ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-25

School District Operational Spending and Student Outcomes: Evidence from Tax Elections in Seven States

Carolyn Abott
St. John's University

Vladimir Kogan
The Ohio State University

St�ephane Lavertu
The Ohio State University

Zachary Peskowitz
Emory University

We use close tax elections to estimate the impact of school district funding increases on operational spending and education outcomes. The analysis indicates that districts where tax levies passed spent 3-5 percent more per pupil annually through 6-8 years after the election. This spending came in the form of higher salaries per employee—as opposed to more teachers or staff—and corresponds to positive achievement effects in districts with a high proportion of impoverished students. Specifically, among districts above the sample median in the proportion of students who qualify for free or reduced-price lunches, the results imply that spending an extra $700 per pupil annually for 6-8 years leads to achievement gains of approximately 0.06-0.08 standard deviations. We find no achievement effects in districts with relatively advantaged students, and there are no attainment effects regardless of district demographics.

VERSION: May 2019

School District Operational Spending and Student Outcomes: Evidence from Tax Elections in Seven States*

Carolyn Abott
Department of Government
St. John's University
abottc@stjohns.edu

Vladimir Kogan
Department of Political Science
The Ohio State University
kogan.18@osu.edu

Stéphane Lavertu†
John Glenn College of Public Affairs
The Ohio State University
lavertu.1@osu.edu

Zachary Peskowitz
Department of Political Science
Emory University
zachary.f.peskowitz@emory.edu

April 16, 2019

Abstract

We use close tax elections to estimate the impact of school district funding increases on operational spending and education outcomes. The analysis indicates that districts where tax levies passed spent 3-5 percent more per pupil annually through 6-8 years after the election. This spending came in the form of higher salaries per employee—as opposed to more teachers or staff—and corresponds to positive achievement effects in districts with a high proportion of impoverished students. Specifically, among districts above the sample median in the proportion of students who qualify for free or reduced-price lunches, the results imply that spending an extra $700 per pupil annually for 6-8 years leads to achievement gains of approximately 0.06-0.08 standard deviations. We find no achievement effects in districts with relatively advantaged students, and there are no attainment effects regardless of district demographics.

Keywords: student achievement, education finance, school districts, tax elections
JEL codes: H11, I28

*Note: We have no conflicts to declare. The Spencer Foundation provided generous funding for this study via a Lyle Spencer Research Award. We are very grateful to James Szewczyk and Matthew Troy for leading much of the data collection, and the 35 undergraduate research assistants who scanned and hand-entered data. We also thank Joe Smith from TexasISD.com, Jared Knowles, Chris Candelaria, Kenneth Shores, and hundreds of officials for providing us with data that made this analysis possible. Finally, we are grateful to those who provided feedback at the annual meeting of the Association for Education Finance and Policy.

†The authors are listed in alphabetical order. Please direct communications to the corresponding author at lavertu.1@osu.edu or by mail: John Glenn College of Public Affairs, The Ohio State University, 110 Page Hall, 1810 College Rd., Columbus, OH, 43210.

1 Introduction

Researchers have long debated the returns to education spending (e.g., see Hanushek, 2003). A wave of recent research, however, provides convincing quasi-experimental evidence that increases in school district funding can improve education outcomes (Jackson, 2018a). Notably, using plausibly exogenous shocks in district revenues from state finance reforms, Jackson et al. (2016), Lafortune et al. (2018), and Candelaria and Shores (2019) find that higher revenues are associated with large achievement and attainment gains—particularly among districts serving low-income students. But there may be diminishing returns to education spending. It remains an open question whether increasing spending would improve student outcomes today, after many finance reforms have dramatically increased spending per pupil and narrowed spending gaps between high- and low-income districts.[1] Likewise, the circumstances under which spending increases have an impact may have changed.

This study estimates the impact of recent increases in education spending across diverse districts in multiple states. Specifically, we use a regression discontinuity design to estimate the impact of local tax referenda on spending and student outcomes in approximately 800 districts across seven states: Arkansas, Louisiana, Michigan, Missouri, Pennsylvania, Texas, and Wisconsin.[2] We focus on district elections held between 2002 and 2008 because they enable us to observe district spending, student achievement, and attainment seven years later, which is often how long it takes to detect the effects of spending increases (e.g., see Lafortune et al., 2018; Candelaria and Shores, 2019).[3] Unlike local bond elections tied to capital spending and restricted state and federal grants, these local tax referenda allow districts to allocate revenues to whatever operational functions they wish. Besides Kogan

[1] Although more recent studies have documented the immediate achievement impact of the Great Recession (Jackson et al., 2018), sudden cuts in spending can have an outsized impact compared to spending increases because of the disruption they can introduce (e.g., see Lavertu and St. Clair, 2018). Thus, the Great Recession may not be ideal for identifying the likely impact of increases in educational spending.

[2] Although we employ data on over 1,200 districts, the effective sample of tax levies typically includes just over 800 unique districts in which tax referenda were within 20 percentage points of passage.

[3] Studies that look at the impact of unanticipated budget cuts, however, often find immediate effects (e.g., see Jackson et al. 2018; Kogan et al., 2017; Lavertu and St. Clair, 2018).

et al. (2017), the few other studies to leverage close elections in a regression-discontinuity framework have focused on bond referenda tied to capital spending in a single state and have found negligible impacts on achievement (e.g., see Cellini et al., 2010 and Martorell et al., 2016; but see Hong and Zimmer, 2016).[4]

Our analysis reveals that, on average, districts in which referenda passed had spent \$300-\$500 more annually per student after seven years. They spent this money on higher salaries per employee (over half of which are teachers) as opposed to employing more teachers or support staff. After a levy passed, district spending on salaries increased steadily with a peak of around \$2,500 more per employee after 6-7 years. On average, across these seven post-election years, districts spent approximately \$1,500 more per employee annually—capturing over half of the increase in operational spending.[5] This average increase in spending generally is not associated with statistically significant changes in student achievement, graduation rates, or our proxy for dropout rates (which compares 12th grade enrollments to 8th grade enrollments four years prior), although achievement estimates often approach common benchmarks for substantive significance.

Further analysis reveals statistically and substantively significant achievement effects in districts that are above the median in the proportion of students eligible for free or reduced-price lunches (FRL). In these districts, levy passage is associated with an average increase in spending of up to \$700 per pupil annually through 6-8 years after the election. The corresponding achievement gains are around 0.06-0.08 standard deviations and often attain conventional levels of statistical significance. These achievement effects are on the lower end of Lafortune et al.'s (2018) estimates and are near the 50th percentile among ed-

[4]Our own analysis of bond referenda across 10 states—including those we list above—indicate large but short-lived effects of bond referenda on capital expenditures. Like the single-state studies we cite, they offer little opportunity to identify the impact of capital investments on long-term student outcomes. Of course, that does not mean that that capital projects do not have an impact on achievement (e.g., see Conlin and Thompson, 2017). Indeed, analyses of adequacy reforms also primarily involve increases in capital expenditures (e.g., see Lafortune et al., 2018).

[5]Salaries include gross payments to employees, including temporary staff (e.g., substitute teachers) and payments for additional duties (e.g., coaching). The remaining operational expenditures (net of salaries) likely capture supplies and purchased services we cannot observe.

ucation interventions in terms of their size and cost (see Kraft, 2018).[6] On the other hand, achievement effects among districts below the median in FRL-eligible students are very close to zero.

Unlike Candelaria and Shores (2019), we do not find significant attainment effects—but we have data limitations that prevent us from drawing firm conclusions. Our evidence on possible mechanisms is also merely suggestive. We observe salary increases that capture over half of increases in operational expenditures. There is some evidence of lower student-librarian ratios, but no significant changes in staffing levels overall—whether or not we limit the analytic sample to districts from states with strong collective bargaining laws. The results align with those of Brunner et al. (2018), who find that revenue increases allocated to higher salaries may nonetheless correspond to higher student achievement. That said, we do not know if the higher salaries are due to hiring more expensive staff (e.g., replacing substitute teachers with credentialed teachers), as opposed to providing raises to current employees. The results also suggest benefits may account for some of the remaining increases in spending (net of salaries), but our estimates are imprecise. Finally, although we cannot observe the supplies or purchased services that may account for the remainder of operational spending net of salaries and benefits, we can rule out spending on construction, equipment, property, and debt service.

Overall, this study provides convincing evidence that spending increases can have a significant impact on districts serving impoverished students. To the extent that these results are generalizable to districts outside of our effective analytic sample, we corroborate recent claims that school finance policy could realize gains by targeting funding to districts serving impoverished students—perhaps without the spending restrictions that accompany some state and federal programs (see Jackson 2018a). In the following sections, we describe our data and empirical strategy, review the results, and offer some concluding thoughts.

[6] If one monetizes the returns to achievement based on increases in students' lifetime earnings, our estimates suggest a benefit-cost ratio just below 1 (see Lafortune et al., 2018, p. 24).

2 Data

First, we describe our tax election data—an original contribution of this study. Second, we describe the publicly available data we use to examine school district finances, staffing, and student composition, achievement, and attainment. Third, through a series of tables and figures we provide a preliminary comparison of spending and other characteristics between districts with passing and failing tax referenda.

2.1 Tax Election Data

We collected information on school district tax and bond referenda held from 2002 to 2015 across 17 states.[7] For the purposes of this analysis, we limit ourselves to tax referenda rather than bond referenda, as the outcomes of bond elections are not predictive of long-term district expenditures.[8] Additionally, because district-level test score data that are comparable across states are available only for school years 2009-2015—and because research indicates that it can take years of increased spending to measurably affect student achievement and attainment—we focus the analysis on tax referenda for which we observe achievement outcomes at least 7 years after the election (i.e., tax elections held in 2008 or earlier). Finally, because public records across many counties make it difficult to determine vote totals in districts that cross county lines, we purged the sample of referenda for which this is a problem.[9] Ultimately, we

[7]States include Arkansas, California, Illinois, Iowa, Louisiana, Maryland, Michigan, Missouri, Nebraska, New York, Ohio, Oklahoma, Pennsylvania, South Carolina, Texas, Virginia, and Wisconsin. We omit Ohio and Oklahoma referenda from the analysis because they are not predictive of long-term operational spending. In Ohio, districts frequently put measures on the ballot simply to maintain funding because many referenda are temporary and all property tax rates are adjusted downward as property values increase. In Oklahoma, various state requirements effectively render levy votes meaningless, as districts have no choice but to approve the maximum allowable rates (OSDE, 2017, p. 2). We omit referenda from Nebraska, Maryland, South Carolina, and Virginia because there are few observations. We omit those from California, Iowa, and New York because referenda from these states are primarily predictive of capital, as opposed to operational, spending. Finally, we omit Illinois because we lack vote totals for the tax measures.

[8]On average across all states, bond elections are predictive of short-term capital spending, but they are not predictive of long-term spending of any type.

[9]Specifically, because our records requests in Arkansas, Michigan, Missouri, and Pennsylvania were at the county level and some districts span county boundaries, we removed observations from districts that span boundaries because it was unclear whether a given county's totals reported all votes across the district or only those votes from the portion of the district in that county. Additionally, numerous election results in Arkansas indicated one voter or some other implausibly low number of voters. Our follow-up communications

ended up with a sample of 1,900 tax elections held from 2002 to 2008 across seven states: Arkansas, Louisiana, Michigan, Missouri, Pennsylvania, Texas, and Wisconsin. (See Table A1 in Appendix A for a brief description of local tax elections in these states.)

[Insert Table 1 about here.]

As Table 1 reveals, there are some differences across states in the number of tax referenda, as well as the proportion of referenda that passed. Most notably, there are many referenda from Pennsylvania and the passage rate is exceptionally low. That is because in 2006 the state sought to provide homeowners with property tax relief and gave districts the option of shifting some of the tax burden from local property taxes to local income taxes to offset losses. All but three of 501 Pennsylvania districts placed such a measure on the ballot in May of 2007, which voters in nearly all districts rejected.[10]

Most of the analysis below restricts the sample to a neighborhood near the passage threshold. Thus, across many of the models we estimate, the effective sample size is closer to 1,000 referenda across 800 districts. Table 1 presents descriptive statistics for the full sample of 1,900 and, in brackets, descriptive statistics for the effective sample within 20 percentage points of the passage threshold—which is a bandwidth that we feature in the analysis. As Table 1 reveals, restricting the sample in this way changes the distribution of observations across states, and in most states the passage rate begins to approach a 50 percent passage threshold.[11] Restricting the bandwidth further leads average passage rates to quickly approach 50 percent across all states (i.e., as our density tests below confirm, near the threshold we have about as many passing and failing referenda). Finally, it is worth

with election administrators in that state indicate that sometimes tax elections were not held and a vote of 100 percent approval was logged to indicate that a district levied a tax. This practice may be related to votes being inconsequential because the tax rate on the ballot remained the same for that year (see Table A1 in Appendix A). To address this issue, we removed tax measures from the sample if the vote total was under 1 percent of the voting age population. Because we focus our analysis in a neighborhood around the cutoff, however, any remaining problematic observations of this kind should not affect the main analysis.

[10]These 2007 referenda from Pennsylvania appear to be revenue neutral. Thus, the handful of passing referenda may not have positive impacts on district revenue—even during the Great Recession. However, the results are similar whether or not we include these observations in the analysis.

[11]The vote threshold is 2/3 for a couple of tax measures in Missouri, where a tax levy would result in an effective tax rate that exceeds 6 percent (see Table A1 in Appendix A).

noting that we lack data from Texas prior to 2006, as there were no tax ratification elections prior to 2006.

2.2 School District Data

The school district data are from publicly available sources. We obtained school district finance, staffing, and student data from the National Center for Education Statistics' Common Core of Data (CCD). The achievement data are from the Stanford Education Data Archive (see Reardon et al., 2018). In particular, we obtained SEDA's 2009-2015 student achievement data for grades 3-8 in mathematics and English/Language Arts (ELA) standardized at the cohort-subject-grade level. In the analysis below, we use a single district-level average of these standardized scores across subjects and grades. Because we observe outcomes 6-8 years after a tax election, focusing on grades 3-8 means we observe outcomes for students who experienced greater spending for all (or nearly all) years of formal schooling.[12]

[Insert Table 2 about here.]

For each tax referendum, we merged in data for the corresponding district from 3 years prior to the election to 9 years after the election. Table 2 presents descriptive statistics for select CCD variables in the year preceding the corresponding tax election. The SEDA standardized test scores (averaged across math and ELA), however, are observed 7 years after the election, as we do not observe test scores prior to the elections. It is worth noting that these post-election achievement variables are highly correlated with the pre-election percentage of district students who qualify for free or reduced-price lunches (FRL). This correlation is evident in Table 2, as states in which districts have the lowest percentage of FRL students (Pennsylvania and Wisconsin) have districts with the highest average student test scores—placing them around 0.2 standard deviations above average—whereas states in which districts have the highest percentage of FRL students (Arkansas and Louisiana) have the lowest test scores—placing them closer to 0.2 standard deviations below average. The

[12] Another advantage of focusing on grade 3-8 achievement data is that some districts in our sample (e.g., some in Missouri) do not include high school grades.

relationship between achievement and spending per pupil is analogous.

Overall, Table 2 reveals that the 1,075 unique districts in our sample spend roughly around the national average per pupil ($12,228 in 2015 dollars[13]) and that we observe somewhat more districts with at least one failing referendum (699) as opposed to districts with at least one passing referendum (512)—even though, as Table 1 indicates, we have more passing referenda than failing referenda in the sample. It is also worth noting that the average referendum passage rates are higher in districts that spend less per pupil, have low-achieving students, have higher enrollments, and are more rural. This is because states that contribute a large number of observations (e.g., Pennsylvania and Wisconsin) had high failure rates. Overall, the table illustrates the diversity of districts in our sample. Because we pool across such diverse states that contribute very different numbers of measures, it is particularly important that we focus on districts very close to the passage threshold and that we reduce the influence of outliers by scaling our variables appropriately. As we discuss below, that is why we scale nearly all variables by student or staff counts and then log them. It is also why we also present the results of models with and without precision weights for SEDA district-level test scores normalized by subject, grade, and cohort.[14]

2.3 Preliminary Look at Changes in Spending and Attainment

Our focus is on operational expenditures, as opposed to capital outlays. These operational expenditures are what the National Center for Education Statistics labels "current expenditures." They include gross salaries and benefits for all district employees—including temporary employees (e.g., substitute teachers)—as well as supplies and purchased services. Figure

[13]We adjusted for inflation using the Consumer Price Index (CPI) published by the U.S. Department of Labor, Bureau of Labor Statistics. For comparability to fiscal education data, we follow the National Center for Education Statistic's practice of using a CPI converted from a calendar year to a school fiscal year basis (July through June). See Digest of Education Statistics 2015, Table 106.70, `http://nces.ed.gov/programs/digest/d15/tables/dt15_106.70.asp`, retreived June 17, 2018.

[14]Reardon et al. (2018) recommend the use of weights to account for variability in the precision of district test score estimates. The standard errors of these estimates do not differ significantly between passing and failing referenda near the passage threshold. Nevertheless, errors are highly correlated with raw counts of district students (over 0.90), so including precision weights based on the squared inverse of these errors could put more weight on big districts away and to the right of the cutoff.

1 presents differences in the natural log of a district's operational expenditures per pupil across four separate years (3 years prior to the election, the year of, 3 years after, and 7 years after) as compared to that district's pre-election baseline (the year prior to the election). Specifically, the y axis captures within-district changes in logged per pupil expenditures and the x axis captures the percent of the vote in favor of passage—centered at the vote threshold necessary for passage. Thus, the figure roughly compares the percent change in operating revenues before and after the election between districts in which referenda passed and those in which referenda failed. We generated the figure using Calonico, Cattaneo, Farrell, and Titiunik's (2017) procedure for calculating bin sizes and fitting the global polynomial, which in this case is cubic. We also restrict the figure to referenda within 20 percentage points of passage, as this is a sample we turn to often in the analysis below.

[Insert Figure 1 about here.]

As Figure 1 illustrates, although there is no discontinuity at the vote threshold the year of the election, by year 7 there is a clear discontinuity of around 0.05. That implies that spending is approximately 5 percent higher in districts that passed referenda 7 years prior, as compared to those that did not. The figures also reveal that this difference comes from greater increases in spending. On average, districts at the vote threshold whose referenda failed increased inflation-adjusted operational spending by just over 5 percent since the election. Districts in which referenda passed, however, increased inflation-adjusted operational spending by about 10 percent since the election.

Figure 2 presents these results over time in terms of annual operating expenditures per pupil for referenda within 20 percentage points of passing (as in Figure 1). Specifically, it presents estimates from OLS models that compare changes in annual expenditures (since the year prior to the election) between districts that did and did not pass their referenda for each year relative to the election—both for the overall sample (Figure 2a) and for districts in Pennsylvania, Michigan, and Wisconsin (Figure 2b). (In the analysis below we often estimate models separately for these states, which have strong collective-bargaining laws.) Specifically,

Figure 2 presents the results of the following model:

$$Y_{ikt} = \tau_k(Pass_i \times \gamma_k) + f(Pass_i, Vote_i, \gamma_k) + \gamma_k + \pi_i + \lambda_t + \epsilon_{ikt} \tag{1}$$

Y captures annual operational expenditures per pupil in the district associated with tax referendum i, in school year t, and year-relative-to-the-election k; $Pass$ indicates whether measure i passed; $Vote$ is the percentage of the vote in favor of passage, centered at the passage threshold; $f(Pass_i, Vote_i, \gamma_k)$ is a quadratic polynomial of the centered vote featuring interactions with the passage indicator and an indicator for each relative year k; and γ, π, and λ capture relative-year, referendum, and school-year fixed effects, respectively. We omit the variable capturing the interaction between passage and the relative-year fixed effect for the year prior to the election ($k = -1$), so that τ_k captures the effects of passage for each pre- and post-election year as compared to that pre-election baseline (as in Figure 1). Standard errors are clustered by school district, and dashed lines in Figure 2 capture 95 percent confidence intervals.

[Insert Figure 2 about here.]

Figure 2a reveals that by year 5, annual operating expenditures are approximately $500 higher per pupil in districts with passing referenda, and that this difference in spending persists nine years after the election. This fact is important, as we are essentially estimating an “intent to treat” effect—that is, any district in which a referendum fails can put a ballot on the measure in subsequent periods. That the spending differences persist indicates that this is not a problem for our analysis. Figure 2b reveals similar estimates for districts in states with strong collective-bargaining laws, though the increase in spending occurs almost immediately, one year after the election. It is also worth noting the small spending increase in the year the election was held. This is to be expected because referenda that passed in July or August, for example, are coded as occurring during the upcoming school year.

We conducted a similar analysis for our attainment variables because we observe them prior to the elections (unlike student achievement, which we do not observe until

2009). First, we calculated graduation rates by dividing the number of diplomas a district awarded in a given year by 8th grade enrollments four years prior.[15] Unfortunately, the graduation rate variable is available only through 2011 and contains many missing values. Thus, to complement this measure, we constructed a proxy for the dropout rate by dividing a district's 12th grade enrollment by its 8th grade enrollment four years prior and subtracting this quantity from 1.[16]

[Insert Figure 3 about here.]

The correlation between graduation rates and our dropout rate is around -0.75 in post-election years. As the figure reveals, the two measures track one another quite well in early post-election years, when they are available for exactly the same districts. In particular, it appears that levy passage leads to an immediate increase in dropout rates and an immediate decline in graduation rates, before the impact on these rates returns to 0 by the third post-election year.[17] However, this short-term result appears attributable to random noise in the data, as our analysis below does not reveal such effects. The figures also reveal that the results begin to diverge after year 3, which is when the number of tax referenda for which we have graduation rate data begins to decline significantly. In the analysis below, we estimate attainment effects using both measures when possible, but we turn to the dropout measure for heterogeneity analyses in which we must maximize sample sizes.

2.4 Primary Outcome Variables

The analysis below examines how sustained increases in district spending translate to salaries, staffing levels, and student outcomes through seven post-election years. That is, because

[15] We follow Heckman and Lafontaine (2010) and use 8th grade enrollments four years prior to avoid downward bias caused by retention in 9th grade. We also code as missing when there are twice as many diplomas as 8th grade students four years prior. We thank Candelaria and Shores (2019) for sharing their enrollment calculations with us, as we could not access the CCD school-grade enrollment data at the time we conducted this analysis.

[16] Actual dropout data are unavailable for our key post-election years (2009-2015). Similar to our coding of graduation rates, we convert to missing observations in which 12th grade enrollments are over twice as large as 8th grade enrollments four years prior.

[17] These results attain conventional levels of statistical significance if we use all three pre-election years as a baseline.

we are concerned with the cumulative effects of increased spending, we are concerned with average increases in inputs across all seven post-election years combined. In some additional analyses, we also pool observations of average inputs through years 6-8 to correspond to the analysis of achievement levels in those years, as pooling helps us deal with noisy estimates of achievement. We do not include the ninth post-election year because, as Figure 2 above indicates, estimates of spending become too noisy as the sample size shrinks.

Specifically, for each tax measure i, the analysis focuses on the average of expenditure or staffing levels (Y) from post-election years $k = 0$ through $k = M$, where $M \in \{6, 7, 8\}$:

$$Z_i^M = \frac{1}{M+1} \sum_{k=0}^{M} Y_i^k \tag{2}$$

We focus first on the analysis of expenditures and staffing through year 7. In these models, we use the average of annual expenditures and staffing for all post-election years through year 7 (i.e., $Z_i^7 = \frac{1}{8}\sum_{k=0}^{7} Y_i^k$).[18] That captures how much average annual spending and staffing students were exposed to through year 7. We also estimate models that pool post-election years 6 through 8—that is, we pool all non-missing values of Z_i^6, Z_i^7, and Z_i^8 for each tax referendum i. Effectively, these models weight observations more heavily from districts for which we observe outcomes in years 6 or 8 after the election than those for which we observe outcomes only in the 7th post-election year.[19]

Unlike the expenditure and staffing variables, the district-cohort achievement estimates and dropout and graduation rates already capture the accumulated impact of greater resources since the election. In particular, because SEDA achievement estimates are based on test scores in grades 3-8, focusing on post-election years 6-8 means that we are focusing on achievement effects among students who experienced differential annual expenditures levels

[18]We include the school year in which elections occur ($k = 0$) because districts with passing or failing levies early in that school year might adjust expenditures or staffing before the school year is over, as Figure 2 indicates. Thus, we scale the measure by the number of post-election years plus 1 ($M + 1$).

[19]Recall that our sample is limited to referenda for which we have achievement data seven years after the election. Although we observe all other variables in years 6 and 8 for these referenda, there are some referenda for which we do not observe achievement (or other variables) in years 6 and 8.

across all or nearly all years of their formal schooling.

[Insert Table 3 about here.]

Table 3 provides a descriptive look at our primary outcome variables in post-election year 7 (school years 2009-2015). The top panel features variables that capture an annual average of spending or staffing from the year of the election *through* seven post-election years. The bottom panel features variables that capture student characteristics and education outcomes *in* the seventh post-election year. The number of districts in the analytic sample is just over 800 for finance and achievement variables, as there are districts for which we observe multiple tax referenda. Missing values of the dropout rate are due primarily to missing or implausible by-grade student counts. The graduation rate data are sparse primarily because we do not observe diploma counts after 2011. We exclude (i.e., code as missing) observations of dropout rates in which either 12th grade enrollments are twice as large as 8th grade enrollments, as well as observations of graduation rates for which diploma counts are over twice as large as 8th grade enrollments four years prior. Removing these outliers yields roughly normal distributions on both measures—in part because there remain implausible values of both measures, as graduation rates can exceed 1 and dropout rates are sometimes negative. We do not top-code graduation rates or bottom-code dropout rates to avoid introducing measurement error in districts that experience enrollment increases or declines.[20]

3 Empirical Strategy

We use a regression discontinuity design to estimate the impact of tax levies on school district spending and education outcomes. First, we estimate covariate-adjusted local linear regression models of the following form:

$$Y_{ik} = \alpha + \tau Pass_i + f(Pass_i, Vote_i) + \beta_1 Y_i^{k=-1} + \beta_2 Y_i^{k=-3} + \epsilon_{ik} \tag{3}$$

[20]Candelaria and Shores (2019) top-code graduation rate data by coding as 1 any rates that exceed 1.

The outcome Y for tax measure i in post-election year $k \in \{6, 7, 8\}$ is a function of $Pass_i$, indicating whether (1) or not (0) a tax measure passed; $f(Pass_i, Vote_i)$, which is a linear function of the running variable, $Vote_i$, with and without an interaction with $Pass_i$; and the outcome observed the year before the election ($Y_i^{k=-1}$) and three years before the election ($Y_i^{k=-3}$). We include as covariates these pre-election values of the dependent variables to increase the precision of our estimates. Because we cannot observe student achievement prior to districts' 2002-2008 tax elections, models in which achievement is the outcome of interest include baseline covariates capturing student demographics known to be correlated with achievement: the fraction of students who are black, the fraction of students who are Hispanic, and the fraction of students eligible for free or reduced-price lunches.[21] Indeed, these baseline demographic variables are highly correlated with achievement in year 7, with correlations in the neighborhood of 0.7.

We employ Calonico, Cattaneo, Farrell, and Titiunik's (2017) "rdrobust" package to estimate these regressions within a restricted bandwidth such that $Vote_i \in [-h, h]$. The bandwidth, h, is Calonico, Cattaneo, and Titiunik's (CCT, 2014) mean-squared-error (MSE) optimal bandwidth. We employ a triangular kernel such that observations near the passage threshold are weighted more heavily, and we report robust standard errors clustered at the district level.

We also report the results of OLS models that focus on referenda within 20 percentage points of passage, which is around the maximum bandwidth selected using CCT's optimal bandwidth selector. We do so to examine the sensitivity of the primary estimates to different model specifications; to compare estimated effects for the same set of measures across all dependent variables; to increase precision with the inclusion of more pre-treatment covariates; and to examine effect heterogeneity using interactions with the treatment indicator. Specifically, we estimate the following OLS model:

[21] Due to missing values of these variables in early years, we include them only for the year prior to the election ($k = -1$).

$$Y_{ikt} = \tau Pass_i + f(Pass_i, Vote_i) + \beta_1 Y_i^{k=-1} + \beta_2 Y_i^{k=-3} + \lambda_t + \mathbf{X}_i' \gamma + \epsilon_{ikt} \tag{4}$$

The outcome Y is for tax measure i, school year t, and year-relative-to-the-election k; λ captures school-year fixed effects; and $\mathbf{X}$ captures a variety of baseline covariates from years prior to the election, including logged operational expenditures per student one and three years prior to the election; fractions of FRL, black, and Hispanic students in the year prior to the election; and logged student counts one and three years before the election. In contrast to equation 3, this model features a quadratic polynomial of the centered vote share variable (as in equation 1), which we also interact with the passage indicator.[22] Once again, we also cluster standard errors by school district.

The parameter τ in both models above is equivalent to the difference in intercepts from two separate regressions estimated using observations on each side of the cutoff. We interpret this parameter as the causal impact of passing a tax referendum for districts at the vote threshold necessary for passage. This interpretation requires the identifying assumption that potential outcomes are continuous through the vote threshold for levy passage. As is common in the literature, we examine the plausibility that this assumption holds by testing for pre-treatment imbalances in covariates, as well as testing for differences in the density of the running variable on either side of the cutoff.

First, we tested whether there are any imbalances in the levels of observed district characteristics one year prior to the election, as well as whether there are differences in trends in the years leading up to the tax election (specifically, a difference between levels one year before the election and three years before). The results in Table B1 in Appendix B compare pre-election levels and trends using the local linear models with CCT's bandwidth. The first

[22] We do not estimate these models with a linear specification of the running variable because of the wide bandwidth. To replicate the results we get from this quadratic specification using a linear specification, one must generally use a bandwidth half as large (within 10 percentage points of passage) as in the first set of models we estimate. But we need the larger sample from the wider bandwidth to conduct much of he analysis below.

two columns focus on the full sample of referenda and indicate six statistically significant differences across levels and trends of 27 separate variables (50 tests in total), although some of these variables capture related quantities (e.g., salaries and benefits, and student-teacher and student-staff ratios). As columns 3 and 4 reveal, some of these imbalances disappear if we restrict the analysis to states with strong collective-bargaining laws (Michigan, Pennsylvania, and Wisconsin), but a couple of other imbalances appear in baseline graduation rates and operational expenditures. (Although we separate out this sample for substantive reasons, it appears that comparing results between these samples could be a good sensitivity check.) The results in Table B2 indicate a similar number of imbalances when we estimate OLS models using a quadratic specification of the running variable and a consistent bandwidth of 20 percentage points across all variables. Overall, there appears to be relative balance between the samples, and our inclusion of baseline covariates should also account for noise that may have led to the few imbalances we detect.[23]

Second, we tested for the plausibility of the continuity of covariates through the passage threshold via a test of the density of the running variable, as per McCrary (2008). As the figure in Appendix C reveals, we are unable to reject the null of a continuous density for the full sample and for the subsample limited to strong collective-bargaining states. However, there appears to be some bunching to the right of the passage threshold for non-CB states. The results are generally similar if we conduct the density test for each state individually. The one exception is that in Michigan there appear to be more failing than passing referenda near the pass threshold (p=0.0531).[24] Once again, these tests suggest that comparing results between the full sample and the CB sample could serve as a good sensitivity analysis.

[23]Because the covariate balance tests generally yield results consistent with the notion that the pre-election covariates we include in the regressions are continuous across the vote threshold needed for passage, including these covariates should introduce no assumptions about the functional form of the underlying regression function (Calonico, Cattaneo, Farrell and Titiunik, 2018). However, if these imbalances indeed are evidence of a violation of RD assumptions, our including baseline covariates means that the causal interpretation of our estimates is dependent on conditional independence assumptions.

[24]Hong and Zimmer (2016) find a greater density on the right side of the threshold for Michigan bond referenda. Indeed, if we combine bond and tax referenda, the density of the running variable appears almost perfectly smooth across the threshold.

4 Results

First, we present the results of models examining the impact of referendum passage on average annual expenditures and staffing *through* the 7th post-election year (Table 4) and student composition, achievement, and attainment *in* the 7th post-election year (Table 5). We then compare effects in high- and low-poverty districts in the 7th post-election year (Table 6), years 6-8 combined (Table 7), and across all available post-election years (Table 8). Finally, we summarize the results of other analyses we conducted.

4.1 Overall Effects

Table 4 presents the results of models estimating the impact of tax levy passage on average annual expenditures and staffing through year 7. We present the results of the first model (equation 3) using Calonico, Cattaneo, and Titiunik's local RD estimator (CCT) and the OLS model (equation 4) using a quadratic polynomial interacted with the passage indicator and using a bandwidth of 20 percentage points across all models. The CCT models include two baseline covariates capturing the dependent variable one and three years prior to the election, respectively. The OLS models also feature these baseline variables, as well as year fixed effects; logged operational expenditures per student one and three years prior to the election; fractions of FRL, black, and Hispanic students in the year prior to the election; and logged student counts one and three years before the election. For each model, we report the results based on district referenda across all states, as well as referenda from the three states with strong collective-bargaining laws (CB).

[Insert Table 4 about here.]

The results in Table 4 reveal that, seven years later, districts in which referenda passed had spent an average of 3-4 percent more per pupil on operations (employees, supplies, and purchased services), but there was no difference in terms of capital expenditures. Half of operational expenditures are classified as instructional expenditures (salaries, supplies, and purchased services related to the interaction of students and teachers), and these expenditures also increased 3-4 percent in three of the four models. These increases in operational expenditures correspond to a 3-4 percent increase in salaries of permanent and temporary employees (including substitute teachers), as opposed to increases in the number of teachers or other employees. The estimates are somewhat larger and more consistent across specifications if we estimate models that pool observations across post-election years 6-8 (see Table D1 in Appendix D). The results that remain somewhat inconsistent across models are for estimates of unlogged expenditures per pupil and salaries per staff. This is likely because of influential outliers, as the models with logged versions of these variables yield results that are relatively consistent across model specifications.

As Figure 4 indicates, salaries per employee grow steadily until they reach a peak of around $2,500 six years after the election. The figure reveals that this translates to just under $300 in salaries per student, which is over half of the $500 in per pupil operational expenditures by that year. Unfortunately, the data do not enable us to observe whether this is due to an increase in salaries of current employees or the employment of new, more expensive staff. For example, a district that relied heavily on substitute teachers may increase salaries per staff by employing full-time teachers. Such a change could increase student achievement significantly without affecting staff-student ratios.

[Insert Figure 4 about here.]

The results in Table 5 indicate no significant changes in district student demographics in the CB sample, which provides us with some confidence that any changes in achievement or attainment we observe are not due to changes in student composition. But the full sample indicates a 1 percentage point increase in the fraction of black students after the election. There are no statistically significant effects in terms of attainment, and only one achievement estimate reaches conventional levels of statistical significance. These results are similar if we pool post-election years 6-8 (see Table D2 in Appendix D).

[Insert Table 5 about here.]

Interestingly, models that exclude precision weights yield achievement estimates that are generally close to zero, but we cannot rule out substantively significant effects among the precision-weighted estimates. These estimates are often in the range of 0.04-0.05 standard deviations, which implies that increasing spending by $1,000 per pupil for seven years yields achievement effects of around 0.10 standard deviations. Among referenda from states with strong collective-bargaining laws, these coefficients are similar whether or not we include pre-treatment covariates; pool post-election years 6-8; or control for contemporaneous student demographics (see Table D3 in Appendix D). Thus, if weighting these measures is indeed important (see Reardon et al., 2018), there may be some substantively significant effects that we are unable to rule out.[25]

4.2 Effects by School District Poverty Level

Multiple quasi-experimental studies have found that impoverished students are those who benefit from increases in district spending. To explore this possibility, we estimate effects separately for districts that are and are not above the sample median in terms of the proportion of their students who are FRL-eligible. To maximize power, we estimate these disaggregated effects simultaneously by adding some interactions to the OLS model in equation 4.[26] Specifically, we estimate the following model:

[25]The analytic weight—which captures the squared inverse of the average standard error of test scores—has a mean of 141, a standard deviation of 161, a minimum of 14.5, and a maximum of 1,149. If we estimate our models with this measure as a dependent variable, the estimated impact of levy passage in year 7 is around 8 (p=0.642) using OLS with a quadratic polynomial, and it is 34 (p=0.281) using CCT's local estimator. These weights are very highly correlated with student counts; we are essentially weighting the models by district size if we employ them. Thus, it appears the differences in results from weighted and unweighted models may be due to heterogeneity in effects between large and small districts. However, estimated effects are similar for districts that are above and below the sample median in terms of district enrollments.

[26]Estimates are very imprecise if we estimate separate models for high-poverty districts. Additionally, as the results in Table D3 indicate, adjusting for baseline covariates to increase precision leads to volatile estimates when we use Calonico, Cattaneo, Farrell, and Titiunik's (2017) local estimator—and they become even more volatile if we limit the analysis to a subset of data.

$$Y_{ikt} = \tau^H HighPass_i + \tau^L LowPass_i + \sigma HighPov_i + f(HighPov_i, Pass_i, Vote_i) + \beta_1 Y_i^{k=-1} + \beta_2 Y_i^{k=-3} + \lambda_t + \mathbf{X}_i'\gamma + \epsilon_{ikt} \quad (5)$$

$HighPov_i$ indicates whether or not a district is above the sample median in the proportion of students who are FRL-eligible.[27] $HighPass_i$ indicates whether or not a referendum passed in a high-poverty district and $LowPass_i$ indicates whether or not a referendum passed in a low-poverty district. $f(HighPov_i, Pass_i, Vote_i)$ captures the polynomial from equation 4 (a quadratic specification of the centered vote interacted with the passage indicator) as well as a version of this quadratic polynomial interacted with the $HighPov_i$ district indicator. Thus, this model features a separate RD specification for high-poverty and low-poverty districts. All other covariates are as in equation 4, and we once again report standard errors clustered by district.

Table 6 and Table 7 present estimates of the impact of referendum passage on operational expenditures, achievement, and dropout rates 7 years and 6-8 years after passage, respectively. They do so for both the full sample (columns 1 and 2) and the sample from states with strong collective-bargaining laws (columns 3 and 4). Additionally, for each sample, tables present the results of models with and without controls for contemporaneous, post-election student characteristics (fractions of black, Hispanic, and FRL-eligible students) in order to examine sensitivity to potential changes in student composition. We have too few observations to estimate these models for districts with graduation rate data, so we report only the attainment results based on dropout rates.

[Insert Table 6 and Table 7 about here.]

Both tables reveal substantively significant achievement effects in high-poverty districts but not in low-poverty districts. Achievement estimates for high-poverty districts

[27] The median is 0.35 for the full sample and 0.28 for the collective-bargaining (CB) sample.

always approach or attain conventional levels of statistical significance, whereas none of the estimates for low-poverty districts approaches conventional levels of statistical significance. Among models that pool post-election years 6-8, achievement estimates among high-poverty districts are always between 0.06 to 0.09 standard deviations, and those for low-poverty districts are always between -0.02 to 0.02 standard deviations—whether or not models include precision weights and whether or not they are restricted to states with strong collective bargaining laws. Moreover, the inclusion of post-election student composition variables only strengthens the precision of these estimates, which supports the interpretation that these results reflect learning gains as opposed to changes in student composition.[28]

Unfortunately, we lack pre-election achievement data, so we are unable to examine whether these results may be due to pre-election imbalances in achievement. However, we do observe achievement in the first six post-election years for a subset of our sample and, thus, can examine whether there are achievement effects in early years among a subset of districts. The top panel (a) of Table 8 presents the results of models 2 and 4 from Table 6 for post-election years 1-2, 3-4, 5-6, and 7-8, respectively. We pool years in this way because we have few observations for early post-election years and to minimize the noise in the student achievement variable. The results indicate no statistically significant effects in years 1-4 among high-poverty districts. Among districts from collective-bargaining states (MI, PA, and WI), estimates in years 1-2 are very close to zero. In both samples, the point estimates in years 3-4 range from 0.02 to 0.06 standard deviations, and then climb to 0.09 standard deviations in post-election years 5-6. On the other hand, impacts on achievement in low-poverty districts are always statistically insignificant and point estimates are consistently close to zero.

[Insert Table 8 about here.]

We must interpret these results with caution, however. Although it appears that districts increase their effectiveness over time—that there are cumulative returns to spending—

[28]Indeed, as Table D4 in Appendix D reveals, passing tax referenda has no significant effects on student composition in high-poverty districts.

the results in panel (a) may be due to changes in the effective sample across columns 1-4. As a robustness check, in panel (b) we report the results of an analysis limited to referenda held in 2007 and 2008, for which we observe achievement in each pair of post-election years (i.e., across columns 1-4). Unfortunately, we do not detect achievement effects that increase over time using this subsample. As in the top panel, the sample of referenda from collective-bargaining states seems to display this property, but the estimates are imprecise.[29] That does not mean that this study's primary results are due to pre-treatment imbalances, of course. But we are unable to rule them out because we are unable to examine achievement effects close to election time for the entire set of high-poverty districts that account for our achievement effects in post-election years 6-8.

Overall, the resuts above indicate that increasing spending by approximately \$500-\$700 per pupil through 6-8 years leads to achievement gains of around 0.06-0.09 standard deviations for students in high-poverty districts, but the corresponding \$350 (or less) in spending increases in low-poverty districts is associated with no gains in achievement. On the other hand, Tables 6-7 indicate no significant effects of passing tax referenda on dropout rates. Indeed, the coefficients are often positive and substantively large. Thus, although our measure is imperfect and noisy, it seems we can rule out substantively significant attainment effects across all district types.

4.3 Additional Analyses

We conducted two additional heterogeneity analyses based on research that indicates minority students and rural communities might also benefit from higher spending. First, we estimated whether there were differences in effects between districts that are more or less racially diverse than the typical district. There were few significant results and no clear patterns, but that may be driven by our inability to conduct an analysis that focuses on districts with a high

[29] These estimates are also volatile and highly sensitive to model specification. But it is worth noting that we are unable to detect statistically significant achievement effects in post-election years 7-8 whenever we use a subset of referenda held in any pair of years (i.e., 2002/03, 2003/04, etc.). It appears we simply lack the sample size to conduct such an analysis.

fraction of Hispanic or black students due to small sample sizes. Second, we estimated models that explored effect heterogeneity across rural and non-rural districts. There is some suggestive evidence of achievement effects in rural districts, but these estimates are generally imprecise (see Table D5 and Table D6 in Appendix D).

5 Conclusion

This study provides evidence that, even after the large increase in K-12 spending associated with state finance reforms in the late 20th Century, additional investments can still have significant impacts on the achievement of disadvantaged students. Indeed, using a different design focused on more recent years, we obtain effect sizes that are very similar to Lafortune et al.'s (2018) difference-in-differences analysis of state finance reforms. Although they are at the lower end of their estimates, that we get similar effects with our design lends confidence to their important, nationwide findings. Moreover, although our results are based on a relatively small number of districts, the districts in our analyses to a large extent mirror the demographic diversity of the United States and are typical in terms of spending levels and student achievement and attainment. To the extent that our results are indeed generalizable, they suggest that policymakers could realize significant gains in student achievement by targeting unrestricted funding more directly to impoverished districts.

In addition to contributing to an emerging and rigorous evidence base on education spending and student outcomes, this study makes several contributions to the debate about the mechanisms that explain student outcomes. Unlike several prominent studies, the effects we detect are based purely on operational expenditures—mostly employee salaries, as opposed to smaller student-teacher ratios. We find no evidence that salary increases undermine education production (e.g., see Hoxby 1996), and, like Brunner et al. (2018), we find achievement effects in states with strong collective-bargaining laws. On the other hand, we are unable to decipher whether districts raised salaries of existing personnel, or whether they replaced them with higher-paid individuals. It could be that poorer districts that passed referenda suddenly were able to hire more effective full-time teachers, as opposed to relying on temporary options or less qualified individuals.

References

Brunner, Eric J., Joshua Hyman, and Andrew Ju. (2018), "School Finance Reforms, Teachers' Unions, and the Allocation of School Resources," Working paper.

Candelaria, Christopher A. and Kenneth A. Shores. (2019) "Court-Ordered Finance Reforms in The Adequacy Era: Heterogeneous Causal Effects and Sensitivity. Education Finance and Policy," *Education Finance and Policy* 14(1):31-60.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014), "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295-2326.

Calonico, S., Cattaneo, M. D., Farrell, M., and Titiunik, R. (2017), "rdrobust: Software for Regression Discontinuity Designs" *Stata Journal* 17(2): 372-404.

Calonico, S., Cattaneo, M. D., Farrell, M., and Titiunik, R. (2018), "Regression Discontinuity Designs Using Covariates," *Review of Economics and Statistics*.

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):215–261.

Conlin, Michael and Paul N. Thompson. (2017). "Impacts of new school facility construction: An analysis of a state-financed capital subsidy program in Ohio."*Economics of Education Review* 59:13–28.

Hanushek, Eric A.(2003). "The Failure of Input-based Schooling Policies," *The Economic Journal* 113(February):F64–F98.

Heckman, James J., and Paul A. LaFontaine. (2010). "The American High School Graduation Rate: Trends and Levels," *Review of Economics and Statistics* 92(2): 244–262.

Hong, Kai and Ron Zimmer. (2016). "Does Investing in School Capital Infrastructure Improve Student Achievement?" *Economics of Education Review* 53(8):143–158.

Hoxby, Carolyn Minter. (1996). "How Teachers' Unions Affect Education Production" *The Quarterly Journal of Economics* 111(3): 671–718.

Jackson, C. Kirabo. (2018a). "Does School Spending Matter? The New Literature on an Old Question," National Bureau of Economic Research, Cambridge, MA, Working Paper No. 25368.

Jackson, C. Kirabo. (2018b). "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes," *Journal of Political Economy* 126(5): 2072–2107.

Jackson, C. Kirabo, Cora Wigger, and Heyu Xiong. (2018). "Do School Spending Cuts Matter? Evidence from the Great Recession." Technical report, National Bureau of Economic Research, Cambridge, MA.

Jackson, C. Kirabo, Rucker C. Johnson, and Claudia Persico. (2016). "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms," *The Quarterly Journal of Economics*, 131(1):157–218.

Kogan, Vladimir, Stéphane Lavertu, and Zachary Peskowitz. (2017). "Direct Democracy and Administrative Disruption," *Journal of Public Administration Research and Theory* 27(3): 381-399.

Kraft, Matthew. (2018). "Interpreting Effect Sizes of Education Interventions," Brown University Working Paper. Downloaded Tuesday, April 16, 2019, from `https://scholar.harvard.edu/files/mkraft/files/kraft_2018_interpreting_effect_sizes.pdf`.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach. (2018). "School Finance Reform and the Distribution of Student Achievement," *American Economic Journal: Applied Economics* 10(2):1–26.

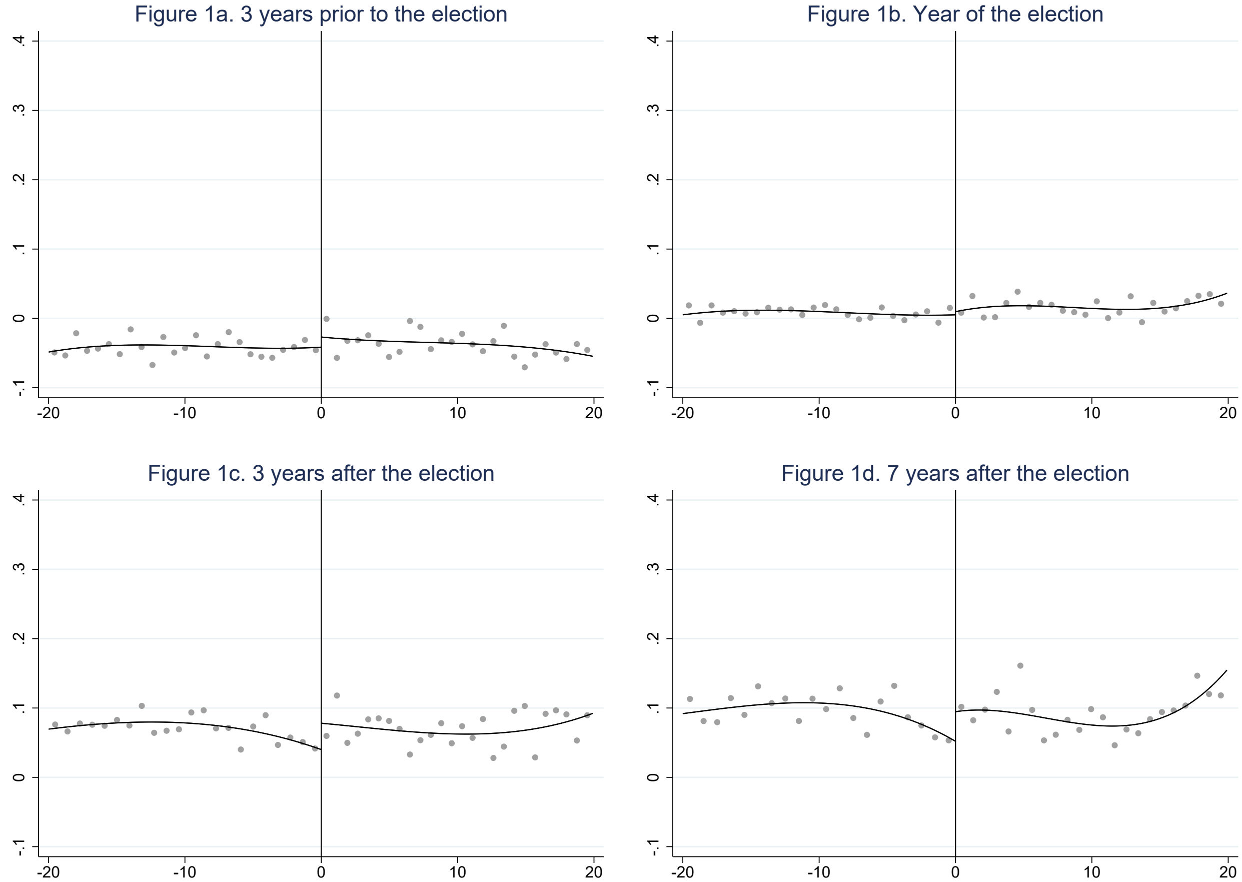
Lavertu, Stéphane and Travis St. Clair. (2018). "Beyond Spending Levels: Revenue Uncertainty and the Performance of Local Governments," *Journal of Urban Economics* 106: 59-80.

Martorell, Paco, Kevin Stange, and Isaac McFarlin. 2016. "Investing in Schools: Capital Spending, Facility Conditions, and Student Achievement," *Journal of Public Economics* 140: 13–29.

Oklahoma State Department of Education (OSDE). 2017. *Oklahoma School Finance Technical Assistant Document*. Financial Services Division. Downloaded April 2, 2019 from `https://sde.ok.gov/sites/ok.gov.sde/files/documents/files/`.

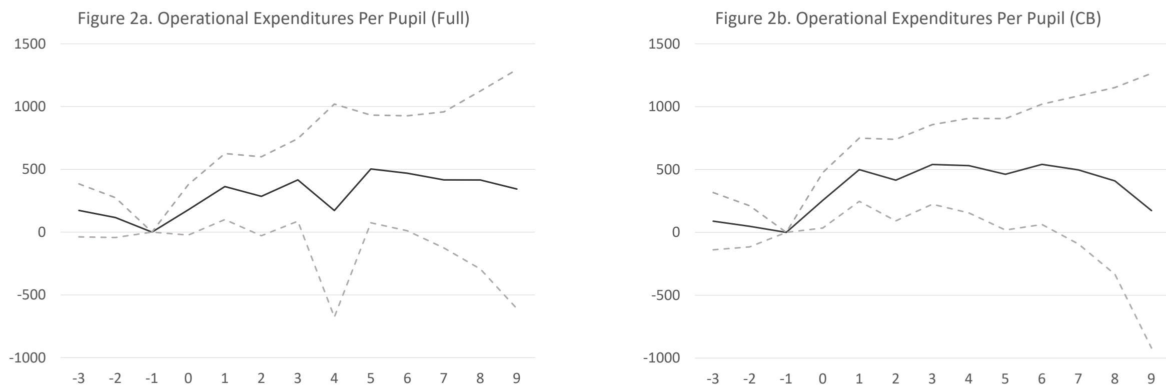
Reardon, S. F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., and DiSalvo, R. (2018). *Stanford Education Data Archive (Version 2.1)*. Retrieved from http://purl.stanford.edu/db586ns4974

Figure 1: Changes in Logged Expenditures Per Pupil



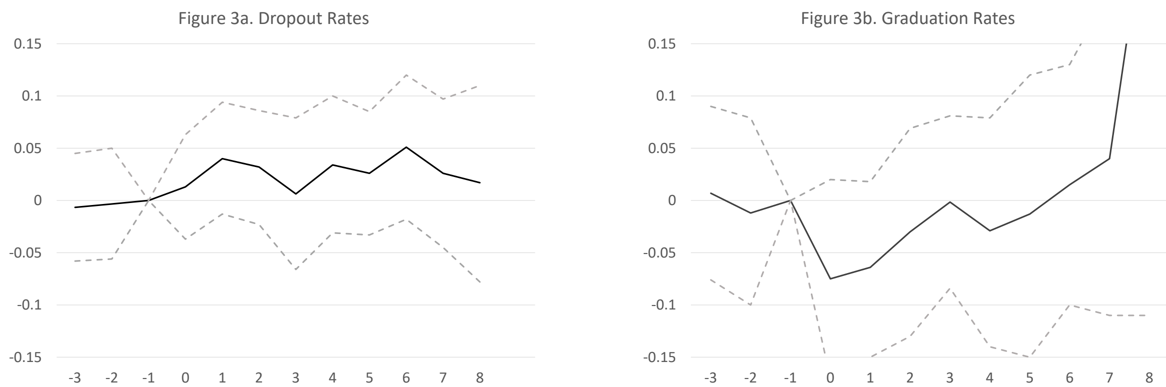
Note: The figure compares changes in the natural log of operational expenditures per pupil between districts in which tax referenda passed (vote margin > 0) and those in which referenda failed (vote margin < 0) 3 years prior to the election (a), the year of the election (b), 3 years after the election (c), and 7 years after the election (d). Specifically, the y axes capture the difference in logged expenditures (2015 dollars) between each of these years and the year before the election. The x axes capture the vote margin—the difference between the percent of votes in favor of passage and the threshold needed for passage (typically 50 percent). The dots are local means and the curves are fitted using a cubic polynomial. Figures are based on the full sample of states but limited to levies within 20 percentage points of passage.

Figure 2: Changes in Per Pupil Operating Expenditures (2015 dollars)



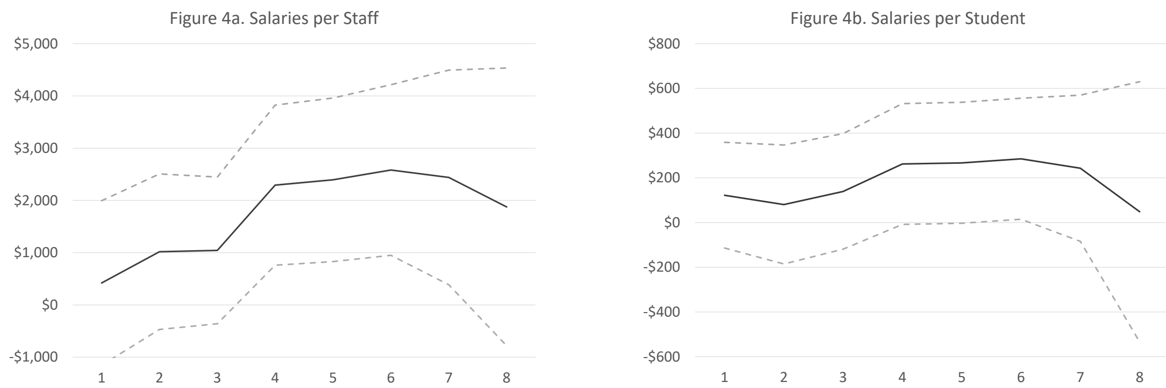
Note: The figures present the estimated impact of tax levies on changes in annual operational expenditures (2015 dollars) using the school year prior to the election as a baseline. The solid lines capture the estimated effects and the dashed lines capture 95 percent confidence intervals. The estimates are based on the sample of tax levies that came within 20 percentage points of passage. The OLS model used to generate these estimates includes district and school-year fixed effects as well as a quadratic polynomial to capture the relationship between the vote margin and educational expenditures. Figure 2a presents estimates for the full sample and Figure 2b presents estimates for "collective bargaining" (CB) sub-sample—Michigan, Pennsylvania, and Wisconsin.

Figure 3: Graduation and Dropout Rates



Note: The figures present the estimated impact of tax levies on changes in our proxy for dropout rates (1 minus the 12th grade enrollment scaled by 8th grade enrollment four years prior) and graduation rates (diplomas awarded scaled by 8th grade enrollments four years prior). The solid lines capture the estimated effects and the dashed lines capture 95 percent confidence intervals. The estimates are based on the sample of tax levies that came within 20 percentage points of passage. The OLS model used to generate these estimates includes district and school-year fixed effects as well as a linear polynomial to capture the relationship between the vote margin and educational expenditures.

Figure 4: Salaries per employee and per student



Note: The figures present the estimated impact of tax levies on changes in salaries per employee (1a) and salaries per student (1b). The solid lines capture the estimated effects and the dashed lines capture 95 percent confidence intervals. The estimates are based on the sample of tax levies that came within 20 percentage points of passage. The OLS models used to generate these estimates includes pre-treatment covariates and a quadratic polynomial, just like those in Table 4.

TABLE 1. Descriptive Statistics for Referenda with Year 7 Observations

	Referendum Count	Percent Passed	Mean Pct Yes Vote	Mean Vote Count	Election Years
Overall	1,900	52.37	52.69	2,405	2002 - 2008
	[1,236]	[46.52]	[49.16]	[2,754]	[2002-2008]
By State					
Arkansas	280	90.71	68.78	570	2002 - 2008
	[148]	[83.11]	[58.32]	[733]	[2002-2008]
Louisiana	183	84.15	66.23	4,228	2002 - 2008
	[100]	[74.00]	[57.43]	[5,272]	[2002-2008]
Michigan	324	85.49	67.40	1,079	2002 - 2008
	[169]	[74.56]	[57.94]	[1,400]	[2002-2008]
Missouri	63	60.32	51.06	5,327	2002 - 2008
	[58]	[60.34]	[50.62]	[5,667]	[2002-2008]
Pennsylvania	625	6.72	35.18	2,643	2002 - 2008
	[415]	[6.75]	[39.07]	[2,416]	[2002-2007]
Texas	117	79.49	62.07	622	2006 - 2007
	[78]	[74.36]	[55.36]	[735]	[2006-2007]
Wisconsin	308	44.48	46.84	3,981	2002 - 2008
	[268]	[48.88]	[48.98]	[4,263]	[2002-2008]

Note: The table provides descriptive statistics for referenda used in the estimation of the effects of expenditures on student achievement. Descriptive statistic in brackets focus on observations within 20 percentage points of the vote threshold for passage, as all analyses are based on effective samples within this range. All failing tax measures from Oklahoma are omitted because districts that proposed them spanned counties, which made it difficult to determine vote counts.

TABLE 2. Descriptive Statistics for Districts with Tax Levies on the Ballot

	District Count	Total Expnd. Per Pupil (2015$)	Oper. Expnd. Per Pupil (2015$)	Pct. Rural	Stdnt Count	Pct. White Stdnts	Pct. F/R Lunch Stdnts	School Count	Teach. FTE	SEDA Test Scores (SDs)
All Refs.	1,075	12,228	10,404	49.53	3,385	82.71	37.86	7.18	226	0.04
By Result										
Passed	512	11,113	9,705	54.57	3,741	77.62	44.32	8.36	251	-0.06
Failed	699	13,453	11,172	43.98	2,994	88.16	31.13	5.88	199	0.14
By State										
Arkansas	78	9.517	8,403	53.57	2,551	78.40	51.17	5.49	174	-0.13
Louisiana	55	9,917	9,017	49.18	9,765	55.74	63.06	20.92	670	-0.28
Michigan	122	11,947	10,210	59.26	2,066	87.23	36.10	5.49	124	-0.06
Missouri	52	11,346	9,222	44.44	3,792	86.63	32.47	7.37	255	0.14
Penn.	487	14,082	11,479	38.72	2,870	90.45	29.53	5.32	194	0.20
Texas	115	11,738	10,023	62.39	2,718	54.05	53.44	6.38	192	-0.07
Wisconsin	166	13,500	11,978	53.90	2,680	92.89	23.39	6.41	185	0.16

Note: The table provides statistics on the districts in which referenda were held. Nearly all statistics are based on averages across referenda in the year prior to an election. The test score data, however, are from the seventh year after referendum passage (school years 2009-2015). Districts that placed multiple referenda on the ballot between 2003 and 2008 figure more prominently in the averages. The mean standardized test scores in the final column are based on 2009-2015 data from the Stanford Education Data Archive (SEDA). All other data are from the NCES Common Core of Data Finance and Universe files. All dollar figures are reported in 2015 dollars, based on school-year (as opposed to calendar year) adjustments for inflation.

TABLE 3. Descriptive Stats for Key Outcome Variables ($k = 7$)

	N	Mean	SD	Min	Max
Averages Through Year 7					
ln(op. exp. per pupil)	1,236	9.31	0.16	8.92	9.99
Op. exp. per pupil	1,236	11,186	1,934	7,477	26,012
ln(salaries per staff)	1,236	10.77	0.18	10.27	11.34
Salaries per staff	1,236	48,835	9,758	28,861	174,256
ln(benefits per staff)	1,236	9.80	0.43	8.55	10.48
Benefits per staff	1,236	19,765	7,222	5,214	78,573
ln(students per staff)	1,236	2.02	0.16	1.50	2.54
Students per staff	1,236	7.75	1.42	4.48	29.80
ln(students per teacher)	1,236	2.66	0.14	2.06	3.16
Students per teacher	1,236	14.44	2.29	7.91	42.56
Values in Year 7					
Fraction FRL	1,228	0.48	0.20	0	1.01
Fraction Black	1,234	0.10	0.20	0	0.97
Fraction Hispanic	1,234	0.08	0.14	0	0.99
Avg. Math/Reading Scores	1,234	0.03	0.32	-1.24	1.04
Dropout Rate	1,030	0.07	0.18	-0.83	0.69
Graduation Rate	163	0.83	0.22	0.35	1.73

Note: These are descriptive statistics for select outcome variables in post-election year 7 ($k = 7$). The top panel is for variables that capture an annual average from the election through year 7. The bottom panel is for variables that capture the level in year 7. The number of districts in the analytic sample is 816, as there are districts for which we observe multiple measures. There are implausible values of dropout rates and graduate rates. Removing such values (e.g., if either enrollment or diplomas are missing) has minimal impact on the results.

TABLE 4. Average Annual Expenditures and Staffing through Year 7

	(1)	(2)	(3)	(4)
ln(op. exp. per pupil)	0.026*	0.028**	0.036***	0.045***
	(0.013)	(0.013)	(0.012)	(0.012)
op. exp. per pupil ($)	289*	300*	402***	526***
	(154)	(153)	(142)	(147)
ln(inst. exp. per pupil)	0.020	0.032**	0.030**	0.042***
	(0.014)	(0.014)	(0.013)	(0.013)
ln(cap. exp. per pupil)	-0.064	-0.096	-0.050	-0.12
	(0.13)	(0.17)	(0.13)	(0.15)
ln(salaries per staff)	0.030**	0.042***	0.032**	0.036**
	(0.014)	(0.016)	(0.013)	(0.016)
salaries per staff ($)	395	1,590	1,613**	2,163
	(1,073)	(1,139)	(813)	(1,347)
ln(benefits per staff)	0.014	0.022	0.024	0.011
	(0.023)	(0.022)	(0.019)	(0.020)
ln(students per teacher)	0.0044	0.0012	-0.0052	-0.016
	(0.013)	(0.015)	(0.012)	(0.014)
ln(students per staff)	0.0066	-0.0030	-0.0013	-0.0085
	(0.013)	(0.016)	(0.013)	(0.015)
States	All	CB	All	CB
Model	CCT	CCT	OLS	OLS
Specification	Linear	Linear	Quadratic	Quadratic

Note: Each coefficient is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. The dependent variables are averages of annual expenditure and staffing variables through 7 years after the election. The CCT models include baseline covariates of the dependent variable one and three years prior to the election. The OLS models include these baseline covariates, as well as year fixed effects, logged operational expenditures per pupil one and three years prior to the election, logged student counts one and three years prior to the election, and student demographics (fraction Black, Hispanic, and FRL) one year prior to the election. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 5. Student Demographics and Academic Outcomes in Year 7

	(1)	(2)	(3)	(4)
ln(student count)	0.022	0.047	0.012	0.011
	(0.028)	(0.034)	(0.026)	(0.029)
Fraction FR Lunch	-0.0098	-0.0042	-0.017	-0.010
	(0.017)	(0.017)	(0.016)	(0.019)
Fraction Black	0.0089**	-0.000062	0.0087*	0.00074
	(0.0043)	(0.0038)	(0.0046)	(0.0036)
Fraction Hispanic	-0.0086	-0.0059	-0.0039	-0.0057
	(0.0054)	(0.0054)	(0.0051)	(0.0060)
Achievement (SEDA, CS)	0.0095	-0.010	0.040	0.018
	(0.035)	(0.040)	(0.036)	(0.038)
Achievement (SEDA, CS)	0.053	0.088**	0.044	0.053
(w/ precision weights)	(0.046)	(0.044)	(0.044)	(0.036)
Dropout rate	0.0033	-0.026	0.0030	0.0013
	(0.032)	(0.031)	(0.031)	(0.034)
Graduation rate	-0.0060	0.016	0.066	0.043
	(0.066)	(0.055)	(0.071)	(0.068)
States	All	CB	All	CB
Model	CCT	CCT	OLS	OLS
Specification	Linear	Linear	Quadratic	Quadratic

Note: Each coefficient is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. The dependent variables capture student demographics, achievement, and dropout rates in the 7th year after the election. The CCT models include baseline covariates of the dependent variable one and three years prior to the election. The OLS models include these baseline covariates, as well as year fixed effects, logged operational expenditures per pupil one and three years prior to the election, logged student counts one and three years prior to the election, and student demographics (fraction Black, Hispanic, and FRL) one year prior to the election. Standard errors clustered at the district level appear in parentheses below coefficient estimates. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 6. Spending and Outcomes through 7 Post-Election Years – By Student Poverty

		(1)	(2)	(3)	(4)
ln(op. exp. per pupil)	High Poverty	0.049***	0.049***	0.043**	0.041**
		(0.018)	(0.018)	(0.019)	(0.019)
	Low Poverty	0.026*	0.026*	0.047***	0.047***
		(0.015)	(0.016)	(0.016)	(0.017)
Op. exp. per pupil	High Poverty	530**	506*	526	599
		(262)	(289)	(393)	(396)
	Low Poverty	193	210	344	259
		(302)	(299)	(230)	(259)
Achievement (SEDA, CS)	High Poverty	0.092*	0.070	0.049	0.047
		(0.053)	(0.045)	(0.057)	(0.050)
	Low Poverty	-0.0093	-0.014	-0.020	-0.037
		(0.048)	(0.040)	(0.053)	(0.050)
Achievement (SEDA, CS)	High Poverty	0.090*	0.042	0.075	0.073*
(w/precision weights)		(0.046)	(0.039)	(0.050)	(0.042)
	Low Poverty	-0.024	-0.030	0.039	-0.0019
		(0.057)	(0.039)	(0.047)	(0.044)
Dropout Rate	High Poverty	0.046	0.045	0.0062	0.0094
		(0.056)	(0.056)	(0.057)	(0.057)
	Low Poverty	-0.032	-0.032	-0.0060	-0.0085
		(0.027)	(0.027)	(0.033)	(0.033)
States		All	All	CB	CB
Post-Elect. Covariates		No	Yes	No	Yes
Model		OLS	OLS	OLS	OLS
Specification		Quad.	Quad.	Quad.	Quad.

Note: Each pair of coefficients ("high poverty" and "low poverty") is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. Poverty status is based on whether or not the fraction of FRL-eligible students at baseline puts a district above the median ("high poverty") or below ("low poverty"). The dependent variables capture average annual expenditures through 7 years after the election, as well as student achievement and dropout rates in the 7th year after the election. The pre-election covariates include are those included in the models reported in Table 3 and Table 4, and they differ for the CCT and OLS models. The post-election covariates capture districts' student demographics (fraction Black, Hispanic, and FRL) during the year in which test scores are observed. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 7. Spending and Outcomes through 6-8 Post-Election Years – By Student Poverty

		(1)	(2)	(3)	(4)
ln(op. exp. per pupil)	High Poverty	0.049**	0.050**	0.053***	0.052***
		(0.019)	(0.020)	(0.019)	(0.019)
	Low Poverty	0.023	0.023	0.036**	0.035**
		(0.015)	(0.015)	(0.016)	(0.016)
Op. exp. per pupil	High Poverty	519*	521*	709*	764**
		(283)	(314)	(390)	(384)
	Low Poverty	311	319	356	241
		(298)	(297)	(228)	(262)
Achievement (SEDA, CS)	High Poverty	0.080	0.069	0.057	0.057
		(0.049)	(0.042)	(0.054)	(0.045)
	Low Poverty	-0.0063	-0.0011	-0.018	-0.027
		(0.042)	(0.034)	(0.044)	(0.040)
Achievement (SEDA, CS)	High Poverty	0.093*	0.060	0.084*	0.084**
(w/ precision weights)		(0.052)	(0.040)	(0.048)	(0.040)
	Low Poverty	-0.024	-0.017	0.025	0.0032
		(0.053)	(0.037)	(0.042)	(0.040)
Dropout Rate	High Poverty	0.050	0.046	0.032	0.030
		(0.046)	(0.045)	(0.049)	(0.048)
	Low Poverty	-0.023	-0.024	-0.011	-0.015
		(0.030)	(0.030)	(0.037)	(0.036)
States		All	All	CB	CB
Post-Elect. Covariates		No	Yes	No	Yes
Model		OLS	OLS	OLS	OLS
Specification		Quad.	Quad.	Quad.	Quad.

Note: Each pair of coefficients ("high poverty" and "low poverty") is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. Poverty status is based on whether or not the fraction of FRL-eligible students at baseline puts a district above the median ("high poverty") or below ("low poverty"). The dependent variables capture average annual expenditures through 6-8 years after the election, as well as student achievement and dropout rates in the 6th, 7th, and 8th year after the election. The pre-election covariates include are those included in the models reported in Table 3 and Table 4, and they differ for the CCT and OLS models. The post-election covariates capture districts' student demographics (fraction Black, Hispanic, and FRL) during the year in which test scores are observed. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 8. Achievement over time – By Student Poverty

		1-2 yrs	3-4 yrs	5-6 yrs	7-8 yrs
a. All Elections (2002-2008)					
All States	High Poverty	0.13	0.064	0.093**	0.048
		(0.094)	(0.074)	(0.045)	(0.039)
	Low Poverty	-0.12	0.0012	-0.023	-0.019
		(0.11)	(0.084)	(0.041)	(0.038)
CB States	High Poverty	-0.011	0.017	0.090*	0.087**
		(0.075)	(0.073)	(0.050)	(0.041)
	Low Poverty	0.0088	0.072	0.012	-0.0036
		(0.059)	(0.048)	(0.044)	(0.040)
b. 2007/2008 Elections					
All States	High Poverty	0.11	0.14*	0.061	0.045
		(0.098)	(0.078)	(0.079)	(0.070)
	Low Poverty	-0.13	-0.035	-0.00093	0.014
		(0.11)	(0.10)	(0.062)	(0.058)
CB States	High Poverty	-0.014	0.034	0.052	0.059
		(0.075)	(0.083)	(0.070)	(0.060)
	Low Poverty	0.00024	0.059	0.051	0.029
		(0.061)	(0.056)	(0.058)	(0.057)
Post-Elect. Covs.		Yes	Yes	Yes	Yes
Model		OLS	OLS	OLS	OLS
Specification		Quad.	Quad.	Quad.	Quad.

Note: Each pair of coefficients ("high poverty" and "low poverty") is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. Poverty status is based on whether or not the fraction of FRL-eligible students at baseline puts a district above the median ("high poverty") or below ("low poverty"). The dependent variables capture student achievement 1-2 years, 3-4 years, 5-6 years, and 7-8 years after the election. The "2007/2008 elections" sample used in panel *b* includes a subset of referenda for which we observe achievement data in all four post-election year bins (i.e., 1-2, 3-4, 5-6, and 7-8). The full sample in panel *a* includes all referenda for which we observe achievement in year 7, but not necessarily all prior years. The pre-election covariates are those included in the models reported in Table 3 and Table 4. The post-election covariates capture districts' student demographics (fraction Black, Hispanic, and FRL) during the year in which test scores are observed. All models include precision weights. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A Tax Levy Information by State

Table A1. Local Tax Levy Information by State

Arkansas	Revenues come primarily from property taxes. There are no local sales or income taxes. There are no limits on local property tax rates that school districts can levy, but there is a minimum of 25 mills for maintenance and operations. A majority of voters must approve increases to property tax rates beyond this minimum. The constitution also mandates that millage rates that are unchanged appear on the ballot every year, but the vote outcome does not affect a district's millage rate.
Louisiana	Revenues come primarily from sales taxes. These are capped at 3 percent for all local entities combined and a majority of voters must approve them. Parish school boards also have the authority to levy a "constitutional" property tax of up to 5 mills (13 mills in Orleans). Districts can supplement this by obtaining voter approval to levy additional property taxes for a specific purpose relating to operations, maintenance, or capital expenses. There is a maximum of 70 mills above the "constitutional" tax and it is limited in duration.
Michigan	Revenues come primarily from property taxes. School districts must get approval from a majority of voters if they wish to exceed caps on local property taxes that the state set in 1994. In general, there is a cap of 18 mills on non-homestead property taxes. A majority of school district voters must approve millage increases for non-homestead properties and must renew these mills over time. Although law generally prohibits property taxes on homesteads (including noncommercial and agricultural property), districts for which the 18 mill cap does not allow them to get to pre-1994 levels may levy additional taxes to get up to that level. There were 52 (of 555) traditional school districts that met this "hold harmless" condition after the law's passage.
Missouri	Revenues come primarily from property taxes. Local school districts can levy taxes of up to \$2.75 per \$100 of assessed valuable without voter approval and must do so to receive increases in state aid. Tax rates above \$2.75 require approval of 50 percent of voters and rates that exceed \$6.00 per \$100 of valuation must be approved by two thirds of voters. There is no limit on tax rates.
Pennsylvania	Revenues come primarily from property taxes. A 2006 law requires voter approval for any proposed tax increase that exceeds an index capturing increases in wages and employment costs for schools. The 2006 law also sought to provide homeowners with property tax relief and gave districts the option of shifting some of the tax burden from local property taxes to local income taxes to offset losses. The vast majority of districts placed such a measure on the ballot in May of 2007, which voters overwhelmingly defeated throughout the state.
Texas	Revenues come primarily from property taxes. Districts adopt maintenance and operations tax rates each year. School districts must hold a tax referendum if school boards adopt a tax rate that exceed the "rollback rate." This rate effectively allows minimal increases in district revenues without voter approval. Voters must also approve any income taxes that a school board proposes.
Wisconsin	Revenues come primarily from property taxes. District can raise property tax revenues up to a state-mandated revenue limit. Districts must obtain approval from a majority of district voters to exceed the state revenue limit. Voters can elect to exceed a revenue limit permanently.

B Covariate Balance Tests

TABLE B1. Pre-Election Balance Tests (Linear, CCT)

	(1) Level	(2) Pre-Trend	(3) Level	(4) Pre-Trend
ln(revenue per pupil)	-0.026	0.0095	0.043	0.014
	(0.031)	(0.012)	(0.028)	(0.012)
ln(local rev. per pupil)	-0.067	0.00050	0.042	-0.0056
	(0.089)	(0.023)	(0.11)	(0.022)
ln(fed. rev. per pupil)	0.098	0.035	0.10	0.071*
	(0.10)	(0.038)	(0.13)	(0.043)
ln(state rev. per pupil)	0.0096	0.072*	-0.0050	0.019
	(0.079)	(0.042)	(0.088)	(0.022)
ln(op. expend. per pupil)	-0.027	-0.015	0.041*	-0.0050
	(0.025)	(0.0094)	(0.025)	(0.0098)
ln(inst. expend. per pupil)	-0.025	-0.025**	0.037	-0.015
	(0.026)	(0.010)	(0.025)	(0.011)
ln(cap. expend. per pupil)	-0.31	-0.42	-0.28	-0.28
	(0.28)	(0.34)	(0.31)	(0.36)
ln(salaries per staff)	-0.067*	0.018	0.00022	0.023
	(0.034)	(0.019)	(0.027)	(0.023)
ln(benefits per staff)	-0.16*	-0.015	-0.039	-0.021
	(0.086)	(0.027)	(0.049)	(0.032)
ln(students per schl)	-0.031	0.00064	-0.013	-0.0044
	(0.071)	(0.021)	(0.085)	(0.022)
ln(students per staff)	-0.049	0.023	-0.043	0.014
	(0.030)	(0.019)	(0.029)	(0.023)
ln(students per teacher)	-0.048*	0.0094	-0.071**	0.0071
	(0.026)	(0.011)	(0.034)	(0.012)
ln(students per aides)	-0.018	0.038	0.064	0.065
	(0.096)	(0.056)	(0.10)	(0.062)
ln(students per counselor)	-0.085	0.040	-0.13	0.0022
	(0.069)	(0.038)	(0.081)	(0.050)
ln(student count)	-0.0034	0.0035	-0.017	0.0027
	(0.20)	(0.0086)	(0.24)	(0.010)
States	All	All	CB	CB
Model	CCT	CCT	CCT	CCT

Note: Each coefficient is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. Results in columns 1 and 2 are based on the full sample of levies, whereas columns 3 and 4 are based on levies from collective-bargaining (CB) states—that is, Michigan, Pennsylvania, and Wisconsin. Differences in levels are estimated based on the year prior to the election, and differences in pre-treatment trends are based on differences between one year prior and three year prior. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE B1. Pre-election balance tests (Linear, CCT)

	(1) Level	(2) Pre-Trend	(3) Level	(4) Pre-Trend
Dropout Rate	-0.011	0.0093	-0.028	0.019
	(0.030)	(0.021)	(0.033)	(0.025)
Graduation Rate	0.038	0.012	0.100*	-0.037
	(0.049)	(0.054)	(0.057)	(0.063)
Fraction Hispanic	0.033	-0.00051	0.017	-0.0011
	(0.025)	(0.0019)	(0.023)	(0.0024)
Fraction Black	0.024	0.0013	0.0089	0.00036
	(0.029)	(0.0017)	(0.015)	(0.0018)
Fraction White	-0.052	0.00026	-0.019	0.0014
	(0.040)	(0.0038)	(0.034)	(0.0047)
Fraction LEP	-0.00066	-0.0071	-0.0020	-0.0087
	(0.0081)	(0.0066)	(0.0097)	(0.0087)
Fraction IEP	0.00032	-0.0095	0.0020	-0.0063
	(0.0062)	(0.0087)	(0.0073)	(0.0071)
Fraction F/R lunch	0.034	0.015	-0.020	0.0026
	(0.036)	(0.011)	(0.035)	(0.012)
Rural	0.016		-0.040	
	(0.091)		(0.097)	
Town	-0.077		-0.10	
	(0.073)		(0.090)	
Suburb	-0.0055		0.062	
	(0.074)		(0.077)	
City	0.051		0.043	
	(0.051)		(0.062)	
States	All	All	CB	CB
Model	CCT	CCT	CCT	CCT

Note: Each coefficient is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. Results in columns 1 and 2 are based on the full sample of levies, whereas columns 3 and 4 are based on levies from collective-bargaining (CB) states—that is, Michigan, Pennsylvania, and Wisconsin. Differences in levels are estimated based on the year prior to the election, and differences in pre-treatment trends are based on differences between one year prior and three year prior. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE B2. Pre-election balance tests (OLS, quad, +/- 20)

	(1) Level	(2) Pre-Trend	(3) Level	(4) Pre-Trend
ln(revenue per pupil)	-0.020	0.0096	0.065**	0.015
	(0.032)	(0.011)	(0.028)	(0.011)
ln(local rev. per pupil)	0.0035	0.0020	0.17	0.0063
	(0.098)	(0.022)	(0.11)	(0.024)
ln(fed. rev. per pupil)	0.11	0.032	0.028	0.068
	(0.12)	(0.043)	(0.13)	(0.051)
ln(state rev. per pupil)	-0.013	0.066*	-0.039	0.019
	(0.081)	(0.039)	(0.080)	(0.019)
ln(op. expend. per pupil)	-0.031	-0.013	0.047*	-0.0050
	(0.030)	(0.011)	(0.027)	(0.011)
ln(inst. expend. per pupil)	-0.036	-0.023*	0.040	-0.015
	(0.032)	(0.012)	(0.029)	(0.014)
ln(cap. expend. per pupil)	-0.14	-0.15	-0.098	-0.23
	(0.24)	(0.28)	(0.29)	(0.36)
ln(salaries per staff)	-0.059*	0.022	0.020	0.046
	(0.034)	(0.024)	(0.028)	(0.034)
ln(benefits per staff)	-0.19**	-0.011	-0.032	-0.00051
	(0.084)	(0.030)	(0.044)	(0.040)
ln(students per schl)	-0.023	-0.014	0.027	-0.0091
	(0.082)	(0.022)	(0.095)	(0.025)
ln(students per staff)	-0.051*	0.022	-0.038	0.035
	(0.029)	(0.023)	(0.029)	(0.034)
ln(students per teacher)	-0.054*	0.014	-0.074**	0.019
	(0.028)	(0.015)	(0.032)	(0.021)
ln(students per aides)	-0.040	0.033	0.030	0.044
	(0.094)	(0.061)	(0.11)	(0.072)
ln(students per counselor)	-0.078	0.049	-0.16**	-0.0025
	(0.071)	(0.048)	(0.078)	(0.058)
ln(student count)	0.17	0.0060	0.24	0.0051
	(0.20)	(0.0098)	(0.22)	(0.012)
States	All	All	CB	CB
Model	OLS	OLS	OLS	OLS

Note: Each coefficient is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. Results in columns 1 and 2 are based on the full sample of levies, whereas columns 3 and 4 are based on levies from collective-bargaining (CB) states—that is, Michigan, Pennsylvania, and Wisconsin. Differences in levels are estimated based on the year prior to the election, and differences in pre-treatment trends are based on differences between one year prior and three year prior. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

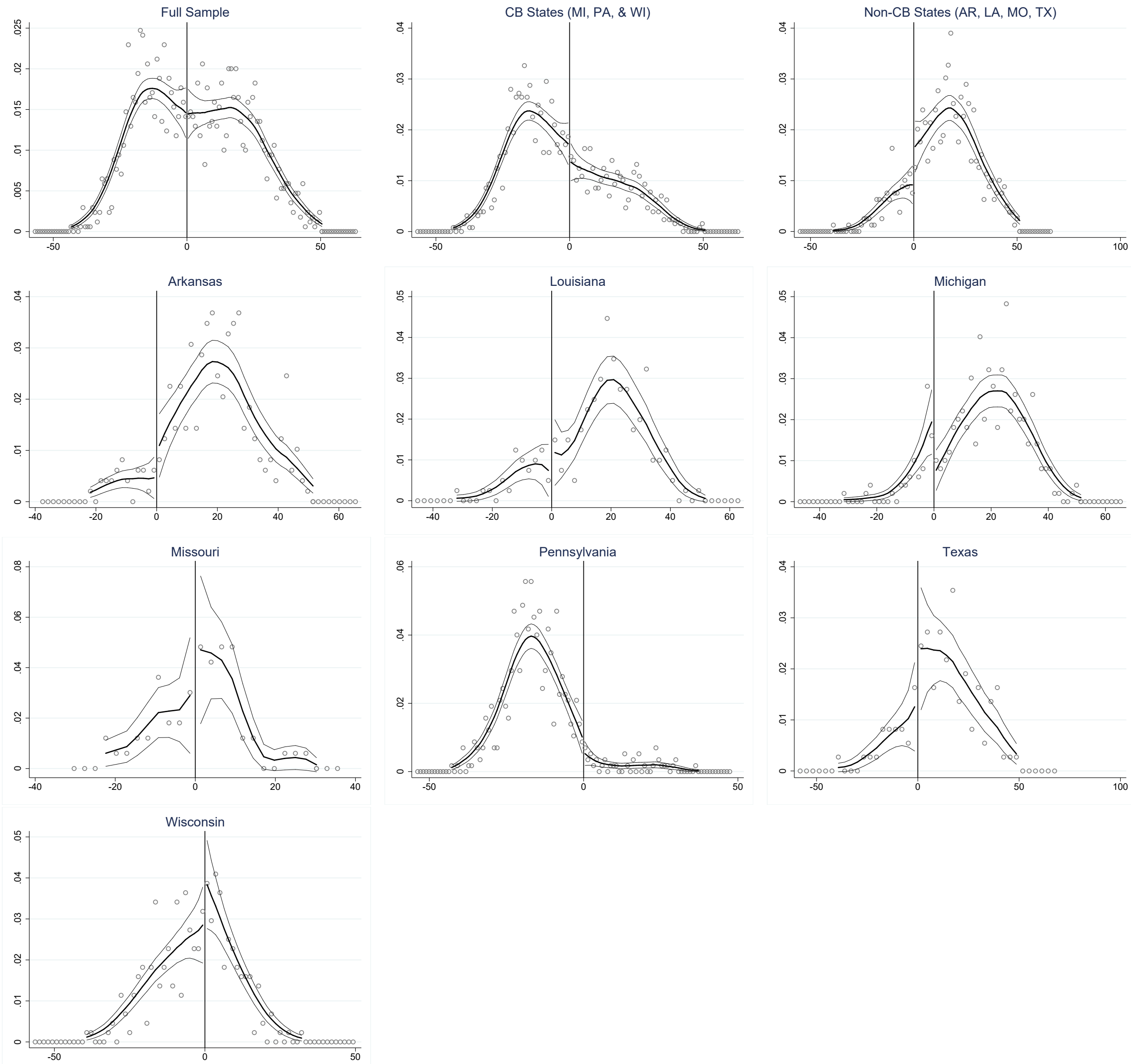
TABLE B2. Pre-election balance tests (OLS, quad, +/- 20), cont'd

	(1) Level	(2) Pre-Trend	(3) Level	(4) Pre-Trend
Dropout Rate	-0.023	0.0032	-0.044	0.0060
	(0.037)	(0.026)	(0.041)	(0.029)
Graduation Rate	0.025	-0.0070	0.089	-0.058
	(0.049)	(0.047)	(0.059)	(0.061)
Fraction Hispanic	0.035	-0.00066	0.022	-0.0012
	(0.027)	(0.0020)	(0.020)	(0.0025)
Fraction Black	0.029	0.00077	0.00031	-0.00099
	(0.033)	(0.0019)	(0.017)	(0.0017)
Fraction White	-0.065	-0.0016	-0.028	-0.0014
	(0.041)	(0.0034)	(0.032)	(0.0043)
Fraction LEP	0.00030	-0.0048	0.0046	-0.0057
	(0.0081)	(0.0054)	(0.0096)	(0.0062)
Fraction IEP	0.00098	-0.012	0.0024	-0.010
	(0.0072)	(0.0089)	(0.0095)	(0.0097)
Fraction F/R lunch	0.014	0.013	-0.058*	-0.0043
	(0.036)	(0.012)	(0.033)	(0.014)
Rural	-0.047		-0.10	
	(0.091)		(0.11)	
Town	-0.048		-0.080	
	(0.080)		(0.093)	
Suburb	0.057		0.13*	
	(0.065)		(0.079)	
City	0.039		0.047	
	(0.055)		(0.061)	
States	All	All	CB	CB
Model	OLS	OLS	OLS	OLS

Note: Each coefficient is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. Results in columns 1 and 2 are based on the full sample of levies, whereas columns 3 and 4 are based on levies from collective-bargaining (CB) states—that is, Michigan, Pennsylvania, and Wisconsin. Differences in levels are estimated based on the year prior to the election, and differences in pre-treatment trends are based on differences between one year prior and three year prior. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C Density Tests

Figure 5: Density Tests



Note: The figure presents the results of McCrary (2008) density tests. Oklahoma had no failing referenda.

D Additional Tables

TABLE D1. Average Annual Expenditures and Staffing through Years 6-8

	(1)	(2)	(3)	(4)
ln(op. exp. per pupil)	0.029**	0.029**	0.035***	0.045***
	(0.013)	(0.013)	(0.012)	(0.012)
op. exp. per pupil ($)	327**	302**	380**	525***
	(154)	(154)	(148)	(150)
ln(inst. exp. per pupil)	0.024*	0.032**	0.029**	0.042***
	(0.014)	(0.013)	(0.013)	(0.013)
ln(cap. exp. per pupil)	-0.047	-0.061	-0.058	-0.12
	(0.14)	(0.17)	(0.12)	(0.15)
ln(salaries per staff)	0.031**	0.041**	0.032**	0.032**
	(0.014)	(0.016)	(0.013)	(0.016)
salaries per staff ($)	275	1,288	1,552*	1,793
	(1,129)	(1,199)	(856)	(1,324)
ln(benefits per staff)	0.018	0.018	0.027	0.015
	(0.022)	(0.023)	(0.019)	(0.019)
ln(students per teacher)	0.0039	-0.00019	-0.0046	-0.015
	(0.013)	(0.015)	(0.012)	(0.013)
ln(students per staff)	0.0058	-0.0057	-0.000026	-0.011
	(0.012)	(0.016)	(0.013)	(0.016)
States	All	CB	All	CB
Model	CCT	CCT	OLS	OLS
Specification	Linear	Linear	Quadratic	Quadratic

Note: Each coefficient is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. The dependent variables are averages of annual expenditure and staffing variables through 6-8 years after the election. The CCT models include baseline covariates of the dependent variable one and three years prior to the election. The OLS models include these baseline covariates, as well as year fixed effects, logged operational expenditures per pupil one and three years prior to the election, logged student counts one and three years prior to the election, and student demographics (fraction Black, Hispanic, and FRL) one year prior to the election. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE D2. Student Demographics and Academic Outcomes in Years 6-8

	(1)	(2)	(3)	(4)
ln(student count)	0.034	0.066*	0.030	0.024
	(0.029)	(0.036)	(0.025)	(0.029)
Fraction FR Lunch	-0.00018	0.00098	-0.010	-0.0046
	(0.014)	(0.015)	(0.014)	(0.015)
Fraction Black	0.0091**	-0.0013	0.0083*	-0.00015
	(0.0044)	(0.0036)	(0.0048)	(0.0034)
Fraction Hispanic	-0.0068	-0.0068	-0.0033	-0.0061
	(0.0053)	(0.0055)	(0.0048)	(0.0058)
Achievement (SEDA, CS)	0.010	0.0077	0.036	0.024
	(0.033)	(0.038)	(0.033)	(0.034)
Achievement (SEDA, CS)	0.063	0.091*	0.045	0.054
(w/ precision weights)	(0.051)	(0.047)	(0.050)	(0.037)
Dropout rate	0.021	-0.0063	0.012	0.012
	(0.033)	(0.031)	(0.028)	(0.031)
Graduation rate	0.018	0.073	0.10**	0.11
	(0.050)	(0.057)	(0.053)	(0.072)
States	All	CB	All	CB
Model	CCT	CCT	OLS	OLS
Specification	Linear	Linear	Quadratic	Quadratic

Note: Each coefficient is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. The dependent variables capture student demographics, achievement, and dropout rates in the 6-8 years after the election. The CCT models include baseline covariates of the dependent variable one and three years prior to the election. The OLS models include these baseline covariates, as well as year fixed effects, logged operational expenditures per pupil one and three years prior to the election, logged student counts one and three years prior to the election, and student demographics (fraction Black, Hispanic, and FRL) one year prior to the election. Standard errors clustered at the district level appear in parentheses below coefficient estimates. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE D3. Student Achievement Sensitivity Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
a. Year 7						
All States	-0.026	0.053	0.0054	-0.013	0.044	0.0085
	(0.092)	(0.046)	(0.032)	(0.095)	(0.044)	(0.029)
CB States	0.044	0.088**	0.039	0.089	0.053	0.032
	(0.13)	(0.044)	(0.030)	(0.13)	(0.036)	(0.031)
a. Years 6-8						
All States	-0.036	0.063	0.024	-0.010	0.045	0.026
	(0.098)	(0.051)	(0.034)	(0.099)	(0.050)	(0.033)
CB States	0.058	0.091*	0.041	0.099	0.054	0.041
	(0.14)	(0.047)	(0.032)	(0.13)	(0.037)	(0.030)
Model	CCT	CCT	CCT	OLS	OLS	OLS
Pre-election Covariates	No	Yes	Yes	No	Yes	Yes
Post-election Covariates	No	No	Yes	No	No	Yes
Weights	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each coefficient is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. The dependent variable is the average of SEDA's student achievement estimate in math and ELA, standardized at the cohort level. Panel *a* presents the results of models based on data 7 years after an election and b presents the results of models based on pooled data 6-8 years after an election. The possible pre-election covariates are those included in the models reported in Table 3 and Table 4 and, as in those tables, they differ for the CCT and OLS models. The post-election covariates capture districts' student demographics (fraction Black, Hispanic, and FRL) during the year in which test scores are observed. All models include precision weights. 95 percent confidence intervals are reported in brackets below the coefficient estimates. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE D4. Changes in Student Composition 6-8 years after levy

		Yrs 6-8 (1)	Yrs 6-8 (2)	Year 7 (3)	Year 7 (4)
ln(stu. count)	High Pov.	0.019	0.042	0.0090	0.035
		(0.034)	(0.041)	(0.038)	(0.040)
	Low Pov.	0.034	-0.0055	0.0077	-0.019
		(0.041)	(0.035)	(0.041)	(0.037)
Frac. Black	High Pov.	0.0096	0.0026	0.0085	0.0041
		(0.0078)	(0.0052)	(0.0073)	(0.0059)
	Low Pov.	0.0073	-0.0032	0.0087*	-0.0032
		(0.0051)	(0.0040)	(0.0052)	(0.0041)
Frac. Hispanic	High Pov.	-0.0054	-0.011	-0.0046	-0.010
		(0.0078)	(0.0095)	(0.0081)	(0.0097)
	Low Pov.	-0.0027	-0.0011	-0.0042	-0.00076
		(0.0059)	(0.0060)	(0.0065)	(0.0066)
Frac. FRL	High Pov.	-0.018	0.0041	-0.027	0.00053
		(0.020)	(0.022)	(0.023)	(0.030)
	Low Pov.	0.0022	-0.0097	-0.0032	-0.017
		(0.020)	(0.019)	(0.023)	(0.020)
States		All	CB	All	CB
Model		OLS	OLS	OLS	OLS
Specification		Quad.	Quad.	Quad.	Quad.

Note: Each pair of coefficients ("high poverty" and "low poverty") is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. Poverty status is based on whether or not the fraction of FRL-eligible students at baseline puts a district above the median ("high poverty") or below ("low poverty"). The dependent variables capture average student demographics 6-8 years after the election (columns 1 and 2) or 7 years after the election (columns 3 and 4). All models include pre-treatment covariates, as in primary specifications in main text. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE D5. Spending and Outcomes through 7 Post-Election Years – By Rurality

		(1)	(2)	(3)	(4)
ln(op. exp. per pupil)	Rural	0.039**	0.039**	0.047***	0.046***
		(0.016)	(0.016)	(0.016)	(0.016)
	Non-rural	0.033**	0.033**	0.047***	0.046***
		(0.016)	(0.017)	(0.018)	(0.018)
Achievement (SEDA, CS)	Rural	0.089*	0.062	0.040	0.035
		(0.051)	(0.046)	(0.058)	(0.057)
	Non-rural	-0.0077	-0.0091	-0.013	-0.029
		(0.051)	(0.039)	(0.051)	(0.042)
Achievement (SEDA, CS)	Rural	0.026	0.026	0.025	0.054
(w/ precision weights)		(0.043)	(0.035)	(0.057)	(0.050)
	Non-rural	0.059	0.0087	0.061	0.028
		(0.060)	(0.039)	(0.043)	(0.038)
Dropout Rate	Rural	-0.024	-0.024	-0.022	-0.023
		(0.038)	(0.038)	(0.041)	(0.041)
	Non-rural	0.023	0.023	0.026	0.028
		(0.050)	(0.049)	(0.056)	(0.055)
States		All	All	CB	CB
Post-Elect. Covariates		No	Yes	No	Yes
Model		OLS	OLS	OLS	OLS
Specification		Quad.	Quad.	Quad.	Quad.

Note: Each pair of coefficients ("rural" and "non-rural") is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. The dependent variables capture average annual expenditures through 7 years after the election, as well as student achievement and dropout rates in the 7th year after the election. The pre-election covariates are those included in the models reported in Table 3 and Table 4, and they differ for the CCT and OLS models. The post-election covariates capture districts' student demographics (fraction Black, Hispanic, and FRL) during the year in which test scores are observed. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE D6. Spending and Outcomes through 6-8 Post-Election Years – By Rurality

		(1)	(2)	(3)	(4)
ln(op. exp. per pupil)	Rural	0.036**	0.036**	0.046***	0.045***
		(0.016)	(0.016)	(0.016)	(0.016)
	Non-rural	0.034**	0.035**	0.047***	0.046***
		(0.017)	(0.017)	(0.018)	(0.018)
Achievement (SEDA, CS)	Rural	0.088**	0.071*	0.051	0.052
		(0.045)	(0.039)	(0.047)	(0.043)
	Non-rural	-0.013	-0.0051	-0.0071	-0.018
		(0.049)	(0.037)	(0.049)	(0.040)
Achievement (SEDA, CS)	Rural	0.011	0.023	0.032	0.056
(w/ precision weights)		(0.045)	(0.036)	(0.054)	(0.046)
	Non-rural	0.062	0.030	0.059	0.037
		(0.067)	(0.044)	(0.046)	(0.037)
Dropout Rate	Rural	0.0043	0.0022	-0.00031	-0.0011
		(0.036)	(0.036)	(0.040)	(0.040)
	Non-rural	0.017	0.015	0.025	0.020
		(0.044)	(0.043)	(0.048)	(0.046)
States		All	All	CB	CB
Post-Elect. Covariates		No	Yes	No	Yes
Model		OLS	OLS	OLS	OLS
Specification		Quad.	Quad.	Quad.	Quad.

Note: Each pair of coefficients ("rural" and "non-rural") is from a separate regression and captures the difference in outcomes between districts in which referenda passed and those in which referenda failed. The dependent variables capture average annual expenditures through 7 years after the election, as well as student achievement and dropout rates in the 7th year after the election. The pre-election covariates are those included in the models reported in Table 3 and Table 4, and they differ for the CCT and OLS models. The post-election covariates capture districts' student demographics (fraction Black, Hispanic, and FRL) during the year in which test scores are observed. Standard errors clustered at the district level appear in parentheses below coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$