American Economic Journal: Applied Economics 2016, 8(4): 100–127
http://dx.doi.org/10.1257/app.20150112

Buy the Book? Evidence on the Effect of Textbook Funding on School-Level Achievement†

By Kristian L. Holden*

This paper considers the effect of textbook funding on school-level test performance by using a quasi-experimental setting in the United States. I consider a lawsuit in California that provided a one-time payment of \$96.90 per student for textbooks if schools fell below a threshold of academic performance. Exploiting this variation with a regression discontinuity (RD) design, I find that textbook funding has significant positive effects on school-level achievement in elementary schools and has a high benefit-per-dollar. In contrast to elementary schools, I find no effect in middle and high schools though these estimates are very imprecise. (*JEL* H75, I21, I22, I24, I28)

Evidence on the effect of school resources on achievement is mixed. In a series of influential reviews, Hanusheck (1981, 1986, 2003) argues that most research on the education production function finds little evidence that improvements in pupil teacher ratios, teacher experience, and teacher qualifications cause improvements in achievement. In contrast, a growing body of literature uses experimental or quasi-experimental methods to provide evidence that school resources affect achievement. Using these methods, Krueger (1999), Angrist and Lavy (1999), and many others provide evidence that reducing pupil-teacher ratios increases achievement.[1] Similarly, improvements in teacher quality and increasing instructional hours have been shown to improve achievement (Rockoff 2004; Rivkin, Hanushek, and Kain 2005; Lee and Barro 2001; Eren and Millimet 2007; Marcotte and Hansen 2010). However, experimental and quasi-experimental settings are rare and little work has focused on capital-related inputs.[2] This paper seeks to fill part of this gap by analyzing the effect of textbook spending on school-level achievement using a quasi-experiment.

* Center for Analysis of Longitudinal Data in Education Research at the American Institutes for Research, 1050 Thomas Jefferson Street NW, Washington, DC 20007 (e-mail: kholden@air.org). I would like to thank Ben Hansen, Jason Lindo, and Glen Waddell for comments and suggestions.

† Go to http://dx.doi.org/10.1257/app.20150112 to visit the article page for additional materials and author disclosure statement or to comment in the online discussion forum.

[1] See also Ding and Lehrer (2010), Urquiola (2006), Dustmann, Rajah, and Soest (2003), Jepsen and Rivkin (2009).

[2] Notable exceptions include the research on school facility investments; see Cellini, Ferreira, and Rothstein (2010) and Jones and Zimmer (2001), and the research on classroom computers by Angrist and Lavy (2002). See Schneider (2002) for a thorough survey on school facilities that includes research outside the field of economics of education.

While researchers think that textbooks can affect achievement in the right setting (for example, see Hanushek 1995), textbooks may have remained unstudied in the United States for two reasons.[3] First, identifying the causal impact of a school input is difficult. If, for example, textbook shortages are associated with local poverty levels, then the fact that schools without textbooks have lower student achievement could actually reflect a causal relationship between poverty and student achievement. Second, there is virtually no data on the stocks of textbooks in schools, meaning even simple correlations between textbooks and achievement cannot be studied directly.

If textbooks affect education outcomes, then the reported shortages of textbooks in the United States are cause for concern. California has experienced several reports of textbook shortages in 1996, 2000, and 2010 (See reports by California Community Foundation 1998, Oakes 2003).[4] Textbook shortages in 2000 were severe enough to motivate students in California schools to file a large lawsuit against the state. Furthermore, textbook shortages are not limited to California. New York City schools experienced sharp declines in textbook funding and corresponding shortages in textbooks in the following years (Stringer 2003). In 1997, 24 percent of New York teachers reported that they could not assign homework because of a lack of textbooks and 21 percent indicated that their classes are disrupted because students had to share textbooks in class (Stringer 2002). Similar issues have been reported in the Houston Independent school district.[5] Schools in Denver, Detroit, New Orleans, Indianapolis, and Rochester have also reported "serious problems" with providing textbooks to students.[6] These reports suggest that textbook shortages are a common and reoccurring problem in many major school districts throughout the United States, yet no one has estimated the impact these shortages have on academic outcomes.

Textbooks may affect achievement through several mechanisms. When reporting shortages, teachers tend to emphasize that textbooks enable a student to complete homework. Descriptive evidence suggests that teachers assign less homework when their students do not have textbooks to take home, and experimental evidence suggests that homework improves student achievement (Paschal, Weinstein, and Walberg 1984).[7] Consequently, a shortage of textbooks may impact student achievement by reducing the amount of homework assigned to students. Additionally, because textbooks facilitate study outside of the classroom, it is likely that the presence of textbooks may have an effect similar to increasing instructional hours in the classroom, which has been shown to increase student achievement (see Eren and Millimet 2007; Marcotte and Hansen 2010). Furthermore, Houtenville and Conway (2008) suggest that parental involvement has large effects on student achievement; and if parents use textbooks to help their children learn at home, then textbook shortages may reduce the effectiveness of parental effort.

[3] Most work textbooks and student achievement come from developing countries. See Glewwe, Kremer, and Moulin (2009); Heyneman, Jamison, and Montenegro (1984); Jamison et al. (1981).

[4] Koskey, Andrea. 2010. "Lack of Books Strains Students." The *San Francisco Examiner*. Oct 5. http://archives.sfexaminer.com/sanfrancisco/lack-of-books-strains-students/Content?oid=2163651.

[5] Lloyd, Jennifer R. 2010. "Textbooks Lagging as New School Year Dawns." *Houston Chronicle*. Oct 5. http://www.chron.com/news/houston-texas/article/New-textbooks-will-be-absent-as-school-starts-2135128.php.

[6] Fatsis, Stefan. 1988. "Urban Schools Facing Textbook Shortage." The *Prescott Courier*. Oct 5.

[7] Descriptive evidence comes from teacher and student testimony in the Williams case.

This paper provides an estimate of the causal effect of textbook funding on test performance in the United States. I estimate these effects by exploiting a quasi-experiment generated by a large lawsuit settlement in California. As part of the Williams settlement, the state allocated a one-time payment of $96.90 per student for textbooks to schools with average test scores below a threshold. My identification strategy leverages this sharp cutoff for textbook funding for a panel of California public schools over a period of ten years (2002–2011).

A key part of this analysis is to verify that textbook funding affects student performance through the purchase of additional textbooks and not through changes in other school inputs. California school districts are not required to report school-level expenditures, so no school-level data on expenditures is available, and it is not possible to evaluate school-level expenditures using the sharp cutoff for school-level funding eligibility. As such, I investigate two other possible channels. First, I find no evidence that the additional textbook funding causes increases in other school inputs, such as the number of teachers employed in each school. Second, I compare trends in district-level financial records to show that districts with schools affected by the Williams funding significantly increased district-level spending on textbooks relative to textbook spending before the Williams funds were disbursed.

The analysis reveals that textbook funding significantly increases standardized test scores for math and reading in elementary schools. My preferred estimates indicate that a one-time increase in funding of $96.90 per student improves school-average test scores by 0.15 school-level standard deviations, which corresponds to about 0.07 student-level standard deviations.[8] This effect is likely driven by improvements in student achievement, as opposed to changes in composition, because the average characteristics of schools are not affected by the change in funding. Furthermore, this estimate is robust across a number of different specifications and corresponds to the timing of the state's disbursal of textbook funding. The magnitude of this effect is comparable to estimates of other school-level education interventions in California, but this textbook intervention has a very high benefit-per-dollar ratio. In contrast, I find no significant effect of textbook funding on middle or high schools. However, these estimates are far less precise; the standard errors of estimated effects are roughly five times as large as those for elementary schools.

The usual diagnostics of RD designs support the validity of these findings, however, there are a few cautionary points that are important to consider. First, there are marginally significant effects on test scores in 2004—prior to schools receiving textbook funding—that correspond to about 0.05 student-level standard deviations. Given that school characteristics are smooth across the threshold, it seems unlikely that schools below the cutoff are higher performing; instead, this improvement may suggest that some schools respond to the announcement of the Williams settlement in anticipation of textbook funding. At worst, if schools below the threshold are slightly higher performing for reasons unrelated to textbook funding, it would be prudent to view the effect of textbook funding as the change in performance from 2004 to 2005. Second, the estimated discontinuity in 2005 is being driven by a

[8] The conversion from school to student-level standard deviations is discussed further in Section IIA.

change in the slope of points near the threshold, which may be less visually compelling than a parallel shift in test scores.[9] Lastly, the estimated effects are particularly relevant for low-performing schools that have few textbooks, and it is not clear how these results would generalize to high-performing schools or schools with greater baseline textbook funding.

This work is related to a large literature examining the effect of litigation on school finance and student achievement. Lawsuits are primarily responsible for changes in state education finance systems and several studies find that litigation raises district resources and reduces funding inequality (see Corcoran and Evans 2008 and Sims 2011, while Springer, Liu, and Guthrie, 2009 find no statistical difference in school finance patterns following litigation), but these studies also find mixed evidence of the effect of these resources on student achievement (positive effects are found by Card and Payne 2002, Downes and Figlio 1997; and negative or zero effects are found by Hoxby 2001 and Husted and Kenny 2000). This paper suggests that increases in school finance that target particular school inputs may be more successful than increases in general funding.

This paper also speaks to growing concerns about the ability to implement large scale policies and achieve outcomes similar to small scale quasi-experiments. For example, Jepsen and Rivkin (2009) provide evidence that the large-scale class size reduction program in California resulted in unanticipated tradeoffs between class size reduction and teacher quality. Similarly, Sims (2008) suggests that California schools strategically structured classes to receive cash payments for class size reduction. Both of these responses result in smaller effects for reducing class size than small scale quasi-experiments suggest. In contrast, this paper uses variation in textbook funding that affected 20 percent of the schools in California and suggests that a large-scale policy for textbook funding can impact student achievement.

The remainder of this paper proceeds as follows. In Section I, I discuss the quasi-experiment that provided textbook funding. In Section II, I describe the data and empirical methodology. In Section III, I present the main results for the analysis of textbook funding on student achievement and investigate potential mechanisms for the estimated effects. Section IV concludes. There are three online Appendices. Appendix A includes falsification tests for individual school characteristics. Appendix B provides alternate specifications, and Appendix C expands on the question of fiscal substitution by presenting more evidence on district spending patterns.

I. Background on the Williams Settlement

This study provides evidence on the effects of textbook funding by focusing on a quasi-experiment generated by Eliezer Williams et al. versus State of California (commonly referred to as the Williams case), which was a class action lawsuit filed on May 17, 2000. Plaintiffs testified that conditions were very poor in the 72 public schools involved in the Williams case. These schools lacked textbooks, qualified instructors, and safely maintained buildings.

[9] As discussed in Section III, this change in the underlying functional form is consistent with marginally higher performing schools benefiting more from textbook funding.

Evidence from the trial suggests that textbooks were in very short supply in all schools named in the lawsuit. Most teachers only had a classroom set, i.e., one set of books for all of a teacher's students, which required multiple students to share the same textbook in class. This lack of textbooks also prevented students from taking textbooks home with them, and several teachers reported that they had to assign less homework as a result. Testimony from students noted that most students shared textbooks in class, sometimes with three to four students per book. The condition of the books was notably poor and significantly outdated. For example, the social studies text that Luther Burbank Middle school students used was so old that it did not reflect the breakup of the former Soviet Union.

The state of California agreed to a settlement in 2004. The settlement established a new standard that "each pupil, including English learners, has a textbook or instructional materials, or both, to use in class and to take home to complete required homework assignments."[10] The state also provided two sources of funding to help schools meet these new standards for textbooks. The first source of funding was the Instructional Materials Fund, which provided all schools with an annual fund for textbooks. Although the program was scheduled to end in 2006, the Williams settlement continued the program with an allocation of $380.3 million. This increased funding for textbooks from $25 per student to $54.22 per student. The second fund, called Instructional Materials–Williams Case (IMWC), was designed to provide additional support for low-performing schools. In addition to the $54.22 per student that every school received, $138 million was allocated to low-performing schools for textbooks.

This paper focuses on the $138 million in IMWC funding that was distributed to low-performing schools. In particular, the settlement used a sharp cutoff for eligibility that provides potentially exogenous variation in the amount of textbook funding for schools. The cutoff restricted IMWC funding to schools in the first two deciles of the academic performance index (API) in 2003. For each type of school—elementary, middle, and high school—a particular API score is chosen as the upper limit for each decile. For example, in 2003, all elementary schools with an API score of 643 or less were within the first or second decile. Thus, a school's API score precisely determined if the school received IMWC funding. Importantly, the Williams settlement did not use this threshold to allocate other types of funding for monitoring, building repair, or other services.[11]

The Williams settlement provides an opportunity to examine the effects of textbook funding in a setting where schools are monitored. The Department of Education adopted two policies to ensure that the new textbook standard was met. First, the Uniform Complaint Process allows students, parents, teachers, and others to submit complaints about textbook insufficiencies and appeal if they are not satisfied. Second, low-performing schools are visited annually by the County Superintendent for an inspection of the stock of textbooks. In particular, the Department of Education

[10] Senate Bill 550, section 18, Education Code Section 60119(c).

[11] If eligibility for other types of funding, such as facility repair funds, also depended on this threshold, then the mechanism for improvements in student achievement would be ambiguous.

requires county oversight for schools within the first three deciles.[12] The county office staff visits schools annually to determine if students had a textbook to use in class and to take home for every subject.

Schools in the third decile are required to meet standards for textbook provision, but are given less funding relative to the second decile schools. Despite oversight, suggestive evidence from site visits and surveys of school administrators indicates that many schools simply failed to meet the state's requirement. This may be because the California Department of Education imposes no substantial repercussions on schools that did not receive additional funding; at a minimum, they are only required to make a plan to remedy insufficiencies. Correspondingly, recent litigation suggests noncompliance. A 2010 lawsuit filed against the state (*Reed v. State of California*) was due in part to a shortage of textbooks. Another lawsuit (*Doe v. State of California*) was based on an investigation that revealed that more than 40 schools were charging illegal fees for course workbooks that are required under the Williams settlement. All of this suggests that tracking textbook funding to textbook purchases is important. To investigate this issue, I use financial records for school districts in Section IIIB to examine textbook spending.

II. Data and Empirical Strategy

The data for this paper are from public schools in California with the exception of charter schools and alternative schools, which were unaffected by the Williams settlement. For each school, from 2002 to 2011, the data contains yearly records of detailed school characteristics. In particular, I focus on five categories of data for these schools: standardized testing in California provides test scores to measure student achievement; API data determines school eligibility for additional textbook funding; race and enrollment data are provided by the California Department of Education; the Common Core of Data provides the fraction of students who are eligible for free and reduced student lunch; and staff and district financial data also comes from the California Department of Education.

A. *Student Achievement and Standardized Tests*

I measure student achievement with standardized test scores from California's Standardized Testing and Reporting (STAR) program. STAR tests are taken by students annually in either April or May, depending on the start date of the school. There are four types of tests in the STAR program: the California Standards (CST), the California Achievement Test (CAT/9), the California Alternate Performance Assessment (CAPA), and the Spanish Assessment of Basic Education (SABE). I focus on the CST to measure student achievement because all students (with the exception of students with disabilities) are required to take it. The other STAR tests are taken by fewer students: CAPA tests students with disabilities; the SABE tests

[12] The cutoff for oversight does not coincide with the cutoff for IMWC funding, and the third decile cutoff provides an opportunity to study the effects of oversight. Separate from this study, my examination of the oversight threshold suggests that there is no effect of oversight on school-average test scores.

Spanish-speaking English learners and these students are required to take the CST in English as well; the CAT/9 only tests grades 3 and 11 for the 2002 to 2011 time period, while the CST tests grades 2 through 11.

The CST covers English-language arts (referred to as reading), mathematics, science, and history–social science for grades 2 through 11. The California Department of Education reports mean-scale scores for schools in addition to the percent of students within the school that meet performance criteria: far below basic, below basic, basic, proficient, and advanced. I use the mean-scale scores to measure school-level improvements and the percent meeting performance standards to examine how the distribution of student scores changes within the school.

I normalize the mean-scale scores to a mean of zero and a standard deviation of one to compare my results to other school interventions. To measure overall achievement, I average math and reading scores (called average score). These measures provide estimated effects in school-level standard deviations, while most research reports student-level. As such, I use the ratio of school-level to student-level standard deviations to convert my school-level estimated effects. Student-level standard deviations for CST are presented in STAR technical reports.[13] I use values reported in 2006, as this is the earliest reported year, which are 83.5 and 60.25 for math and reading mean scale scores. School-level standard deviations come from the sample: 34.43 and 27.24 for math and reading mean scale scores.

B. *API*, *Textbook Funding*, *and Fiscal Substitution*

The distribution of textbook funding is determined by California's API score. A school's API score is a weighted average of test scores from the STAR program and it is used to measure a school's accountability to Adequate Yearly Progress (AYP) of the No Child Left Behind program, with an API of 800 being the target performance level. It is also used to roughly compare a school's academic performance relative to other schools for California programs (such as the California Open Enrollment program examined in Holden 2013). The Williams settlement used API scores in 2003 to allocate funding for textbooks to "low-performing" schools.

The Williams settlement distributed textbook funding as a deterministic function of California's Academic Performance Index in 2003. All schools received a base amount of \$54.22 per student for textbooks each year, and elementary, middle, and high schools with API scores at or below 643, 600, and 584, respectively, receive an additional one-time payment of \$96.90 per student. I merge the 2003 API file to the STAR test score data so that schools have the same API score for each year of STAR test data.

I use data from the California Department of Education's API files to identify each school's API score relative to the cutoff for funding. The probability that a school qualifies for additional funding is a deterministic function of API score in 2003:

$$(1) \quad FundingPerStudent_{i2005} = \begin{cases} \$54.22 + \$96.90 & \text{if} \quad APINORM_{i2003} \leq 0 \\ \$54.22 & \text{if} \quad APINORM_{i2003} > 0 \end{cases},$$

[13] See http://www.cde.ca.gov/ta/tg/sr/technicalrpts.asp.

where I define $APINORM_{i2003}$ as the distance between school i's API score and the API cutoff for the twentieth percentile.[14] To construct the "first stage" of the funding for textbooks to schools, I use reported apportionments from the California Department of Education. In Section III, I use this data to show how funding was distributed and that the API threshold was strictly enforced.

A key part of this analysis relies on verifying that IMWC funding did not cause fiscal substitution. I use two datasets to investigate this issue. First, I use the Standardized Account Code Structure (SACS) to identify district-level spending on textbooks. Second, I use a rich set of school characteristics to investigate the possibility of fiscal substitution. In particular, I examine full-time equivalency (FTE), experience, and experience within-district for teachers, administrators, and pupil-service staff.

C. *Estimating Effects on Student Achievement*

I use the "sharp" RD implied by equation (1) to estimate effects of textbook funding on student outcomes at the eligibility cutoff (Trochim 1984). I use the following regression equation to estimate the effects of funding for textbooks on school outcomes:

$$(2)\quad TestScore_{ist} = \alpha + \delta \times 1(APINORM_{i2003} \leq 0) + m(APINORM_{i2003}) + u_{it},$$

where $m(\cdot)$ is a flexible, continuous function of a school's normalized API score in 2003. The coefficient of interest is δ, the estimated impact of textbook funding.[15] As mentioned previously, school districts spend the funding at different points in time. As such, my estimates can be interpreted as a sharp intent-to-treat effect and I estimate equation (2) separately for each year following the settlement.

One practical issue is how to model $m(\cdot)$. As suggested by Imbens and Lemieux (2008), I estimate the discontinuity using local linear regressions with rectangular kernel weights and heteroskedasticity-robust standard errors with bandwidths chosen by the optimal selection procedure presented in Calonico, Cattaneo, and Titiunik (2014), and show that these estimates are similar for a wide range of bandwidths.[16] In online Appendix B, I present similar results for local quadratic regressions and controls for observable school characteristics. Additionally, given the discrete nature of API scores, I present estimates with standard errors clustered on the running variable as suggested by Lee and Card (2008) and estimates with standard errors clustered by school district.

My empirical approach is motivated by the idea that schools with API scores just above the cutoff provide a good counterfactual for schools with API scores just below the cutoff. More precisely, identification relies on the assumption that school characteristics should be smooth through the cutoff (Porter 2003). This assumption

[14] This cutoff is 643 for elementary schools, 600 for middle schools, and 584 for high schools.

[15] Positive estimates of δ correspond to improvements in performance.

[16] My regression equation is given by $TestScore_{it} = \alpha + \delta \times 1(APINORM_{i2003} \leq 0) + \beta \times APINORM_{i2003} + \gamma \times APINORM_{i2003} \times 1(APINORM_{i2003} \leq 0) + u_{it}$ and restricted to a bandwidth of 19.099 API above and below the cutoff.

is plausible, as the bill that introduced the cutoff was proposed on August 24, 2004, and 2003 API scores were determined from tests taken in the spring of 2003. Thus, it is unlikely that the cutoff was known before 2003 API scores were determined. However, as it is possible that information about the cutoff was available before the bill was introduced, I investigate the smoothness of observable school characteristics through the threshold. The conclusion from this analysis is that my estimates for textbook funding are valid RD estimates.

The identification strategy provides estimates that are local to the eligibility threshold, and in this setting, schools that are near the cutoff are important for policy because they are regularly targeted for academic improvement. Table 1 compares the full sample to schools within a 19.099 API bandwidth of the cutoff. The first three rows show that the cutoff was chosen to provide resources for low-performing schools; particularly, schools near the threshold have lower math and reading scores as well as lower API scores. As we may expect, these low-performing schools tend to have higher enrollment and more Hispanic students than the average California school. Additionally, there are more students who are eligible for free or reduced school lunch within these schools. This high fraction of FRLP eligibility suggests that schools near the threshold tend to be in higher poverty areas, suggesting a correlation between low-performance and poverty. The larger schools require more teaching, administrative, and pupil-service staff.

III. Results

I present the results of my analysis in three parts. I begin by examining the validity of the RD design in this setting. Then, I examine the assignment of textbook funding to schools and the corresponding effects on textbook spending. Next, I present the main results, the effect of textbook funding on achievement. Lastly, I explore potential mechanisms for improvements in student achievement.

A. *Validity*

Nonrandom sorting is the main concern in RD designs in which those who could be effected by the policy under consideration know the eligibility cutoff. In this case, nonrandom sorting would occur if schools just above the cutoff actively influenced their 2003 API score to receive additional textbook funding. Nonrandom sorting is unlikely in this setting because the individuals affected by the policy did not know the eligibility cutoff. In particular, the bill that introduced the cutoff was proposed on August 24, 2004, and API scores were determined from tests taken in the spring of 2003. However, information about the cutoff may have been spread before the announcement of the bill in 2004. Alternatively, policy makers may have chosen the twentieth percentile as the cutoff because schools around this cutoff vary substantially in the absence of the Williams settlement. For these reasons, I investigate the identifying assumption that school characteristics are smooth through the threshold.

A key part of my analysis is to leverage the timing of the Williams settlement. If the RD design is valid in this setting, student achievement and school characteristics

Table 1—Summary Statistics

Variable	Full sample	19.099 API bandwidth around eligibility cutoff
Math score	359.89	334.45
Reading score	344.39	322.01
API score in 2003	727.68	645.15
Total enrollment	663.84	738.06
Percent Hispanic	47.01	72.18
Percent white	31.29	10.89
Percent other	19.19	15.55
Percent eligible For free or reduced lunch	54.38	80.20
FTE for teachers	32.29	36.80
FTE for admin	1.78	2.12
FTE for pupil service	1.20	1.32
Pupil/teacher	21.12	20.31
Observations	54,803	6,685

Notes: Listed means are not weighted. The full sample includes 54,803 school-year observations for public schools in California from 2002 to 2011. The restricted sample shows statistics of schools within 19.099 API from the cutoff for additional textbook funding. API score, as constructed by the California Department of Education, is a weighted average of standardized test scores.

should be smooth through the cutoff before the textbook funding is disbursed in 2005. Given the large set of characteristics available, I create a single index by regressing characteristics on pretreatment test scores, and evaluate the smoothness of this index.[17] This index greatly improves the precision of the falsification tests to help addresses concerns that there may be insufficient power to detect economically significant differences between schools to the left and right of the threshold. Figure 1 shows the average value of this index as a function of API in 2003. Each dot represents the average value across schools in bins of 3 API; there are around 42 schools in each bin per year. Visual inspection of Figure 1 indicates that schools on each side of the cutoff have similar characteristics, which suggests that the RD design is valid in this setting. Estimates of potential discontinuities are presented in Table 2. Each row shows the estimated discontinuity in indices for average test scores, math scores, and reading scores, while columns present different sets of characteristics used to construct the index.[18] Each entry in Table 2 indicates that there are no statistically significant differences in each index; furthermore, point estimates are very close to zero as the largest difference is only 0.02 of a school-level standard deviation, which is roughly equivalent to 0.005 of a student-level standard deviation.

The smoothness of some teacher characteristics merits further discussion. Table 3 presents estimated discontinuities in FTE, years of experience, and years in the school district. While most estimates are not significant, average experience

[17] Tests for the smoothness of individual characteristics are included in online Appendix A.

[18] Student characteristics include total enrollment, percent of white students, percent of Hispanic students, and percent of free or reduced lunch eligible students. Teacher characteristics include total FTE for teacher, administrators, and pupil-service staff; average teacher experience; and average teacher experience within the school district. Prior test score includes the corresponding test subject reported in the row for the preceding year.

Table 2—Regression Discontinuity Estimates of Smoothness of Predicted Student Achievement in Elementary Schools

	(1)	(2)	(3)
Discontinuity of predicted, pretreatment test scores			
Average score	−0.002	−0.007	0.012
	(0.008)	(0.008)	(0.014)
Math	0.000	−0.000	0.023
	(0.005)	(0.006)	(0.019)
Reading	−0.005	0.015	0.001
	(0.018)	(0.019)	(0.021)
Outcome predicted using:			
Student characteristics	Yes	Yes	Yes
Teacher characteristics	No	Yes	Yes
Prior test scores	No	No	Yes

Notes: Each entry is an estimated effect from a linear regression with flexible slopes, rectangular kernel weights, and a bandwidth of 19.099. Robust standard errors are displayed in parentheses. Estimates include 2002, 2003, and 2004 for a total of 1,608 observations. Average score is school average of math and reading scores. Student characteristics include total enrollment, percent of white students, percent of Hispanic students, and percent of free or reduced lunch eligible students. Teacher characteristics include total FTE for teacher, administrators, and pupil-service staff; average teacher experience; and average teacher experience within the school district. Prior test score includes the corresponding test subject reported in the row for the preceding year.

Table 3—Regression Discontinuity Estimates of Smoothness of Predicted Student Achievement in Elementary Schools

	Pretreatment			Posttreatment				
	2002 (1)	2003 (2)	2004 (3)	2005 (4)	2006 (5)	2007 (6)	2008 (7)	2009 (8)
Potential effects on staff characteristics								
Teacher FTE	2.63	2.83	2.64	2.24	2.85	2.52	1.95	2.35
	(2.64)	(2.56)	(2.50)	(2.36)	(2.08)	(1.91)	(1.85)	(1.76)
Administrator FTE	0.01	0.20	−0.12	−0.13	−0.13	−0.08	−0.09	−0.17
	(0.15)	(0.16)	(0.18)	(0.20)	(0.18)	(0.18)	(0.19)	(0.19)
Pupil staff FTE	0.36	0.03	−0.05	−0.10	−0.12	0.02	0.00	−0.27
	(0.19)	(0.19)	(0.21)	(0.26)	(0.21)	(0.18)	(0.20)	(0.19)
Years of experience	−1.07	−1.07	−0.65	−0.59	−0.73	−0.55	−0.54	−0.52
	(0.50)	(0.50)	(0.51)	(0.50)	(0.47)	(0.49)	(0.49)	(0.49)
Years in district	−0.64	−0.82	−0.39	−0.57	−0.52	−0.47	−0.36	−0.36
	(0.45)	(0.43)	(0.47)	(0.46)	(0.42)	(0.45)	(0.48)	(0.47)

Notes: Each entry is an estimated effect from a linear regression with flexible slopes, rectangular kernel weights, and a bandwidth of 19.099. Robust standard errors are displayed in parentheses. Columns 1–8 show estimated effects for individual years, with 536 observations each. Average score is school average of math and reading scores.

is significantly lower in 2002 and 2003 for schools below the cutoff. This could be because a very large number of characteristics were tested, with 90 separate tests for student and teacher characteristics, and we may expect a few tests to produce significant differences even if characteristics are smooth through the cutoff. There are two other reasons why these significant differences may not be very concerning. First,

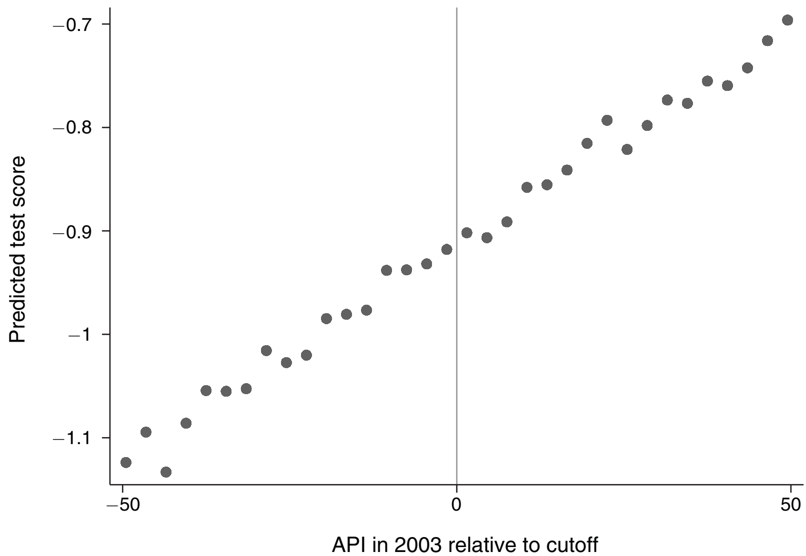


Figure 1. Predicted Pretreatment Test Scores Using School Characteristics for Elementary Schools

Notes: This figure shows a predicted test score outcome as a function of API in 2003. The vertical line shows the threshold for eligibility. Predicted test scores are constructed by regressing the characteristics mentioned in each panel on pretreatment test scores; years 2002, 2003, and 2004. Student characteristics include total enrollment, percent of white and Hispanic students, and percent of students eligible for free or reduced lunch. Teacher characteristics include FTE for teachers, administrators, and staff, teacher experience, and experience within-district.

they do not coincide with the introduction of the Williams settlement funding and likely cannot explain the improvement in performance from 2004 to 2005. Second, prior work suggests that experienced teachers tend to be more effective, and since treated schools have lower average experience, this would potentially work against finding a positive effect.

If nonrandom sorting were a problem, we may expect to see a discontinuity in the distribution of schools at the cutoff, as a disproportionate number of schools would fall just below the cutoff relative to the number of schools just above the cutoff. Figure 2 shows the distribution of schools around the cutoff as a function of API score in 2003. This figure shows that the distribution of schools is smooth through the cutoff. Using each of these cells as an observation, Table 4 shows estimates from local linear regressions with rectangular kernel weights for various bandwidths and bin sizes.[19] The estimated discontinuity is not significant, and the distribution of schools is smooth across the cutoff.

As a whole, these tests support the validity of the research design. Predicted pretreatment test scores and the distribution of schools are smooth through the cutoff. As such, the changes in school outcomes across the threshold presented in previous sections can be attributed to the funding for textbooks provided by the Williams settlement.

[19] This is similar to the test proposed by McCrary (2008).

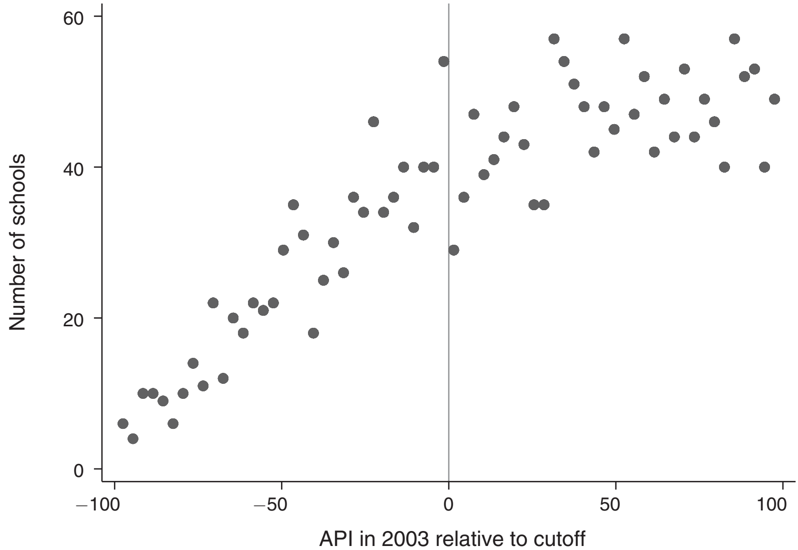


Figure 2. Distribution of Elementary Schools Around Eligibility Threshold

Notes: Each point shows the number of schools with an API score of "*x*" in 2003. The vertical line shows the threshold for eligibility.

Table 4—Regression Discontinuity Estimates of Smoothness in the Distribution of Schools

API cells	1	1	2	2	5	5
Bandwidth	20	50	20	50	20	50
	(1)	(2)	(3)	(4)	(5)	(6)
1(API in 2003 <= 643)	2.34	1.19	4.61	2.44	13.80	6.72
	(2.36)	(1.48)	(2.59)	(2.25)	(11.49)	(7.34)
Constant	13.57	13.51	27.26	27.04	65.05	66.88
	(1.86)	(1.11)	(2.15)	(1.35)	(10.35)	(5.93)
Observations	40	100	20	50	8	20

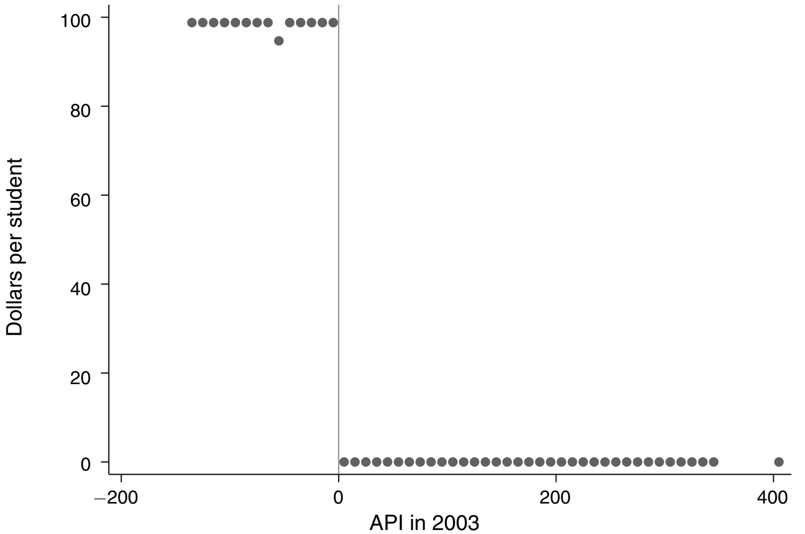
Notes: Robust standard errors are displayed in parentheses. Estimates are based on linear regressions with rectangular kernel weights.

B. *Assignment of Textbook Funding*

I begin by showing the effect of the API cutoff on the allocation of textbook funding to school districts. Panel A of Figure 3 shows textbook funding from the Williams case as a function of API in 2003 and the vertical line represents the cutoff for eligibility. Consistent with the program's description in Section I, there is a sharp change in the allocation of textbook funding across the cutoff. Schools at or below their API cutoff receive the IMWC one-time payment of $96.90 per student.[20] Only one school did not receive funding according to the schedule. I now turn to whether

[20] For context, the Williams settlement also allocated $54.22 of textbook spending per student each year for all schools, regardless of API score in 2003.

Panel A. School-level assignment of IMWC textbook funding per student in 2005



Panel B. District-level spending on textbooks per student by number of IMWC schools

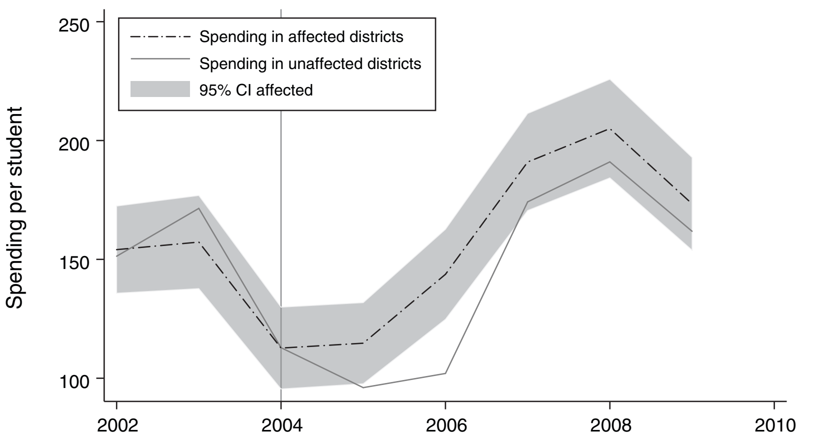


Figure 3. Assignment and Spending of Textbook Funding, School Level Eligibility, District-Level Spending

Notes: Panel A shows Williams Settlement textbook funding as a function of API score in 2003 normalized to zero at the threshold. The vertical line shows the threshold for eligibility. Schools at or below the threshold receive an additional $96.90 per student from the IMWC fund. Schools are averaged into five API score bins. Panel B shows district-level spending on textbooks per student over time. The dashed line shows spending in districts with at least one school that qualifies for IMWC funding. The solid line shows spending in districts with no qualifying schools. The vertical line indicates the last school year prior to the allocation of IMWC funding.

the stock of textbooks likely increased in treated schools. The ideal comparison would be to replicate the RD figure from panel A, however, school-level expenditure data is not collected in California.[21]

[21] This is most likely because school districts are the financially responsible entity.

Next, I investigate the effect of the program on textbook spending at the district level using a difference-in-differences style approach. In panel B of Figure 3, I examine aggregate textbook spending for two types of districts: the dashed line represents spending on textbooks for districts with at least one school that qualified for IMWC textbook funding and the solid line shows spending on textbooks for districts with no qualifying schools and therefore should not be affected by IMWC textbook funding. Importantly, pretreatment trends in spending are similar for both groups: textbook spending is not statistically significantly different in 2002 through 2004 between both types of districts and track closely despite fluctuations from year-to-year. After the funding was distributed, districts with at least one qualifying school had persistently higher textbook spending relative to districts that have no qualifying schools. Therefore, IMWC funding likely increased textbook spending and the stock of textbooks in treated schools, although the large confidence intervals on spending make it difficult to say more; at best, we can rule out that more than half of the textbook funds were diverted to other types of spending for 2006, but not for other years.

One possible concern with this approach is that the increase in textbook spending for affected districts may be the result of unobserved financial changes that are unrelated to textbook spending. To further investigate this possibility, I examine non-textbook spending in online Appendix C. Figure C.1 shows categories of spending that should not be affected by IMWC funding, including equipment, instructional aids, services provided through sub-agreement, operations and housing, and travel and conferences. Each panel suggests that affected and unaffected districts have similar trends in non-textbook spending before and after the IMWC funding was provided.

Additionally, if fiscal substitution is causing some other input to have an effect on test scores, we may expect to observe large changes in the observable characteristics of schools (e.g., class size, teacher characteristics, student sorting), which can be examined using a school-level RD analysis. To illustrate, Table 3 shows potential discontinuities in inputs like hours worked by teachers and staff. None of the post-treatment estimates suggest that textbook funding had significant impacts on these characteristics. Figure 4 presents a corresponding visual inspection of teacher FTE in 2005. Furthermore, substitution likely cannot explain the magnitude of the main effects because the program provided little funding relative to interventions like class size reduction. For example, if IMWC funds caused the hiring of more teachers, it would only reduce average class size by around 0.25 students, which would likely have a very minimal impact on student performance. Figure 3, panel B also suggests that the difference in spending declines over time as schools spend their one-time payment. An examination of IMWC funding expenditures, as opposed to total textbook spending, suggests that the majority of program funding is spent in the first 3 years (about 76 percent of total IMWC funds), while a substantial fraction (about 24 percent) is not spent until after 2007. As such, I examine the effects of textbook spending over time as well as the cumulative impact.

C. *Main Results: Student Achievement in Elementary Schools*

I estimate the effects of textbooks on student achievement by comparing test score outcomes before and after IMWC funding was provided to schools. For prior

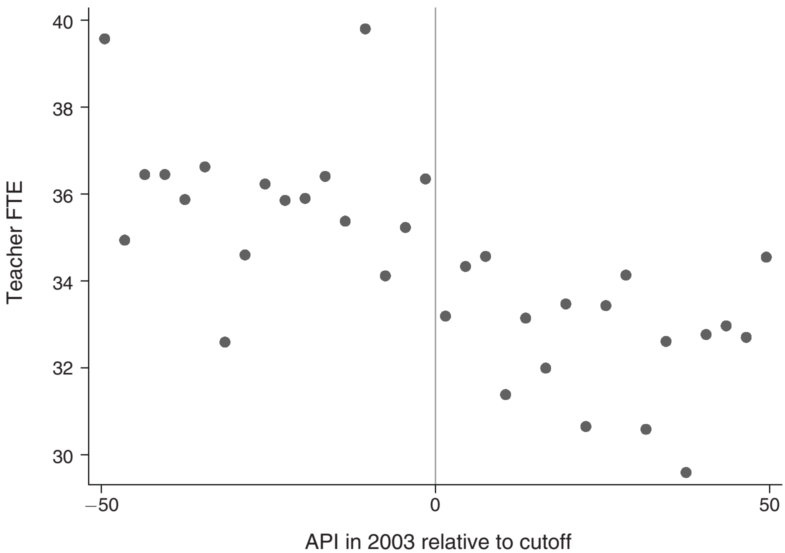


FIGURE 4. NO APPARENT SUBSTITUTION EFFECTS ON TEACHER FTE

Notes: Each panel shows a school's total teacher FTE in 2005 as a function of API score in 2003 normalized to zero at the threshold. The vertical line shows the threshold for eligibility. Each dot shows an average over three API scores. The estimated discontinuity is 2.24, and not statistically significant.

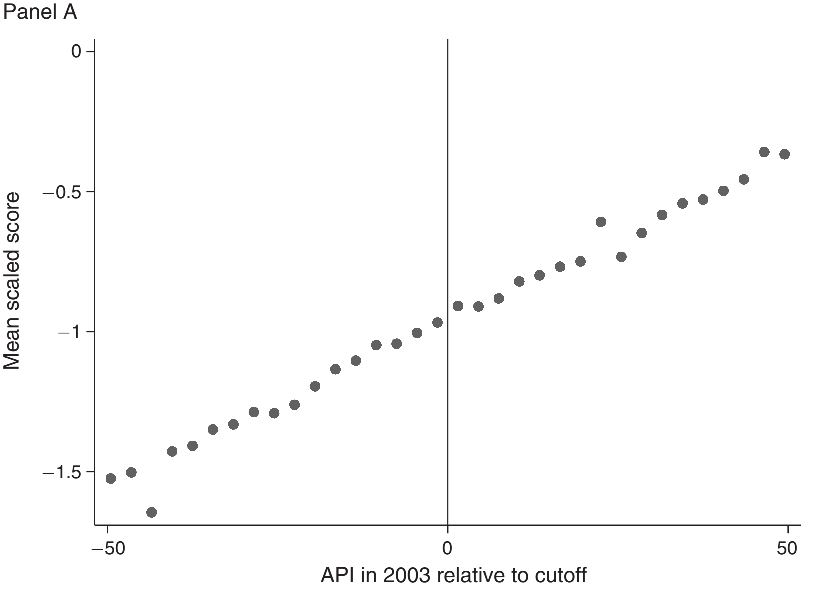
achievement, I use school-average test scores in 2003 since this year is unambiguously pre-Williams settlement; in contrast, achievement in 2004 may be affected if low-performing schools preemptively increased textbook funding in anticipation of Williams funds.[22] Panel A of Figure 5 shows school-level test scores in 2003, before the IMWC textbook funding is assigned, and panel B shows school-level test scores in 2005, after the IMWC textbook funding is distributed, both as a function of 2003 API.[23] Visual inspection suggests that test scores are discontinuously higher after IMWC textbook funding is assigned while pre-treatment outcomes are similar.

The effect is most noticeable for the schools within 10 points to the left of the cutoff, but the number of observations in this range is substantial. There are 150 elementary schools that serve a total of 105,052 students and received roughly $10 million in textbook funding as part of the intervention. Another notable feature is the increase in the slope of the underlying functional form just below the cutoff. This suggests the underlying relationship between test scores and API in 2003 has changed since the introduction of textbook funding. It is difficult to say anything causal, but the increase in slope is consistent with marginally higher performing schools in 2003 benefitting more from textbook funding in 2005 than lower performing schools in 2003.[24] This could be because higher performing schools

[22] Appendix C shows achievement figures for 2002, 2003, and 2004. Visual inspection suggests that there are no apparent differences, though effects are marginally significant in 2004.

[23] Each panel shows the school-level average of reading and math test scores across all grades. A separate figure for reading and math scores is presented in online Appendix C.

[24] Schools within ten points to the left of the cutoff do not appear to be clustered on a particular characteristic. The falsification tests included in online Appendix A suggest that characteristics are similar at the cutoff, and visual



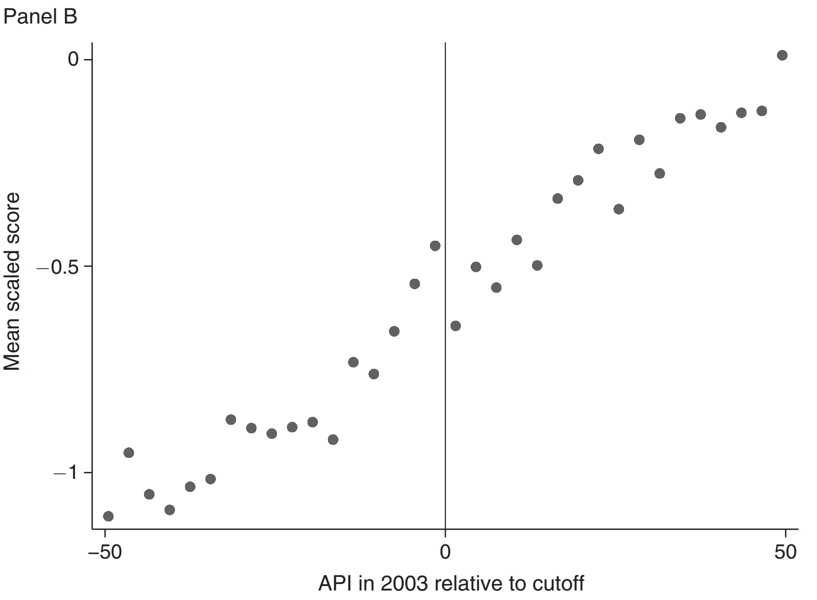


Figure 5. Effect of Textbook Funding on Student Achievement in Elementary Schools

Notes: Both panels show average test scores at the school as a function of API score in 2003. The vertical line shows the threshold for eligibility.

make better use of textbooks, or that higher performing students in these schools benefit more than lower performing students, consistent with findings in developing countries (e.g., Glewwe, Kremer, and Moulin 2009).

inspection of pretreatment characteristics in Figure A.2 and Figure A.3 suggests there is also no apparent clustering around ten points to the left of the cutoff.

Table 5—Regression Discontinuity Estimates of Effect of Textbook Funding on Student Achievement in Elementary Schools

	All post years	Pretreatment			Posttreatment			
		2002	2003	2004	2005	2006	2007	2008
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average score	0.15	0.05	0.01	0.10	0.20	0.16	0.18	0.12
	(0.05)	(0.05)	(0.03)	(0.05)	(0.07)	(0.09)	(0.09)	(0.10)
Math	0.12	0.01	−0.04	0.09	0.22	0.18	0.16	0.09
	(0.05)	(0.06)	(0.05)	(0.07)	(0.09)	(0.10)	(0.11)	(0.11)
Reading	0.17	0.09	0.02	0.10	0.19	0.15	0.20	0.15
	(0.05)	(0.07)	(0.05)	(0.06)	(0.08)	(0.09)	(0.09)	(0.10)

Notes: Each entry is an estimated effect from a linear regression with flexible slopes, rectangular kernel weights and a bandwidth of 19.099. Robust standard errors are displayed in parentheses. Estimates in all post years pool all observations from 2005 to 2009 for a total of 3,750 observations. Columns 2–9 show estimated effects for individual years, with 536 observations each. Average score is school average of math and reading scores.

In Table 5, I present estimates for the observed discontinuity seen in panel B of Figure 5, as well as math and reading scores by year. Each entry represents a separate regression, where the row identifies the dependent variable measuring student achievement and the column represents the year.[25] Similar to panel A of Figure 5 and the falsification tests in online Appendix B, columns 2 to 4 present the estimated discontinuities in test scores as a falsification test, while columns 1 and 5 to 9 present estimated program effects. The estimates in column 5 indicate an effect on student achievement of 0.20 standard deviations in the first year that is statistically significant at the 1 percent level for average scores. The estimated effect declines to 0.16 and 0.18 standard deviations in 2006 and 2007 and is not significant in later years. Similar effects are found within both math and reading scores, with slight differences. Math scores show slightly larger effects than reading scores in the first few years, but decline below the reading effects in 2007 and 2008.

To examine how the effects of textbook funding have changed over time, Figure 6 shows the estimated effects by year, including effects in 2009, 2010, and 2011. The figure suggests that effects decline over time, which is reasonable given the details of the program. First, the pattern of effects on student achievement is similar to pattern the spending of additional textbook funding. Spending of the additional textbook funding is high in 2005, 2006, and 2007, and falls to around 5 percent in 2008 and 2009. This suggests that a constant source of funding of textbooks may be required to prevent large changes in the stock of textbooks. While detailed information on the loss rate of textbooks is not available, L.A. County has a general rule that 10 percent of the textbook stock will need to be replaced every year. Back of the envelope calculation suggests that around $16,000 would be needed each year to maintain the stock of textbooks. In fact, we may be concerned if a one-time

[25] Each column presents estimates for a different year. All effects are estimated with linear regressions with a bandwidth of 19.099 API and heteroskedasticity-robust standard errors. Estimates from flexible quadratic regressions, district fixed effects, and multiple methods of clustering standard errors are available in online Appendix B; each suggest similar results.

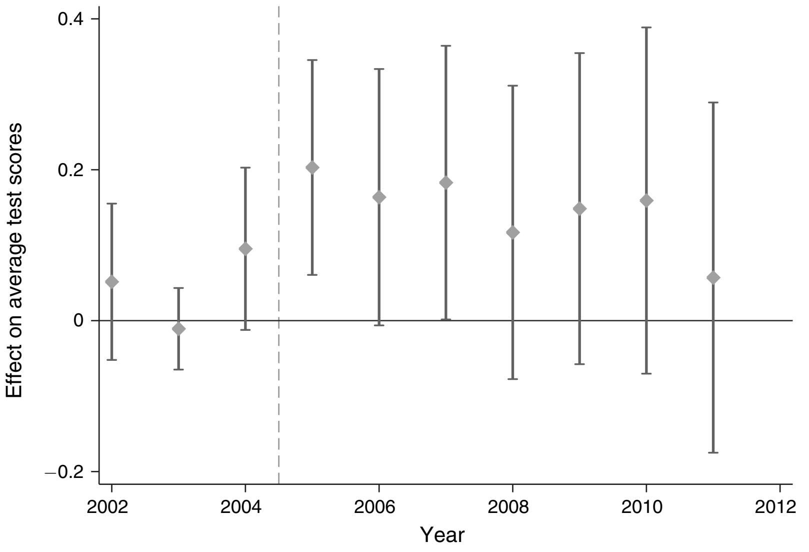


Figure 6. Effect of Textbook Funding on Student Achievement in Elementary Schools by Year

Notes: Each panel shows the average test score in each year as a function of API score in 2003 normalized to zero at the threshold. The vertical line shows the threshold for eligibility. Each dot shows an average over three API.

payment for textbooks leads to persistent differences in test scores. Second, it may be that the marginal benefits of textbook funding may decrease quickly as the stock of textbooks increases. During this period, all schools receive an additional $54.22 each year for textbooks, and thus, schools that did not qualify for IMWC funding may be increasing their stock of textbooks as well. After four years, all schools may have sufficient numbers of textbooks, eliminating the advantage of IMWC textbook funding.

How large are these estimates compared to other school-level interventions? Two studies are ideal for comparing estimates of textbooks to class size reduction. Jepsen and Rivkin (2009) and Unlu (2005) estimate the effect of a massive class size reduction program for the same set of elementary schools in California.[26] Jepsen and Rivkin (2009) find that a 10-student reduction in class size improved reading and math scores by 0.10 and 0.06 standard deviations of the school test distribution, while Unlu (2005) finds larger estimates of 0.3 and 0.2 standard deviations. I find an effect of around 0.20 standard deviations of improvement in student achievement in 2005. While both interventions have roughly similar effects on students, the class size reduction program exceeded $1.7 billion (or $1,024 per pupil) annually while the entire IMWC program cost $138 million (or $96.90 per pupil). This suggests that the textbook intervention had a very high benefit per dollar.

[26] This program, called the Class Size Reduction program, started in 1996 and is still in use. Importantly, eligibility does not depend on the threshold used for the IMWC funding; instead, eligiblility is nearly universal.

While very few studies have focused on textbooks in the United States, there are several recent randomized experiments that have provided additional textbooks to students in developing countries.[27] Jamison et al. (1981) study an experiment in Nicaragua that provided additional textbooks to first-grade students, and found that math scores increased by approximately 0.33 standard deviations. Another study by Heyneman, Jamison, and Montenegro (1984) found that textbook provision to first and second-grade students in the Philippines increased math and reading scores by around 0.3 standard deviations. A study by Glewwe, Kremer, and Moulin (2009) suggests more nuanced results; textbooks were provided in English, and while this is the language of instruction, English is the third language of most students; only the highest performing students benefited from the additional textbooks with students in the highest quintile of pretest performance improving test outcomes by about 0.22 standard deviations. In total, these results are qualitatively similar to the findings from the Williams case, though larger in magnitude, perhaps suggesting diminishing returns to textbooks, as untreated students in developing countries tend to have very few textbooks. For example, 80 percent of the untreated Kenyan classes had less than 1 English textbook for every 20 students, and most untreated classes in the Nicaragua study had about 3 books for 20 students.

In Table 6 and Figure 7, I explore the distribution of the effect on student achievement by examining the percent of students that meet various performance standards. Each row in Table 6 shows effects for a different dependent variable. For example, the entry in the first row of column 1 suggests that treated schools had 0.77 percentage points fewer students in the lowest performance category. Similarly, the entry in the second row of column 2 shows that treated schools had 1.07 percentage points fewer students classified as below standard. The last two rows of column 1 suggest that treated schools increased the percentages of proficient and advanced students by 1.09 and 1.15 percentage points.[28] These estimates suggest that textbook interventions not only improve test performance for low-performing students, but they also increase the percent of high performing students in the school.

In Table 7, I examine disaggregated, grade-level test score outcomes. Each column represents a cohort of students who take the four tests in elementary school; moving from left to right through columns 1 to 4, cohorts are progressively more impacted by the program. For example, the cohort in column 1 only receives textbook funding in their last year while the cohort in column 4 is affected each year. Each entry in a row represents the estimated effect from a separate regression. For example, the cohort in column 1, row 1 has an insignificant effect of -0.01 in grade two, consistent with prior falsification tests. Effects are concentrated in grades three and four while effects for grade five are not significant in any year. The patterns of effects within cohorts are interesting, though the disaggregated data is fairly noisy, so it is difficult to say whether the patterns are significant. Early exposure

[27] For a thorough review of previous work in developing countries, see Heyneman, Farrell, and Sepulveda-Stuardo (1981).

[28] The predicted effect of textbook funding on intermediate performance categories is ambiguous. For example, if textbooks improve student achievement, we may find a reduction in the percent of students in the basic category because these students are now proficient or advanced, or an increase in percent basic because of students improving from the far below basic and below basic categories.

Table 6—Regression Discontinuity Estimates of Effect of Textbook Funding on the Distribution of Student Achievement in Elementary Schools

	All post years	Pretreatment			Posttreatment			
		2002	2003	2004	2005	2006	2007	2008
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect on the percent of students in each category								
Far below	−0.77	−0.45	0.32	−0.66	−1.11	−0.96	−0.98	−0.31
Standard	(0.28)	(0.55)	(0.31)	(0.44)	(0.55)	(0.55)	(0.53)	(0.52)
Below	−1.07	−0.11	−0.04	−0.65	−1.22	−0.99	−0.94	−0.99
Standard	(0.35)	(0.57)	(0.43)	(0.55)	(0.59)	(0.62)	(0.70)	(0.72)
Meets	−0.41	0.08	−0.44	0.02	−0.69	−0.60	−0.19	−0.40
Standard	(0.26)	(0.54)	(0.46)	(0.56)	(0.56)	(0.50)	(0.54)	(0.58)
Proficient	1.09	0.11	−0.27	0.59	1.63	1.34	0.73	0.71
	(0.31)	(0.50)	(0.31)	(0.43)	(0.53)	(0.60)	(0.63)	(0.65)
Advanced	1.15	0.31	0.30	0.67	1.38	1.21	1.39	0.98
	(0.39)	(0.27)	(0.29)	(0.42)	(0.57)	(0.70)	(0.78)	(0.83)

Notes: Each entry is an estimated effect from a linear regression with flexible slopes, rectangular kernel weights, and a bandwidth of 19.099. Robust standard errors are displayed in parentheses. Column 1 shows estimated effects across all treatment years for a total of 2,680 observations. Columns 2–8 show estimated effects for individual years, with 536 observations each.

is consistent with improvements in reading; for example, third grade reading test scores increase across cohorts (effects for third grade correspond to columns 3–4: 0.13, 0.21 standard deviations). However, this pattern is also consistent with larger program effects in years 2006 and 2007, which could be due to greater textbook spending in 2006 as seen in Figure 3 panel B. In comparison, third grade math scores are fairly constant for treated cohorts (effects for third grade correspond to columns 3–4: 0.18, 0.19 standard deviations).

For robustness, Figure 8 shows the sensitivity of the estimated effect to different choices of bandwidth. Panel A shows that point estimates are fairly consistent and significant for bandwidths smaller than 25 API, while after this point, estimated effects decline. In online Appendix B, Figure B.1, I demonstrate that linear specifications with wider bandwidths do not appear to fit the curvature in the data. In contrast, a quadratic specification appears to fit the data well for wider bandwidths. As such, I present point estimates for quadratic specifications in panel B, which provide consistent bandwidths beyond 25 API. In any case, the optimally chosen bandwidth of 19.099 provides a similar, significant point estimate regardless of whether the specification is linear or quadratic.

One may be concerned at the apparent discontinuity observed at negative ten on the running variable. In Table 8, I investigate whether this gap, and others at 5 API intervals, represent significant discontinuities in average school test scores both before and after funding for textbooks is allocated. Each estimate uses the previous linear regression specification, but assumes a running variable that is shifted from the true cut point. Before the funding is allocated, estimates from row 1 suggest that only 1 point is significant at the 10 percent level. After the funding is allocated, only the true cutoff has a statistically significant discontinuity. Similarly, one may also be concerned

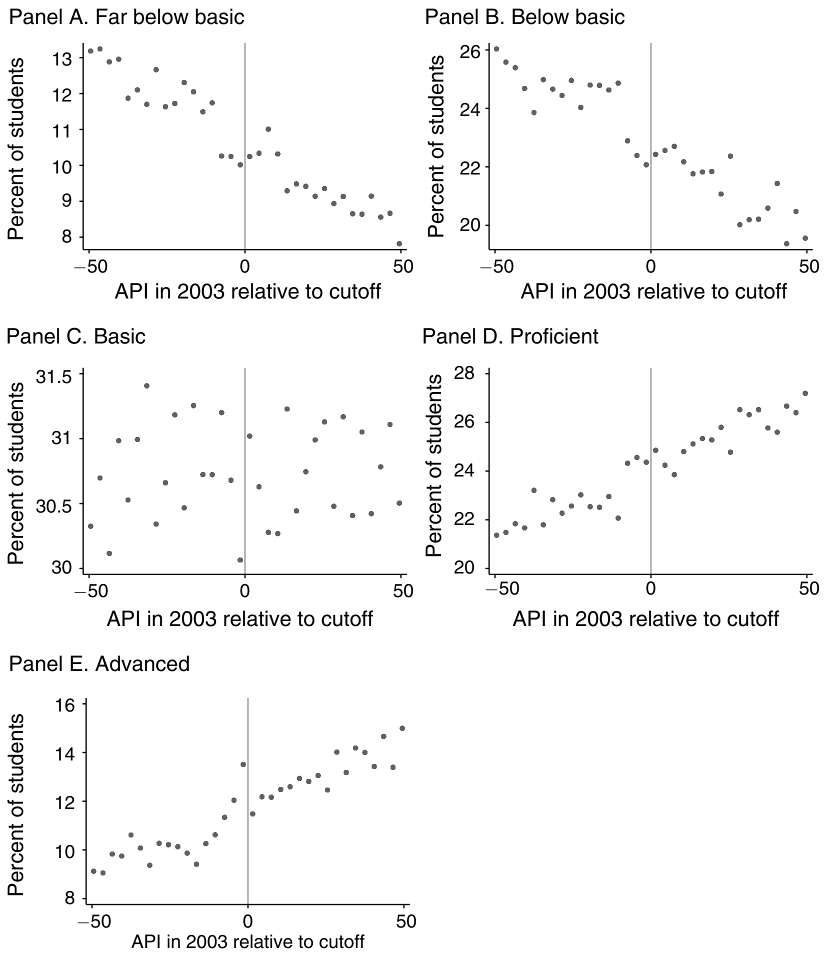


Figure 7. Effect of Textbook Funding on the Distribution of Student Test Scores in Elementary Schools—Percent of students in Performance Category

Notes: Each panel shows the percent of students in the school that are in the relevant performance category: far below basic, below basic, basic, proficient, and advanced—as a function of API score in 2003. The vertical line shows the threshold for eligibility. Each panel shows raw cell means (points) for each API score.

that decile cutoffs in 2003 may have effects on school outcomes independent of the effect of textbook funding. This does not appear to be the case, as there are no significant effects in 2004 at the second decile cutoff, and other decile cutoffs do not present any significant effect either before or after the assignment of textbook funding.

D. *Main Results: Middle and High Schools*

In Table 9, I present estimated effects on middle school and high school student achievement. The estimates for middle schools in panel A and high schools in panel B suggest that there are no statistically significant effects for any year. If

Table 7—Regression Discontinuity Estimates of Effect of Textbook Funding on Math and Reading Scores by Grade

	Math scores				Reading scores			
	Cohort entered second grade in year:				Cohort entered second grade in year:			
	2002	2003	2004	2005	2002	2003	2004	2005
Cohort:	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Pretreatment test score differences:								
Tested in year								
2002	−0.01				−0.00			
	(0.08)				(0.07)			
2003	−0.04	0.07			−0.05	0.01		
	(0.06)	(0.07)			(0.06)	(0.07)		
2004	0.08	0.10	0.04		0.07	0.03	0.02	
	(0.06)	(0.07)	(0.09)		(0.05)	(0.06)	(0.08)	
Posttreatment test score effects:								
2005	0.06	0.16	0.18	0.15	0.04	0.07	0.13	0.14
	(0.10)	(0.08)	(0.08)	(0.12)	(0.06)	(0.05)	(0.07)	(0.09)
2006		0.03	0.14	0.19		0.00	0.07	0.21
		(0.10)	(0.09)	(0.10)		(0.07)	(0.06)	(0.08)
2007			−0.03	0.11			0.01	0.15
			(0.10)	(0.10)			(0.06)	(0.07)
2008				−0.03				0.03
				(0.11)				(0.08)
2009								

Notes: Each entry is an estimated effect from a linear regression with flexible slopes, rectangular kernel weights, and a bandwidth of 19.099. Each column represents a cohort of students who are tested four times in elementary school; for example, each column begins with test score effects in grade two. Robust standard errors are displayed in parentheses.

anything, effects for high schools are somewhat positive in 2005 and 2006, though all standard errors are very large.

Why are there no effects for middle and high schools? Notably, there are far fewer middle and high schools compared to elementary schools. In California, there are roughly five times as many elementary schools than high schools or middle schools, and the small sample size is further reduced when focusing on a bandwidth of observations near the cutoff.[29] As a result, the standard errors of the estimated effects are very large and we cannot rule out economically significant effects on middle schools or high schools.

On the other hand, older students may not have benefited from textbook provision. This could occur for several reasons. Older students may not have developed the study habits necessary to use textbooks because they did not have access to textbooks in earlier grades. Teachers of older students may not adapt their curriculum to include new textbooks. Or, perhaps, middle schools and high schools place greater importance on maintaining the stock of textbooks and additional textbook funding did not increase student access to textbooks.

[29] Several researchers note that RD designs have low statistical power: see Deke and Dragoset (2012) and Schochet (2008).

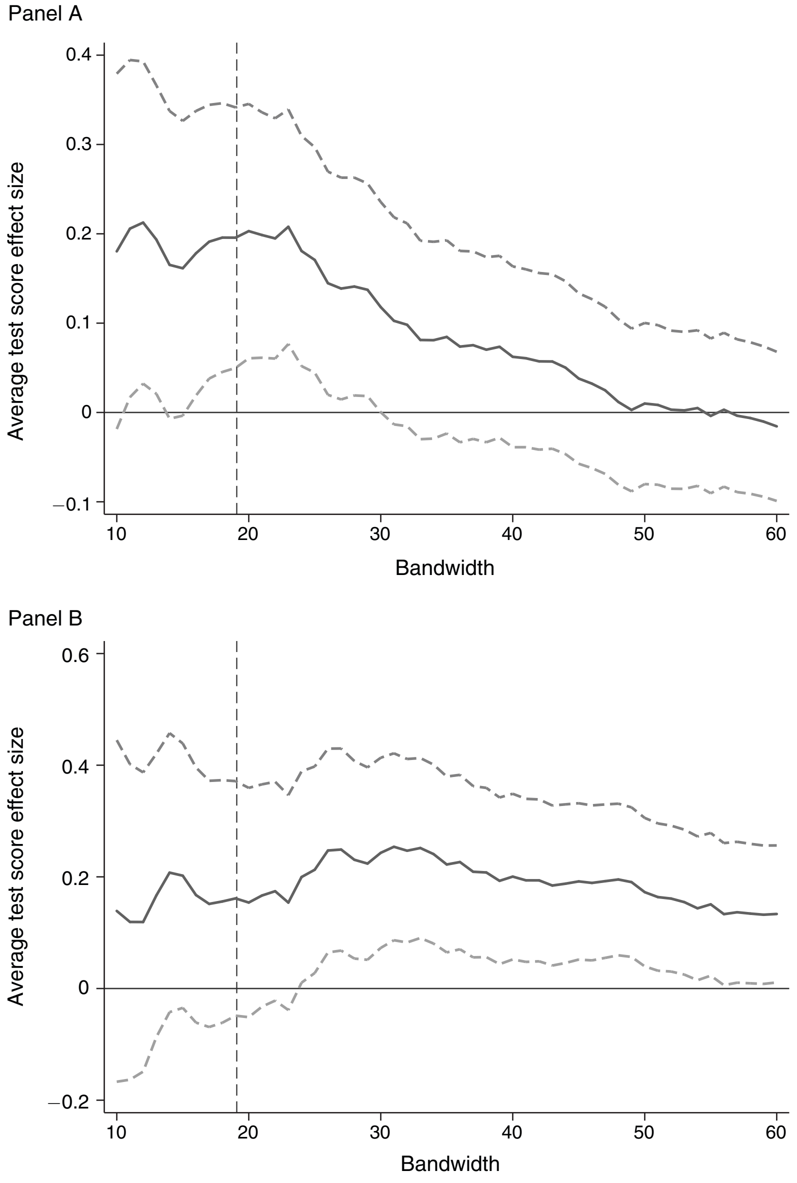


Figure 8. Robustness of Regression Discontinuity Estimates in Elementary Schools to Bandwidth Choice

Notes: The black line indicates the estimated effect on student achievement from a local linear regression with a bandwidth on the *x*-axis. The estimated dashed lines indicate 95 percent confidence intervals for the point estimate using robust standard errors. The vertical dashed lines indicate the optimal bandwidth of 19.099.

IV. Conclusion

In this paper, I identify the effect of textbook funding on student achievement by focusing on a quasi-experiment generated by the Williams settlement. Exploiting an eligibility threshold for textbook funding, I use a RD design to estimate the effect of

Table 8—Falsification Tests for Effects at False Cutoffs and Other Deciles

Assumed cutoff location (relative to actual cutoff)	−10	−5	0 (true cutoff)	+5	+10
	(1)	(2)	(3)	(4)	(5)
Effect in year 2004	−0.06	−0.09	0.10	−0.03	−0.10
	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
Effect in year 2005	−0.07	0.11	0.20	−0.09	−0.12
	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)
Observations	547	555	536	530	526
Effects at each decile	1	2 (true cutoff)	3	4	5
Effect in year 2004	−0.04	0.10	0.05	0.04	−0.01
	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)
Effect in year 2005	−0.12	0.20	0.02	0.03	−0.11
	(0.08)	(0.07)	(0.07)	(0.06)	(0.07)
Observations	404	536	591	620	544

Notes: Robust standard errors are displayed in parentheses. Estimates are based on linear regressions with rectangular kernel weights and a bandwidth of 19.099.

Table 9—Regression Discontinuity Estimates of Effect of Textbook Funding on Student Achievement in Middle Schools and High Schools

	All post years	Posttreatment 2005	2006	2007	2008	2009
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Effects on middle school student achievement						
Average score	−0.07	−0.15	−0.15	−0.15	−0.05	−0.05
	(0.09)	(0.16)	(0.15)	(0.15)	(0.18)	(0.20)
Math	−0.09	−0.16	−0.29	−0.29	−0.03	−0.10
	(0.11)	(0.27)	(0.24)	(0.23)	(0.25)	(0.28)
Reading	−0.05	−0.14	−0.02	−0.01	−0.07	0.00
	(0.10)	(0.11)	(0.13)	(0.14)	(0.17)	(0.18)
Panel B. Effects on high school student achievement						
Average score	−0.06	0.06	0.11	0.02	−0.07	−0.08
	(0.10)	(0.22)	(0.21)	(0.24)	(0.24)	(0.26)
Math	−0.06	0.09	0.07	−0.03	0.01	−0.13
	(0.13)	(0.31)	(0.28)	(0.33)	(0.34)	(0.38)
Reading	−0.06	0.04	0.14	0.07	−0.14	−0.03
	(0.11)	(0.19)	(0.19)	(0.21)	(0.20)	(0.21)

Notes: Each entry is an estimated effect from a linear regression with flexible slopes, rectangular kernel weights, and a bandwidth of 19.099. Robust standard errors are displayed in parentheses. Estimates in all years pool all observations from 2005 to 2009 for a total of 915 observations for middle schools and for high schools. Columns 2–6 show estimated effects for individual years, with 130 observations for middle schools and for high schools. Average score is school average of math and reading scores.

a one-time payment of $96.90 per student for textbook funding on student achievement. My findings suggest that textbook funding has a significant, positive effect on student achievement. I also find that the benefits of the one-time payment decline four years after funding is disbursed. Textbook funding improves student achievement at all performance levels and I find no evidence that textbook funding caused fiscal substitution.

My results also suggest that textbooks are a very cost-effective way to improve test scores compared to class size interventions. The estimates from Krueger (1999) suggest that reducing class size leads to a 0.2 standard deviation improvement in reading and math scores and costs approximately $3,501 per year for each student (Krueger 2003). My results suggest a student-level effect in math and reading around 0.074 standard deviations and 0.067 standard deviations, respectively. While substantially smaller than the effect of class size reduction, textbooks cost only $96.90 per year for each student, leading to a very high benefit per dollar.

The purpose of this paper, however, is not to discourage or replace class size reductions. We still know very little about the mechanisms that lead textbooks to affect student achievement. In particular, the marginal benefit textbook interventions may be zero when students have a book to use in class and one to take home. Instead of suggesting substitution between school inputs, I view these results as suggesting that the benefit of providing textbooks appears to exceed the very low cost of provision.

With such high spending per student, why are there textbook shortages in the United States? One clue may lie in the pattern of textbook spending: textbook spending appears to be counter-cyclical and highly volatile compared to spending on teacher salaries. It could be that budget shortages cause school districts to cut spending and textbook purchases are flexible, while teacher salaries are likely not. Alternatively, it could also be that school districts are attempting to replace textbooks with other mediums, such as laptops or tablets. These mediums may reduce the price of electronic print materials, though the high cost to replace a laptop or tablet could lead to more severe shortages or restrict take-home use. In short, more research is needed to understand why shortages happen, and whether electronic substitutes are viable solutions to shortages.

I view this paper as a first step toward understanding the effects of textbooks in the United States. There is still much to learn about textbook interventions, including further study of the effects of textbook provision on middle and high schools. Additionally, schools considered in this paper are monitored; further research can investigate if monitoring is an important policy to include with textbook provision.

REFERENCES

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): 533–75.

Angrist, Joshua, and Victor Lavy. 2002. "New Evidence on Classroom Computers and Pupil Learning." *Economic Journal* 112 (482): 735–65.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Card, David, and A. Abigail Payne. 2002. "School finance reform, the distribution of school spending, and the distribution of student test scores." *Journal of Public Economics* 83 (1): 49–82.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein. 2010. "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design." *Quarterly Journal of Economics* 125 (1): 215–61.

Corcoran, Sean P., and William N. Evans. 2008. "Equity, adequacy, and the evolving state role in education finance." *Handbook of Research in Education Finance and Policy*, edited by Helen F. Ladd and Edward B. Fiske, 332–56. New York: Routledge.

Deke, John, and Lisa Dragoset. 2012. "Statistical Power for Regression Discontinuity Designs in Education: Empirical Estimates of Design Effects Relative to Randomized Controlled Trials." http://files.eric.ed.gov/fulltext/ED533141.pdf.

Ding, Weili, and Steven F. Lehrer. 2010. "Estimating Treatment Effects from Contaminated Multiperiod Education Experiments: The Dynamic Impacts of Class Size Reductions." *Review of Economics and Statistics* 92 (1): 31–42.

Downes, Thomas A., and David N. Figlio. 1997. "School Finance Reforms, Tax Limits, and Student Performance: Do Reforms Level Up or Dumb Down?" Institute for Research on Poverty (IRP) Discussion Paper 1142-97.

Dustmann, Christian, Najma Rajah, and Arthur van Soest. 2003. "Class Size, Education, and Wages." *Economic Journal* 113 (485): F99–120.

Eren, Ozkan, and Daniel L. Millimet. 2007. "Time to learn? The organizational structure of schools and student achievement." *Empirical Economics* 32 (2–3): 301–32.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics* 1 (1): 112–35.

Hanushek, Eric A. 1981. "Throwing money at schools." *Journal of Policy Analysis and Management* 1 (1): 19–41.

Hanushek, Eric A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24 (3): 1141–77.

Hanushek, Eric A. 1995. "Interpreting recent research on schooling in developing countries." *World Bank Research Observer* 10 (2): 227–46.

Hanushek, Eric A. 2003. "The Failure of Input-Based Schooling Policies." *Economic Journal* 113 (485): F64–98.

Heyneman, Stephen P., Joseph P. Farrell, and Manuel A. Sepulveda-Stuardo. 1981. "Textbooks and Achievement in Developing Countries: What We Know." *Journal of Curriculum Studies* 13 (3): 227–46.

Heyneman, Stephen P., Dean T. Jamison, and Xenia Montenegro. 1984. "Textbooks in the Philippines: Evaluation of the Pedagogical Impact of a Nationwide Investment." *Educational Evaluation and Policy Analysis* 6 (2): 139–50.

Holden, Kristian L. 2013. "Signals of School Quality and School-Choice: Regression Discontinuity Evidence from California's Open Enrollment Program." PhD diss. University of Oregon.

Holden, Kristian L. 2016. "Buy the Book? Evidence on the Effect of Textbook Funding on School-Level Achievement: Dataset." *American Economic Journal: Applied Economics*. http://dx.doi.org/10.1257/app.20150112.

Houtenville, Andrew J., and Karen Smith Conway. 2008. "Parental effort, school resources, and student achievement." *Journal of Human Resources* 43 (2): 437–53.

Hoxby, Caroline M. 2001. "All School Finance Equalizations Are Not Created Equal." *Quarterly Journal of Economics* 116 (4): 1189–1231.

Husted, Thomas A., and Lawrence W. Kenny. 2000. "Evidence on the Impact of State Government on Primary and Secondary Education and the Equity-Efficiency Trade-Off." *Journal of Law and Economics* 43 (1): 285–308.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics* 142 (2): 615–35.

Jamison, Dean T., Barbara Searle, Klaus Galda, and Stephen P. Heyneman. 1981. "Improving elementary mathematics education in Nicaragua: An experimental study of the impact of textbooks and radio on achievement." *Journal of Educational Psychology* 73 (4): 556–67.

Jepsen, Christopher, and Steven Rivkin. 2009. "Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size." *Journal of Human Resources* 44 (1): 223–50.

Jones, John T., and Ron W. Zimmer. 2001. "Examining the impact of capital on academic achievement." *Economics of Education Review* 20 (6): 577–88.

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114 (2): 497–532.

Krueger, Alan B. 2003. "Economic Considerations and Class Size." *Economic Journal* 113 (485): F34–63.
Lee, Jong-Wha, and Robert J. Barro. 2001. "Schooling quality in a cross-section of countries." *Economica* 68 (272): 465–88.
Lee, David S., and David Card. 2008. "Regression discontinuity inference with specification error." *Journal of Econometrics* 142 (2): 655–74.
Marcotte, Dave E., and Benjamin Hansen. 2010. "Time for school." *Education Next* 10 (1): 52–59.
McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142 (2): 698–714.
Millot, Benoit, and Julia Lane. 2002. "The efficient use of time in education." *Education Economics* 10 (2): 209–28.
Oakes, Jeannie. 2003. "Education Inadequacy, Inequality, and Failed State Policy: A Synthesis of Expert Reports Prepared for Williams v. State of California." *Santa Clara Law Review* 43 (4): 1299–1398.
Paschal, Rosanne A., Thomas Weinstein, and Herbert J. Walberg. 1984. "The Effects of Homework on Learning: A Quantitative Synthesis." *Journal of Educational Research* 78 (2): 97–104.
Porter, Jack. 2003. "Estimation in the Regression Discontinuity Model." http://econ.ucsb.edu/~doug/245a/Papers/Estimation%20in%20RD%20Model.pdf.
Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73 (2): 417–58.
Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94 (2): 247–52.
Schneider, Mark. 2002. "Do School Facilities Affect Academic Outcomes?" http://www.ncef.org/pubs/outcomes.pdf.
Schochet, Peter Z. 2008. "Technical Methods Report: Statistical Power for Regression Discontinuity Designs in Education Evaluations." National Center for Education Evaluation and Regional Assistance (NCEE) 2008-4026.
Sims, David. 2008. "A strategic response to class size reduction: Combination classes and student achievement in California." *Journal of Policy Analysis and Management* 27 (3): 457–78.
Sims, David P. 2011. "Lifting All Boats? Finance Litigation, Education Resources, and Student Needs in the Post-'Rose' Era." *Education Finance and Policy* 6 (4): 455–85.
Springer, Matthew G., Keke Liu, and James W. Guthrie. 2009. "The impact of school finance litigation on resource distribution: A comparison of court-mandated equity and adequacy reforms." *Education Economics* 17 (4): 421–44.
Stringer, Scott. 2002. *Reading is* Fund*amental: A Report by Assembly Member Scott Stringer on the Textbook Crisis Facing Our Schools*. New York State Assembly. New York, April.
Stringer, Scott. 2003. *From Bad to Worse: One Year after the Release of "Reading is Fundamental*." New York State Assembly. New York, April.
Trochim, William M. K. 1984. *Research Design for Program Evaluation: The Regression-Discontinuity Approach,* Vol. 6. Beverly Hills: Sage Publications.
Unlu, Fatih. 2005. "California class size reduction reform: New findings from the NAEP." http://www.classsizematters.org/wp-content/uploads/2011/04/California_CSR_Fatih_Unlu2.pdf.
Urquiola, Miguel. 2006. "Identifying class size effects in developing countries: Evidence from rural Bolivia." *Review of Economics and Statistics* 88 (1): 171–77.

This article has been cited by:

1. Carrie Conaway, Dan Goldhaber. 2020. Appropriate Standards of Evidence for Education Policy Decision Making. *Education Finance and Policy* **15**:2, 383-396. [Crossref]
2. Qiang Gao, Mingfeng Lin, D. J. Wu. 2019. Minuscule but Impactful: The Effects of Educational Crowdfunding on Student Performance. *SSRN Electronic Journal* . [Crossref]
3. W. Jill Fitzgerald, Jeff Elmore, Melody Kung, A. Jackson Stenner. 2017. The Conceptual Complexity of Vocabulary in Elementary-Grades Core Science Program Textbooks. *Reading Research Quarterly* **52**:4, 417-442. [Crossref]