MD: Center for the Social Organization of Schools, Johns Hopkins University.
- Balfanz, Richard, Lisa Herzog, and Douglas Mac Iver. 2007. Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist* 42(4):223–235.
- Card, David, and David Lee. 2008. Regression discontinuity inference with specification error. *Journal of Econometrics* 142(2):655–674.
- Cascio, Elizabeth U., and Douglas O. Staiger. 2012. Knowledge, tests, and fadeout in educational interventions. NBER Working Paper No. 18038.
- Cavanagh, Sean. 2006. Students double-dosing on reading and math. *Education Week* 25(40):1.
- Chall, Jeanne, and Vicki Jacobs. 2003. Poor children's fourth-grade slump. *American Educator* 27(1):14–15.
- Cortes, Kalena, Joshua Goodman, and Takako Nomi. 2012. Doubling up: The long run impacts of remedial algebra on high school graduation and college enrollment. Paper presented at the Association for Education Finance and Policy 37th Annual Conference, Boston, March.
- Cortes, Kalena, Takako Nomi, and Joshua Goodman. 2013. A double dose of algebra. *Education Next* 13(1):70–76.
- Darling-Hammond, Linda. 2004. Standards, accountability, and school reform. *Teachers College Record* 106(6):1047–1085.
- Dole, Janice, Gerald Duffy, Laura Roehler, and David Pearson. 1991. Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research* 61(2):239–264.
- Edmonds, Meaghan, Sharon Vaughn, Jade Wexler, Colleen Reutebuch, Amory Cable, Kathryn Klinger Tackett, and Jennifer Wick Schnakenberg. 2009. A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research* 79(1):262–300.
- Fitzpatrick, Maria, David Grismmer, and Sarah Hastedt. 2011. What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. *Economics of Education Review* 30(2):269–279.

- Gillet, Nicolas, Robert Vallerand, and Marc-André Lafreniere. 2012. Intrinsic and extrinsic school motivation as a function of age: The mediating role of autonomy support. *Social Psychology of Education* 15(1):77–95.
- Goodman, Joshua. 2014. Flaking out: Student absences and snow days as disruptions of instructional time. NBER Working Paper No. 20221.
- Graham, Steve, and Michael Hebert. 2012. A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review* 81(4):710–744.
- Hamilton, Laura, and Daniel Koretz. 2002. About tests and their use in test-based accountability systems. In *Making sense of test-based accountability in education*, edited by Laura S. Hamilton, Brian M. Stecher, and Stephen P. Klein, pp. 13–49. Santa Monica, CA: RAND Corporation.
- Hansen, Benjamin. 2011. School year length and student performance: Quasi-experimental evidence. Working Paper 2269846, Social Science Research Network.
- Imbens, Guido, and Karthik Kalyanaraman. 2012. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies* 79(3):933–959.
- Jennings, Jennifer, and Jonathan Bearak. 2010. State test predictability and teaching to the test: Evidence from three states. Paper presented at the Annual Conference of the American Sociological Association, Atlanta, August.
- Koretz, Daniel. 2003. Using multiple measures to address perverse incentives and score inflation. *Educational Measurement: Issues and Practice* 22(2):18–26.
- Koretz, Daniel. 2005. Alignment, high stakes, and the inflation of test scores. In *Uses and misuses of data in accountability testing*, edited by Joan L. Herman and Edward H. Haertel, pp. 99–118. Chicago: National Society for the Study of Education.
- Lavy, Victor. 2010. Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. NBER Working Paper No. 16227.
- Lee, Jihyun, Wendy Grigg, and Patricia Donahue. 2007. *The nation's report card: Reading 2007* (NCES 2007–496). Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences.
- Marcotte, Dave E., and Steven W. Hemelt. 2008. Unscheduled school closings and student performance. *Education Finance and Policy* 3(3):316–338.
- Mazzolini, Barb, and Samantha Morley. 2006. A double-dose of reading class at the middle and high school levels. *Illinois Reading Council Journal* 34(3):9–25.
- McCrary, Justin. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2):698–714.
- McLaughlin, Milbrey, and Lorrie Shepard. 1995. *Improving education through standards-based reform. A Report by the National Academy of Education Panel on Standards-Based Education Reform*. Stanford, CA: National Academy of Education.
- Murnane, Richard, and John Willett. 2011. *Methods matter*. New York: Oxford University Press.

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.

Paglin, Catherine. 2003. Double dose: Bethel school district's intensive reading program adds beefed-up instruction for at-risk readers from day one. *Northwest Education* 8(3):30–35.

Papay, John P., Richard J. Murnane, and John B. Willett. 2010. The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis* 32(1):5–23.

Porche, Michelle, Stephanie Ross, and Catherine Snow. 2004. From preschool to middle school: The role of masculinity in low-income urban adolescent boys' literacy skills and academic achievement. In *Adolescent boys: Exploring diverse cultures of boyhood*, edited by Niobe Way and Judy Y. Chu, pp. 338–360. New York: New York University Press.

Rouse, Cecilia, and Alan Krueger. 2004. Putting computerized instruction to the test: A randomized evaluation of a "scientifically based" reading program. *Economics of Education Review* 23(4):323–338.

Schochet, Peter, Thomas Cook, Jonathan Deke, Guido Imbens, J. R. Lockwood, Jack Porter, and Jeffrey Smith. 2010. *Standards for regression-discontinuity designs*. Available http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf. Accessed 4 December 2013.

Schlottter, Martin, Guido Schwerdt, and Ludger Woessmann. 2011. Econometric methods for causal evaluation of education policies and practices: a non-technical guide. *Education Economics* 19(2):109–137.

Shen, Ce, and Joseph J. Pedulla. 2000. The relationship between students' achievement and their self-perception of competence and rigour of mathematics and science: A cross-national analysis. *Assessment in Education: Principles, Policy & Practice* 7(2):237–253.

Sims, David. 2008. Strategic responses to school accountability measures: It's all in the timing. *Economics of Education Review* 27(1):58–68.

Snow, Catherine, and Elizabeth Moje. 2010. Why is everyone talking about adolescent literacy? *Phi Delta Kappan* 91(6):66–69.

Stock, James, Jonathan Wright, and Motohiro Yogo. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20(4):518–529.

Tatum, Alfred. 2008. Toward a more anatomically complete model of literacy instruction: A focus on African American male adolescents and texts. *Harvard Educational Review* 78(1):15–180.

Taylor, Eric. 2012. Allocating more of the school day to math: Regression-discontinuity estimates of returns and costs. Paper presented at the Association for Public Policy Analysis and Management Conference, November, Baltimore.

Vaughn, Sharon, Janette Klingner, Elizabeth Swanson, Alison Boardman, Greg Roberts, Sarojani Mohammed, and Stephanie Stillman-Spisak. 2011. Efficacy of

collaborative strategic reading with middle school students. *American Educational Research Journal* 48(4):938–964.

Wang, Ming-Te, and Rebecca Holcombe. 2010. Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal* 47(3):633–662.

Wanzek, Jeanne, and Sharon Vaughn. 2008. Response to varying amounts of time in reading intervention for students with low response to intervention. *Journal of Learning Disabilities* 41(2):126–142.

What Works Clearinghouse (WWC). 2012. *Find what works*. Available <http://ies.ed.gov/ncee/wwc/FindWhatWorks.aspx?o=6&n=Reading%2fWriting&r=1>. Accessed 30 October 2013.