

Performance Federalism and Local Democracy: Theory and Evidence from School Tax Referenda*

Vladimir Kogan
Department of Political Science
Ohio State University
2140 Derby Hall
154 North Oval Mall
Columbus OH, 43210
kogan.18@osu.edu

Stéphane Lavertu
John Glenn School of Public Affairs
Ohio State University
110 Page Hall
1810 College Road
Columbus, OH 43210
lavertu.1@osu.edu

Zachary Peskowitz
Department of Political Science
Ohio State University
2140 Derby Hall
154 North Oval Mall
Columbus OH, 43210
peskowitz.4@osu.edu

*Replication data for this study will be available upon publication at <http://glennschool.osu.edu/educational-governance/>. We thank the Ohio State University Center for the Study of Democracy for research funding; Cory Barnes, T.J. Beavers, Annie Curie, Ray Gans, William McCracken, Kyle Larson, and Chaekwang You for their indispensable research assistance; and the Ohio School Board Association for giving us access to their database. We also thank Chris Achen, Rob Bond, Craig Burnett, Greg Caldeira, Alex Hirsch, John Holbein, Matt Hitt, Carl Klarner, Todd Makse, Nolan McCarty, Marc Meredith, William Minozzi, Tom Nelson, Jan Pierskalla, and Lapo Salucci for helpful comments. The authors' names are in alphabetical order, reflecting equal contributions.

Abstract

Federal governments are increasingly employing empirical measures of lower-level government performance to ensure that provincial and local jurisdictions pursue national policy goals. We call this burgeoning phenomenon *performance federalism* and argue that it can distort democratic accountability in lower-level elections. We estimate the impact of a widely publicized federal indicator of local school district performance — one that we show does not allow voters to draw valid inferences about the quality of local educational institutions — on voter support for school tax levies in a U.S. state uniquely appropriate for this analysis. The results indicate that a signal of poor district performance increases the probability of levy failure — a substantively large and robust effect that disproportionately affects impoverished communities. The analysis employs a number of identification strategies and tests for multiple behavioral mechanisms to support the causal interpretation of these findings.

Running Head: Performance Federalism

Word count: 8,758

Keywords: federalism, performance measurement, voting behavior, new public management, education politics, No Child Left Behind

Introduction

Federal systems around the world have evolved significantly in recent decades, with increasing interactions and sharing of responsibility between different levels of government. In tandem with these changes, reformers inspired by private-sector management practices have sought to transform administrative states from rigid, rule-bound hierarchies to flexible organizations focused on the efficient pursuit of public goals. These reforms have been motivated by a principal-agent perspective that emphasizes the delegation of authority to administrative agents and the measurement of administrative outcomes so that agents may be incentivized to pursue the objectives of political principals. Scholars have examined extensively how such “performance management” reforms are transforming public administration, but little research has focused on the unintended political consequences of such reforms for federal systems in which higher-level governments seek to shape the incentives of lower-level governments responsible for administering shared programs.¹ This is an important research gap because the adoption of “performance federalism” will surely grow as the technological capacity to realize performance-based accountability schemes in multi-level institutional contexts continues to improve.

This study focuses on the impact of performance federalism on local democratic politics in the United States. The U.S. federal government has established performance-based incentives for state and local governments administering a multitude of federally and jointly funded programs, most notably those dealing with job training (e.g., see Heckman et al. 2011) and K-12 public education (e.g., see Manna 2006, 2011). Historically, the federal government has moved from imposing prescriptive rules and procedures for the use of federal funds to delegating significant discretion to state and local governments, partly in the hope of improving programmatic efficiency but at the risk of state and local governments using those funds to pursue something other than federal goals (Conlan, 1998). The appeal

¹Two notable exceptions are research by Boyne et al. (2009) and James and John (2007). Bertelli and John (2010) consider the multi-level politics of performance measurement from a principal-agent perspective, but they do not examine the impact of performance designations on local democracy.

of performance measurement is that the federal government can minimize the tradeoff between administrative efficiency and shirking by ameliorating the informational asymmetry underlying the principal-agent problem. But the consequences for state and local democracy are likely significant. As we demonstrate, federal performance measures can result in voters drawing invalid inferences about the performance of local governments in realizing their public goals, distorting and undermining democratic accountability in local elections.

We examine these consequences by estimating the impact of federal performance measures in local school tax referenda in Ohio from 2003 to 2012. Doing so provides several advantages. First, during these years, the federal No Child Left Behind Act (NCLB) imposed a performance metric that provided a clear informational signal of whether or not schools made “adequate yearly progress” (AYP) toward meeting federal policy goals. Second, Ohio provides tremendous variation because of its size and an unusual state legal context that limits local discretion over school tax referenda, requiring school districts to go to the ballot more frequently and with less discretion over timing than is the case in other states. Third, the Ohio data allow us to employ a panel design and to evaluate the quality of the performance indicators available to voters, two important contributions to the emerging literature on the political implications of performance measurement. In short, this is an unusually fertile period and place to identify the causal impact of performance federalism on local elections and public policy outcomes. Although our empirical analysis focuses on a single state to maximize the strength of our research design and the quality of available data,² the main findings speak to core questions of democratic accountability in multi-level systems.

The empirical analysis provides a striking illustration of the informational impact of federal performance measures in local elections and the extent to which such measures can undermine electoral accountability. We show that school districts that failed to meet the federal government’s AYP goal — a standard which, as we demonstrate below, yields a highly problematic metric for judging the performance of local school districts — saw the

²The value-added measures we utilize to evaluate the construct validity of AYP are publicly available for a sufficient number of years only in a handful of states.

probability of passing their levies fall by more than 10 percent. The substantively large effect reduced local school district revenue by more than 13 percent on average, potentially undermining the quality of public education. Moreover, because the problematic AYP metric effectively served as an indicator of student socioeconomic status instead of district or school quality, these perverse effects disproportionately affected impoverished communities.

The paper proceeds as follows. First, we briefly review the existing literature and make the case that a new kind of federalism — which we refer to as *performance federalism* — is emerging and likely to affect electoral accountability in state and local elections. Second, we provide background on the federal No Child Left Behind performance-based accountability system and describe the inferential problems associated with using federal school district designations to assess the performance of local districts. Third, we delineate our hypotheses and describe the empirical approach. Fourth, we demonstrate the impact of these performance measures on the passage of local school tax levy proposals, test for the behavioral mechanisms of our theory (for example, by examining the timing and novelty of the informational signals), and conduct placebo tests that support the causal interpretation of our findings. Finally, we conclude by examining the substantive significance of the findings and offer some concluding thoughts about the emerging importance of performance federalism.

Performance Federalism and Local Elections

America's federal system is increasingly marked by the intermingling of policy authority and fiscal resources between different levels of government (Walker, 1995). Scholars have offered a number of explanations for growing federal involvement in functions traditionally performed by state and local governments, including federal officials' desire to take credit for the provision of popular public services (Volden, 2005) and their desire to overcome the institutional collective action problems that plague state and local policymaking on issues ranging from environmental protection to economic redistribution (Oates, 1972; Peterson

and Paul, 1989).

But federal policymakers face significant principal-agent problems as they seek to influence public programs administered at the state and local levels (Scholz and Wei, 1986). Merely outlining the goals that state and local actors ought to pursue as they spend federal funds is unlikely to elicit maximal effort toward realizing those goals. State and local agents possess significant expertise that federal officials wish to harness, but the discretion that these agents need in order to mobilize their expertise may also enable them to pursue their own, divergent objectives. In an effort to address this problem, federal principals have increasingly sought to measure the extent to which state and local actors — both elected and appointed — realize federal policy goals. Such a strategy can, in theory, enable the federal government to design mechanisms to incentivize state and local agents to pursue federal policy goals without binding them with discretion-limiting regulations (Holmstrom, 1979). This institutional logic, inspired by the economics of organization and often promoted under the banner of “new public management,” has motivated sweeping administrative reforms across the globe (Pollitt and Bouckaert, 2004). And dramatic improvements in our ability to collect and analyze data increasingly make such performance federalism a viable strategy.

The devolution of federal authority, growing interrelationships between levels of government, and adoption of performance management approaches have all attracted substantial attention from the scholarly community (e.g., Conlan 1998; Heinrich and Marschke 2010), but their collective implications for local democracy have not been explored. Incentive mechanisms tied to performance measures can motivate lower-level actors to pursue federal policy goals, but such metrics may also create significant informational spillovers that reach ordinary voters. Indeed, performance measures collected for purposes of federal oversight can influence voter behavior in state and local elections in unintended ways that may actually undermine federal objectives. Such consequences are particularly likely in American local elections, because they are often characterized by low levels of voter engagement and the absence of information about the performance of incumbent officials and public agencies.

While some local contests feature lively campaigns, result in substantial spending, and attract media attention, the vast majority do not. In this information-poor context, voters weighing down-ballot races must either invest substantial time and personal effort to investigate local issues and candidates — which, rationally, few do (Downs, 1957) — or turn to heuristics and cues (Lau and Redlawsk, 2001). Because the vast majority of local elections in the U.S. are nonpartisan, voters lack easy access to partisan labels that prove so informative in state and federal races. Many instead turn to alternative cues, often relying on candidate race and ethnicity (Lorinkas, Hawkins and Edwards, 1969; Matson and Fine, 2006), incumbency (Schaffner, Streb and Wright, 2001), occupation (McDermott, 2005), and other shortcuts. Shared policy authority among multiple actors within and between levels of government also complicate the cognitive task of evaluating government performance by requiring voters to correctly apportion credit and blame for observed policy outcomes (Brown, 2010; Malhotra and Kuo, 2008).

In principle, federally imposed performance measures can improve democratic accountability by clarifying responsibility and generating widely disseminated, easy-to-interpret benchmarks that voters can use as the basis for retrospective evaluations (Berry and Howell, 2007). In practice, however, performance measures can also *impede* accountability at the local level if such measures are poorly designed or misinterpreted and misapplied by voters. As we demonstrate with an examination of the performance measurement system required by the No Child Left Behind Act, this possibility indeed became a reality in U.S. public education.

No Child Left Behind Act of 2001

The provision of K-12 public education has historically been a local responsibility in the U.S., with increasing state and, later, federal participation over the past five decades. By the 1980s and 1990s, amid growing concern that poor school performance undermined America's

economic prosperity and global competitiveness, federal policymakers began to demand more rigorous assessment and reporting of student educational outcomes (Manna, 2011, p. 7-9). In 2001, federal lawmakers seeking to hold states and local school districts accountable for eliminating achievement gaps and raising overall educational performance triumphed with the No Child Left Behind Act (NCLB). Though the federal share of public education spending climbed only from around 7 percent to around 10 percent of total K-12 spending, the accountability system that NCLB imposed as a condition for receiving this money marked perhaps the largest increase in the federal role in public education since the 1960s (Manna, 2006).

Specifically, NCLB required states to set annual performance targets that increased steadily through the 2013-2014 school year, by which time the law required 100 percent of students to be “proficient” in mathematics and reading. Schools and districts whose students met the yearly proficiency rate targets would be labeled as having made “adequate yearly progress” (AYP) toward full proficiency by 2014. NCLB also required states to develop academic content standards and tests aligned to those standards so that student progress could be tracked. For purposes of determining whether individual schools and districts made AYP, districts had to administer mathematics and reading tests in grades 3 through 8 and once in high school. For schools and districts to satisfy the AYP standard in a given year, every student subgroup (according to race, gender, economic disadvantage, disability, and English-language status) had to meet the AYP proficiency rate target for that year.³

Unlike previous federal education reforms, NCLB featured strict enforcement mechanisms to incentivize schools and districts to meet the escalating proficiency requirements. First, the law required states to publicize school and district AYP designations (“met” or “not met”) on report cards released just prior to the following school year. The federal government also required districts to directly notify parents whose children attended schools or

³There were some additional ways beyond meeting the overall AYP proficiency rate targets that student subgroups could avoid failing the AYP proficiency requirement. Meeting the proficiency target for each subgroup was also not sufficient to “meet” AYP, with the law also requiring minimum attendance, test participation, and graduation rates.

districts that failed to meet the AYP performance requirement for two consecutive years. These requirements meant that many educational stakeholders received a signal of school and district quality based on the dichotomous AYP designation. Second, NCLB required invasive administrative interventions when schools or districts failed to meet AYP proficiency targets for at least two years in a row.⁴

Volatility and Validity of NCLB’s AYP Performance Measures

The implementation of NCLB performance standards resulted in significant variation over time and across states in AYP performance targets and the resulting measures of student performance (Davidson et al., 2013). One reason is that the U.S. Department of Education allowed states to alter over time how they calculated AYP designations and the conditions under which administrative interventions were required. Another reason for the variation is that states possessed significant discretion in measuring and defining student educational performance. For example, states set their own cut scores for identifying “proficient” students on state tests, which resulted in many states relaxing their proficiency requirements in order to help schools and districts avoid failing AYP. Moreover, working within general federal guidelines, states set their own yearly AYP proficiency schedule. Many states, for example, back-loaded dramatic increases in the proficiency rate required to meet AYP to the last few years leading up to 2014, in the hope that, before that time, federal policymakers would relax the requirement of 100 percent proficiency or lessen sanctions associated with failing to meet it.

These features of NCLB implementation led to dramatic year-to-year and within-state volatility in the performance standards to which schools and districts were held accountable. Figure 1 illustrates the volatile AYP proficiency requirements in Ohio, the focus of our empirical analysis. The figure plots the math and reading proficiency rates for Ohio students from 2004 through 2011, and it juxtaposes these proficiency rates with the climbing profi-

⁴See Manna (2011, p. 26-29) for a description of what these administrative interventions looked like.

ciency target required for making AYP along with the overall rate at which districts failed to make AYP. The figure also plots the reputable National Assessment of Education Progress (NAEP) proficiency rates in the same subject areas. As the figure indicates, between 2004 and 2005 — when AYP sanctions were set to kick in for schools that failed AYP two years in a row — Ohio adjusted its proficiency cut score from one that identified about 35 percent of students as proficient (similar to the NAEP level) to one that identified about three-quarters of students as proficient. After that, however, the overall proficiency rate used to determine AYP held steady.

[Insert Figure 1 here]

The figure also documents how, in accordance with federal requirements, the AYP proficiency rate targets increased over time. For example, a district with 65 percent of students demonstrating proficiency in reading would meet the district-wide AYP goal in 2005 but would fall short with identical performance in 2006. Yet, while initial increases in the required proficiency rates corresponded to an increasing number of districts falling short of AYP, the failure rate actually decreased dramatically in 2008 in spite of a growing proficiency target. The central reason for this divergence appears to be that Ohio was granted an exemption that enabled schools and districts to meet AYP requirements if the state projected that students would meet proficiency requirements down the road. In other words, tinkering and politically motivated schedules for increasing proficiency targets led to unpredictable variability in the rate of AYP failure over time. This volatility in district AYP designations occurred in spite of relatively constant levels of actual student proficiency as documented by scores on separate NAEP exams. Schools and districts moved in and out of their AYP “not met” designations for reasons that were difficult to anticipate and that were largely unrelated to changes in their performance in educating students.⁵

⁵In our sample, a significant number of districts moved both in *and* out of AYP failure status during this period. Overall, 63 percent of the districts in our sample saw their AYP status change at least once. Conditional on having AYP status change at least once, a district’s designation changed 3.4 times on average during the course of our full panel (including years with no levies on the ballot). In the full panel, we observe AYP status change compared to the previous year in 31.6% of district-years.

Although undesirable from the point of view of voters who might wish to rely on AYP as an accurate measure of school quality, the arbitrary nature of AYP provides an important source of causal identification in our design. That district AYP designations changed year-to-year due to changing standards and definitions helps ensure that they are not correlated with time-varying, difficult-to-measure contextual factors that may independently affect voters' subjective evaluations of local schools and that are not captured by district fixed-effects. In other words, an inherent problem with the AYP measure, from the point of view of using it as a tool to improve accountability, allows for an ideal empirical test of the impact of federal performance measures on voter behavior in local elections.

In addition to the problems described above, mathematics and reading proficiency rates — the key performance measures used as the basis of NCLB's accountability framework — are themselves invalid measures of school and district performance. To gauge how successful a school or district is in carrying out its mission to educate students, one must measure how much students actually learn during a school year. Proficiency rates, on the other hand, largely capture the aptitude students possess *prior* to entering school (see Downey, von Hippel and Hughes 2008). Thus, the most effective schools — those that post the biggest gains in student achievement — may have failed AYP simply because the students they serve started off at a lower base.⁶ Indeed, across all 50 states, there is a strong positive relationship between AYP failure and the proportion of a school's student population that is economically disadvantaged or members of a racial or ethnic minority group (Davidson et al., 2013). During our period of study, the probability of failing AYP in any given year was more than 0.67 among Ohio districts identified by the state as serving a high number of poor students, compared to 0.42 for wealthier districts. In other words, the risk of failure among high-poverty districts was nearly 60 percent greater than in other districts.

In the Supporting Information, we provide direct evidence that higher rates of AYP failure among economically disadvantaged districts were due primarily to lower levels of

⁶Note that this does not pose a problem for our design, because the socio-economic status of district students changes very slowly over time and is captured in statistical models featuring district fixed-effects.

baseline proficiency attributable to students’ socioeconomic status rather than the quality of their schools. One piece of evidence is the relationship between district AYP status and “value added” measures of district quality that Ohio introduced into its own, separate state accountability system in 2007. The value-added method compares the gains a district’s students made on state exams with those of students in other districts.⁷ Instead of absolute levels of proficiency, these value-added scores reward districts for growth in student performance. Districts whose students made lower-than-expected gains are labeled “below” and those whose students made expected gains or significantly exceeded expectations are labeled “met” and “above,” respectively.

[Insert Figure 2 here]

Figure 2 contrasts district AYP designations with Ohio’s value-added benchmarks for the years 2007 to 2012, the only period for which both measures are available. As the figure illustrates, there is only a weak relationship between whether a district meets federal AYP targets and how much progress its students post on state exams (Goodman-Kruskal $\gamma = 0.26$). Overall, about 19 percent of districts that met AYP over this period actually received the “below” value-added designation. By contrast, almost 70 percent of districts that fell short of AYP nevertheless met or exceeded their value-added benchmark.

Summary

In summary, the NCLB measurement approach suffered from three serious shortcomings that undermined its potential for helping voters accurately assess the performance of local school districts. First, the AYP designations were based in large part on arbitrary, changing, or politically determined benchmarks that varied across states and within states over time. Second, even ignoring these limitations, the underlying metric used for calculating AYP designations — absolute levels of student proficiency — was largely determined before students

⁷SAS Inc. generated the “value-added” estimates for Ohio. For technical reports describing the SAS EVAAS method for estimating school and district value-added, consult the SAS and ODE websites.

ever set foot in a classroom and only partially (if at all) reflected the effectiveness of their teachers or the quality of their schools.

Nevertheless, the federal benchmarks attracted substantial publicity and attention. That districts and schools failing to make AYP were required to notify parents of their status likely made this a particularly salient consideration in the minds of many voters. Although earlier research found only limited evidence that voters rely on test-score performance to hold incumbent education officials accountable at the ballot box (Berry and Howell, 2007), the prominence of annual AYP designations made it easy for voters to incorporate these signals into their retrospective judgments. Indeed, recent studies have shown strong evidence that voters update their beliefs about school quality in response to changing designations on widely disseminated official school “report cards” (e.g., Chingos, Henderson and West 2012; Jacobsen, Saultz and Snyder 2013).

Interestingly, it appears that few federal policymakers explicitly considered how the law’s reporting requirements would affect local politics. The issue did not come up at the five hearings held on education reform in 2000 and 2001—the congressional session during which NCLB was adopted—including two hearings held after its passage that focused on the law’s implementation.⁸

Empirical Strategy

We examine how the flawed AYP performance measure affected voter behavior in local elections.⁹ Specifically, our empirical analysis focuses on property tax levies considered by voters during the period that NCLB was operational in Ohio. Districts place such levies on the

⁸Speaking at a hearing on April 23, 2002, Sen. Judd Gregg (R-N.H.), did suggest that the law would “involve the empowerment of parents when they find their children in schools which have failed and which continue to fail.” Like other contemporary observers, however, Sen. Gregg envisioned that such empowerment would come primarily through “exit,” allowing parents to take advantage of the law’s school choice provisions to transfer their kids to better-performing schools, rather than “voice,” with local voters attempting to change local education policy through elections (Hirschman, 1970).

⁹Although NCLB required AYP designations for both schools and districts, we focus on the latter in our empirical analysis because levies are adopted at district level.

ballot regularly to fund their operating or capital needs. In Ohio, funding approved through such levies represents one-third of local school budgets, just above the national average.¹⁰ In addition to being a large state with many, diverse school districts — as well as a population that is demographically similar to the country as a whole — Ohio provides an ideal research setting due to state laws that prevent property taxes from growing automatically when property values increase, ensuring that school levies appear on the ballot quite frequently as district costs rise and new revenues are required. Many local levies also remain operational only during an exogenously determined period of time, requiring districts to return to the voters with replacement levies. These institutional features allow us to rule out the possibility that districts can be too strategic in their timing of levy proposals¹¹ and provide us with a sufficient number of observations (nearly 4,000 levy votes from 2003 to 2012) to estimate within-district effects over time.

Data

There are currently 611 school districts in Ohio, although two additional districts were operational during at least one year in our period of study, 2003 to 2012. The empirical analysis employs data on the 556 districts that placed at least one levy on the ballot during these years. Districts not included in the analysis were primarily those that operate at Ohio’s minimum statutory tax rate floor, set at 20 mills.¹² Importantly, districts from all of Ohio’s major metropolitan areas are included in the analysis.

The data employed in the analysis are publicly available and were obtained from the websites of a number of organizations. AYP designations, state performance designations, and student enrollment and demographic characteristics came from the Ohio Department of

¹⁰Ohio district financial forecast tables available on the Ohio Department of Education website indicate that districts relied on local property taxes for 33.3 percent of their total revenue during the 2008-2011 period. A survey conducted in 2004 revealed that, on average, school districts across the country relied on local property taxes for 29 percent of total funding (McGuire and Papke, 2008).

¹¹We show in the Supporting Information that there is no evidence of AYP failure affecting the timing or size of levy proposals.

¹²A 1 mill levy raises \$1 per \$1,000 of assessed value.

Education. The variables that capture student demographic characteristics — percent poor, percent black, percent Hispanic, percent English language learners, and percent disabled — were created according to federal guidelines established for the purpose of disaggregating academic achievement statistics by student subgroup. District expenditure data were taken from the National Center for Educational Statistics while the revenue and property values data are from the Ohio Department of Taxation. Details about each levy, including the electoral outcome, were obtained from the Ohio School Board Association. The remaining variables, including rates of home ownership and the proportion of school-age students attending private schools, were extracted from the American Community Survey.¹³

In addition to student demographic characteristics and several measures of district finances, our full specifications also control for each district’s rating under Ohio’s own school accountability system. These rankings incorporated a number of measures of student performance, including many of the metrics used to calculate AYP. Starting in 2008, Ohio’s rating system also incorporated district value-added benchmarks. Unlike the dichotomous federal AYP designation, Ohio places districts into one of six categories, ranging from “Academic Emergency” to “Excellent with Distinction.”

The purpose of including covariates capturing district and community demographics, property values, existing millage rates, and district expenditures is to account for factors that may affect or reflect voters’ inclination to approve school levies. The set of covariates we employ is unusually complete, capturing factors known to affect voting in local elections and attitudes toward tax levels. Nevertheless, our preferred specifications also include year and district fixed-effects to capture potentially omitted factors, and we provide direct evidence of the causal mechanisms to further rule out the possibility of spurious results.¹⁴

¹³Because these Census variables are time-invariant, we exclude them when reporting the fixed-effects specifications.

¹⁴The district fixed-effects design, which leverages *change* in AYP status for identification, also accounts for pre-existing differences in voter beliefs about their district quality, which may be correlated with AYP status in any given year.

Hypotheses

The analysis examines the effect of districts receiving a designation of “AYP not met,” as opposed to “AYP met,” on the probability of levy passage.¹⁵ It may seem ambiguous how AYP failure should shape voter support for local school funding. One might expect, for example, that failure would increase the willingness of local taxpayers to fund local schools at higher levels, either as a sincere response meant to improve student outcomes or a self-interested goal of protecting their property values. Research indicates that voters do not respond to performance information in this way, however. Few voters appear to make a coherent connection between the taxes they pay and the quality of services they expect to receive (e.g., Beck, Rainey and Traut 1990; Benton and Daley 1992; Lowery and Sigelman 1981). Many, it seems, expect “something for nothing” (Sears and Citrin, 1982). Studies consistently show that support for tax increases and other financial contributions is positively associated with citizen satisfaction with existing public services (e.g., Collins and Kim 2009; Figlio and Kenny 2009; Glaser and Hildreth 1999; Simonsen and Robbins 2003). Consequently, we expect that a signal of negative school district performance indicated by AYP failure reduces voter support for proposed school levies.

In addition, we hypothesize that the timing of the dissemination of annual AYP designations affects the extent to which voters incorporate this information into their decision-making. Both theoretical and empirical literature on voting decisions illustrate how voters’ use of performance information is subject to behavioral biases, such as myopia. Recent performance information, such as macroeconomic outcomes in the six months before an election, has a disproportionate effect on electoral outcomes (e.g., Healy and Lenz 2014; Huber, Hill and Lenz 2012; Kramer 1971). We expect voters to exhibit similar myopia in our local context. In particular, we hypothesize that voters put greater weight on a district’s most recent AYP designation and that this information is most electorally relevant when the AYP

¹⁵In other words, the comparison is between a district failing and meeting the AYP requirement. This should be kept in mind when interpreting the effects substantively.

designations are released in close temporal proximity to the levy election. Empirical support for these hypotheses would also provide some evidence that district AYP status affects levy passage through voter attitudes rather than alternative mechanisms, such as the strategic use of federal performance measures by interest groups campaigning against local tax increases.

Finally, we hypothesize that the magnitude of the electorate’s response to the first AYP failure is greater than that of subsequent failures. This hypothesis is based on Bayesian learning models of voter behavior (Achen, 1992; Gerber and Green, 1998).¹⁶ These models indicate that initial information shocks produce a larger shift in beliefs than additional information reinforcing these messages.

Statistical Model

In the body of the paper, we examine the empirical relationship between AYP failure and the probability of levy passage.¹⁷ Our basic specification models levy passage probability as a linear function of AYP failure, a large battery of district covariates, and year fixed effects, γ_t . The year fixed effects allow statewide shocks in voter willingness to support taxes to evolve nonparametrically over the course of our sample period. Formally, our base specification is:

$$y_{it} = \alpha + \tau \text{AYP Failure}_{it} + \beta \text{Covariates}_{it} + \gamma_t + \epsilon_{it}$$

In addition to our rich set of district covariates that include demographic, economic, and educational data for all districts, we further account for unobserved time-invariant school district characteristics by estimating district fixed-effects models. The fixed effects specifi-

¹⁶As a complementary mechanism, local media organizations may cover the first incidence of AYP failure with a great deal of intensity and dismiss subsequent failures as old news that is less worthy of media coverage.

¹⁷We focus on the dichotomous dependent variable of levy passage or failure, rather than the percent of votes cast in support of each levy, because we are interested in the substantive policy consequences of district AYP status. For robustness, however, we replicate our main results using the percent of “yes” votes cast for each levy. The results, which we report in the Supporting Information, are substantively the same as those of the dichotomous models.

cations can be written as the linear model:

$$y_{it} = \alpha_i + \tau \text{AYP Failure}_{it} + \beta \text{Covariates}_{it} + \gamma_t + \epsilon_{it}$$

where α_i are the district fixed effects. To remove the α_i nuisance parameters from the estimation, we transform the data through demeaning and estimate an OLS regression on the demeaned data. The fixed-effect specifications only exploit within-district variation so it is less plausible that unobserved district characteristics are falsely driving the association between AYP status and levy passage probability.¹⁸ In all of our specifications, we report heteroskedasticity robust standard errors that are clustered at the district level.

After evaluating the relationship between AYP failure and levy passage probability, we directly test the more subtle informational predictions about voter response to AYP failure. To investigate the possibility of myopia in voters' use of district performance information, we first estimate the effect of AYP failure in the previous year on levy passage in the current year. In our OLS specification, we estimate the model:

$$y_{it} = \alpha + \tau \text{AYP Failure}_{it-1} + \beta \text{Covariates}_{it} + \gamma_t + \epsilon_{it}$$

A second test of the myopia hypothesis is to determine if AYP failure has a larger effect for levy elections held in the fall, shortly after district designations are released in late August, as opposed to earlier in the year. In the November elections, voters have just learned whether their school district met or failed AYP. For elections that take place earlier in the calendar year, voters have had a longer period of time to forget whether the school district met AYP. In Ohio, general elections are held every November — for state and federal offices in even-

¹⁸It is important to note that the fixed effects also account for other aspects of the local political context — such as the strength of the local teachers' union, for example — that do not change over the duration of our panel.

numbered years and local offices in odd-numbered years — while primary elections occur in either May (for odd-numbered years) or March (for even-numbered years).¹⁹

We then investigate how voters respond to the informational shock that occurs when a school district fails to meet the AYP standard for the first time. As discussed above, voters’ attitudes toward and evaluations of the school district might exhibit a more pronounced response the first year a school district fails AYP compared to subsequent years. We model this hypothesis as an additive effect in which the coefficient on the first instance of AYP failure is allowed to differ from the coefficient on subsequent instances of AYP failure. Specifically, we use the specification:

$$y_{it} = \alpha + \tau_1 \text{First AYP Failure}_{it} + \tau_2 \text{Subsequent AYP Failures}_{it} + \beta \text{Covariates}_{it} + \gamma_t + \epsilon_{it}$$

Note that we can separately identify τ_1 and τ_2 due to the presence of districts that fail to meet AYP in multiple years in our sample.

To further test the credibility of our estimates, we conduct placebo tests using future realizations of AYP failure as our independent variable of interest. Future years’ AYP status is clearly not known to voters when they cast their ballots and should have no effect on levy passage once we include other district characteristics in our specification. For these placebo tests, we estimate variants of the following specification:

$$y_{it} = \alpha + \tau \text{AYP Failure}_{it+1} + \beta \text{Covariates}_{it} + \gamma_t + \epsilon_{it}$$

If the results are driven by unobserved factors that are correlated with AYP failure and citizens’ vote choices, we would expect the placebo test to also generate a spurious association

¹⁹School districts also occasionally place levies on the ballot during special elections held in February and August, although there are very few such proposals in our dataset.

between future AYP failure and levy passage probability.

Results

[Insert Table 1 here]

The base linear probability specifications are reported in Table 1. The first four columns report cross-sectional results with an increasingly large battery of controls that account for both district-level and levy-level factors. The fifth column reports the fixed-effects specification with the complete battery of controls. In all models, AYP failure is associated with a large and statistically significant decrease in levy passage probability, ranging from approximately 4.9 to 9 percentage points.²⁰ In the fixed-effects specification, which provides the most convincing identification of a causal effect, AYP failure decreases levy passage probability by approximately 4.9 percentage points.²¹ With voters approving approximately 52 percent of the levies in our sample, this translates into a substantively meaningful effect that has important implications for public school finance (see the Supporting Information). Because the probability of AYP failure was more than 50 percent greater in high-poverty districts than in low-poverty districts — and because the former also started off with lower levels of local funding — the federal designation exacerbated the disparity in local tax resources between advantaged and disadvantaged districts.²²

²⁰The probability of levy passage in districts that met the AYP standard was 0.57. This conditional mean probability of levy passage was much higher than in districts where the AYP standard was not met (0.48), but it is sufficiently far from 1 to suggest that ceiling effects do not bias our estimates.

²¹In the Supporting Information, we also investigate whether our findings hold once we exclude the lowest-performing districts. The results are quite similar to our main findings and provide no evidence that voters in districts with more moderate academic outcomes respond differently to the news of AYP failure.

²²As a robustness check, we account for the binary nature of the dependent variable by estimating these specifications with both probit and conditional logit models instead of the linear probability model. The results are presented in the Supporting Information and are substantively unchanged. We also estimated two additional specifications, interacting the proposed millage with the property values in the district and allowing the effect of AYP failure to vary with levy size. The interaction between millage size and property taxes captures the absolute, out-of-pocket cost of each levy for the typical voter in the district. The effects of AYP in both specifications are substantively unchanged, with both interaction terms failing to reach conventional levels of statistical significance. Finally, we re-estimated the models using only district observations after 2007, the years for which value-added test scores are available. In this subset, the effect of failing to meet

[Insert Table 2 here]

We next turn to investigating the more subtle informational predictions about voter responses to AYP failure. We first examine how voters incorporate AYP status from the previous year in their voting decisions. The first column of Table 2 reports a cross-sectional regression with the full battery of controls and the indicator for AYP failure in the previous year as the independent variable of interest. The point estimate is close to zero and statistically insignificant, which is consistent with evidence that voters' retrospective evaluations of governmental performance tend to discount information from the past. In the next column, we add district fixed effects to the specification and reach a similar conclusion. In the final two columns, we report finite-distributed lag specifications that include both lagged AYP failure and contemporaneous AYP failure as independent variables of interest. The third column is cross-sectional while the fourth column includes district fixed effects. In both specifications, the coefficient estimate on lagged AYP failure is statistically insignificant. Strikingly, the coefficient estimate on contemporaneous AYP failure is negative, large in magnitude, and statistically significant, further corroborating our primary result. The fixed-effects specification indicates that failing AYP within the past year results in a decrease of 6.9 percentage points in the probability of levy passage.

[Insert Table 3 here]

To provide a second test of voter myopia we separately examine the response to AYP failure in elections held in November and those scheduled earlier in the year. Nearly all levy elections held prior to November take place in either the summer or spring, prior to the dissemination of district report cards in late August. For most non-November levies, the latest AYP information was released at least seven months prior. Theories of myopia predict that voters will respond more intensely to performance information disseminated in close proximity to the election.

AYP was negative, significant, and substantively large. By contrast, the less salient but arguably more valid value-added scores have no electoral consequences.

The first column of Table 3 restricts the sample to elections held before November. The point estimate for AYP failure is still negative, although it is much smaller in magnitude and is not statistically distinguishable from zero at conventional significance levels. The second column restricts the sample to November elections and shows that the coefficient estimate for AYP failure is larger in magnitude than in our pooled specifications and that it is statistically significant. The third and fourth columns report the results of models that include the entire sample and an interaction between AYP failure and a variable indicating whether or not the election was held in November.²³ In both specifications, the joint effect of AYP failure on levy passage probability is more than 7 percentage points when the election is held in November. While this is not quite statistically distinguishable from the effect for elections held in other months, the point estimates are consistent with the myopia documented using lagged realizations of AYP. Because the composition of the electorate changes between elections in a way that might moderate the effect of performance information, these results are merely suggestive.

[Insert Table 4 here]

As predicted by Bayesian theories of voter behavior, voters may respond differently to the first incidence of AYP failure compared to subsequent failures. To investigate this possibility, we create two new indicator variables capturing the first incidence of AYP failure and an indicator of subsequent failures, and we include these as independent variables in our models. While the statistical power of this test is quite limited, especially in the district fixed-effects specifications (because we can only observe at most one first incidence of AYP failure for each district) the results reported in Table 4 suggest that the first AYP failure has a particularly large effect on the likelihood of levy passage. We cannot statistically distinguish the coefficient estimates on the first and subsequent AYP failure variables, but we do find that the magnitudes of the point estimates are larger for the first AYP failure. The results

²³We also include an indicator for November elections in the specification to control for unobservable differences in the probability of passing a levy across election months.

are consistent with a mechanism whereby voters react more sharply to the new, perhaps surprising information about district performance than to subsequent reminders of the same information.

[Insert Table 5 here]

Finally, we conduct a placebo test that uses *next year's* AYP status as the independent variable of interest. We report two fixed-effects specifications. In the first column of Table 5, we report estimates of the model with future AYP failure as the only AYP variable in the specification. In column two, we report estimates with both future AYP failure and current AYP failure as independent variables. In both, we fail to reject the null hypothesis that future AYP failure is uncorrelated with levy passage in the current year. These null placebo test results corroborate the credibility of our research design and increase our confidence in the validity of the results.

Conclusion

The touted goals of the federal No Child Left Behind Act were to promote academic achievement and close the achievement gap between student subgroups. This study documents how the flaws in the construction and use of NCLB's AYP performance metric may have undermined these objectives. First, the dissemination of AYP designations is unlikely to have promoted academic achievement because the metric did not accurately reflect the quality or performance of local school districts. As this study illustrates, district AYP status served as a proxy for poverty and was only minimally correlated with Ohio's value-added measure of student achievement in math and reading (the subject areas emphasized in the AYP metric).

Second, the response of local voters to districts' AYP designations undermined the goal of narrowing the achievement gap between economically advantaged and disadvantaged students. The probability of passing a school district levy declined by more than 10 percent

when districts failed AYP, resulting in a large financial penalty that disproportionately affected districts in impoverished communities and only widened the resource disparity between districts. Recent research indicates that this likely affected student achievement. For example, Jackson, Johnson and Persico (2014) document the importance of financial resources for promoting student achievement in impoverished districts, and Cohen and Moffitt (2009) conclude that limited resources are largely responsible for these districts struggling to implement federal educational programs meant to close the achievement gap.

Our findings also shed light on an important debate in the education policy community about how voters respond to information about school performance. Many proponents of education accountability reforms argue that performance information can help empower parents and target remedial efforts toward schools and districts facing the greatest achievement hurdles. By contrast, some NCLB skeptics have argued that the law actually had the opposite effect, undermining voter support for public education by presenting local schools in a negative light (e.g., Ladd 2012, p. 215). Our study provides evidence that is more consistent with the latter perspective. Our findings suggest that, at least among districts failing AYP, the law's performance benchmarks affected voter attitudes toward local school districts in a negative way and led to a reduction in district resources available to pursue academic missions. To the extent that levy failures leave lasting negative effects on academic achievement, the political impact of the federal law appears to have undermined, rather than supported, struggling districts serving low-performing student populations.²⁴

This study also documents the importance of considering the validity of performance measures when examining their use by voters. Invalid measures may not be too problematic when decision-makers are highly informed and understand how much weight to assign to them. The performance management literature indicates that those who manage public agencies — and, therefore, have access to large amounts of quantitative and qualitative in-

²⁴We focus only on the public finance implications, but it is possible that AYP may have also affected parent willingness to donate to schools (Figlio and Kenny, 2009) and to volunteer or serve on local PTAs for example.

formation about their own organizations — are less inclined to use performance information in making management decisions if they perceive that information to be invalid (Moynihan and Lavertu, 2012). Voters, however, may lack other sources of information about public agencies and may simply assume that a summary designation (e.g., that a school has “met” AYP standards) provides a valid characterization of performance. The performance measurement literature has examined the perverse organizational responses to invalid measures (Heinrich and Marschke 2010) but there has been very little examination of the consequences when external and relatively uninformed audiences are exposed to invalid measures.

Finally, our findings highlight an important but frequently overlooked aspect of inter-governmental relations. In federal systems, different levels of government influence each other not only directly — through the adoption of policies that alter the set of feasible options available to other jurisdictions — but also indirectly through the electoral channel, by changing the types of considerations voters bring to the polls and the salience of individual issues. Although the fact that presidential approval affects election outcomes in state and local elections is well established (e.g., Carsey and Wright 1998), our study suggests that there exist many other political interdependencies between governments with overlapping constituencies. Policy actions by one level of government can thus produce electoral reactions in another, with consequences that may not be anticipated by officials at either level.

Of course, our results do not represent a general indictment of performance measurement and its application in federal contexts. Given low levels of voter knowledge, it is possible that well-designed and clearly explained measures may improve democratic accountability and transparency, providing constituents with important information about the performance of their government and offsetting other detrimental biases in voter behavior. But our study offers lessons that should inform the design of such rating systems. First, we show that the effects of performance information diminish substantially after the initial informational “shock.” Second, we provide suggestive evidence that the impact of district AYP failure on voter decision-making was contingent on the timing of local elections. Both sets of findings

suggest that the effects are short-lived and point to the need for performance measures to incorporate information over longer time periods to counteract well-established cognitive biases that lead to voter myopia. These findings also suggest that national variation in local political institutions and informational contexts complicates the use of performance measurement as a federal policy and management tool. Those who design performance measurement systems should keep in mind the potentially heterogeneous effects of these schemes in light of how voters use information. Finally, it is important to take account of how federal goals map onto local policy objectives, as local governments may experience goal displacement if performance measures cause local stakeholders to focus only on dimensions of import to federal overseers.

The measurement of government performance and the wide dissemination of performance information are likely to continue to proliferate in tandem with rapid advances in statistics and technology. Our results highlight the potential downside of such performance federalism, illustrating how flawed measures may undermine political accountability in state and local democracy and undercut the very policy objectives these accountability systems are designed to achieve. Researchers should investigate the political and democratic impacts of performance federalism so that these significant issues may be sorted out.

References

- Achen, Christopher H. 1992. "Breaking the Iron Triangle: Social Psychology, Demographic Variables and Linear Regression in Voting Research." *Political Behavior* 14(3):195–211.
- Beck, Paul Allen, Hal G. Rainey and Carol Traut. 1990. "Disadvantage, Disaffection, and Race as Divergent Bases for Citizen Fiscal Policy Performance." *Journal of Politics* 52(1):71–93.
- Benton, J. Edwin and John L. Daley. 1992. "The Paradox of Citizen Service Evaluations and Tax/Free Preferences: The Case of Two Small Cities." *American Review of Public Administration* 22(4):271–287.
- Berry, Christopher R. and William G. Howell. 2007. "Accountability and Local Elections: Rethinking Retrospective Voting." *Journal of Politics* 69(3):844–858.
- Bertelli, Anthony M. and Peter John. 2010. "Government Checking Government: How Performance Measures Expand Distributive Politics." *Journal of Politics* 72(2):545–558.
- Boyne, George A., Oliver James, Peter John and Nicolai Petrovsky. 2009. "Democracy and Government Performance: Holding Incumbents Accountable in English Local Governments." *Journal of Politics* 71(4):1273–1284.
- Brown, Adam R. 2010. "Are Governors Responsible for the State Economy? Partisanship, Blame, and Divided Federalism." *Journal of Politics* 72(3):605–615.
- Carsey, Thomas M. and Gerald C. Wright. 1998. "State and National Factors in Gubernatorial and Senatorial Elections." *American Journal of Political Science* 42(3):994–1002.
- Chingos, Matthew M., Michael Henderson and Martin R. West. 2012. "Government Service Quality: Evidence from Public Schools." *Quarterly Journal of Political Science* 7(4):411–445.

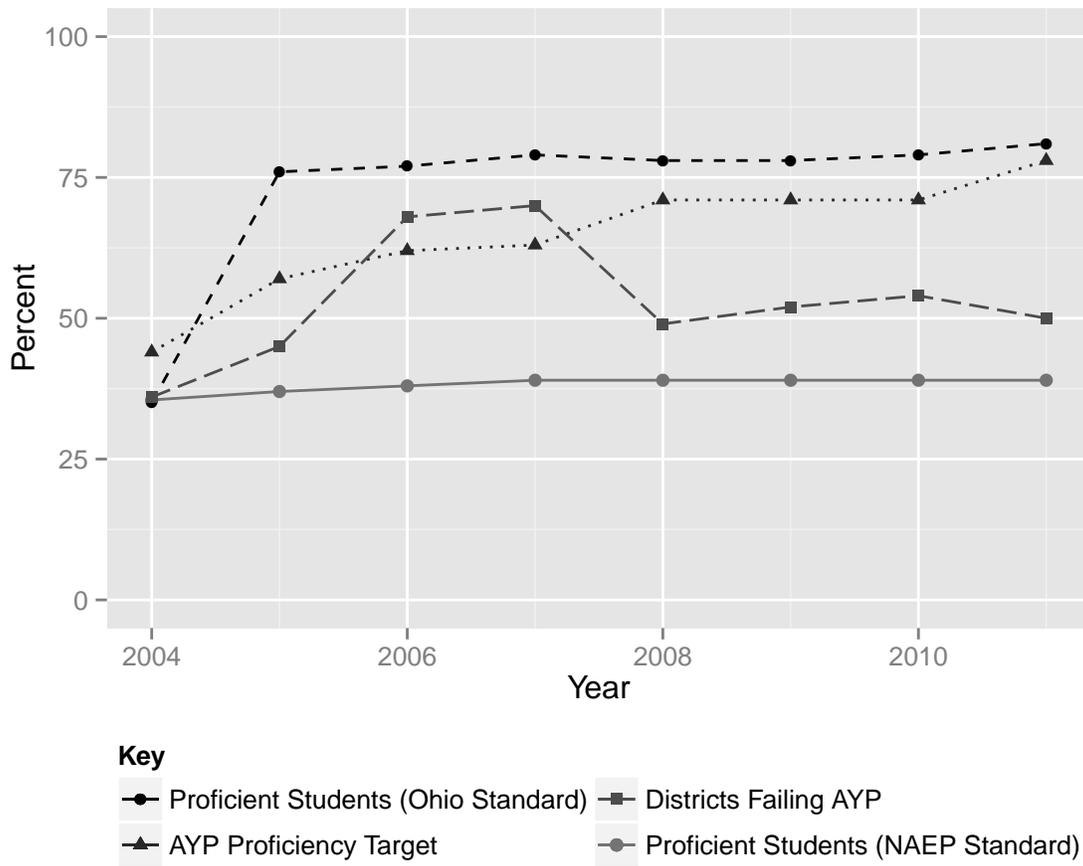
- Cohen, David K. and Susan L. Moffitt. 2009. *The Ordeal of Equality: Did Federal Regulation Fix the Schools?* Cambridge, MA: Harvard University Press.
- Collins, Brian K. and Hyun Joon Kim. 2009. "Are Satisfied Citizens Willing to Pay More? Public Sector Consumerism as Equitable Social Exchange." *Public Money and Management* 29(2):109–116.
- Conlan, Timothy. 1998. *From New Federalism to Devolution: Twenty-Five Years of Intergovernmental Reform*. Washington, DC: Brookings Institution.
- Davidson, Elizabeth, Randall Reback, Jonah Rockoff and Heather L. Schwartz. 2013. "Fifty Ways to Leave a Child Behind: Idiosyncrasies and Discrepancies in States' Implementation of NCLB." NBER Working Paper No. 18988.
- Downey, Douglas B., Paul T. von Hippel and Melanie Hughes. 2008. "Are 'Failing' Schools Really Failing? Using Seasonal Comparison to Evaluate School Effectiveness." *Sociology of Education* 81(3):242–270.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York, NY: Harper.
- Figlio, David N. and Lawrence W. Kenny. 2009. "Public Sector Performance Measurement and Stakeholder Support." *Journal of Public Economics* 93(9-10):1069–1077.
- Gerber, Alan and Donald P. Green. 1998. "Rational Learning and Partisan Attitudes." *American Journal of Political Science* 42(3):794–818.
- Glaser, Mark A. and W. Bartley Hildreth. 1999. "Service Delivery Satisfaction and Willingness to Pay Taxes: Citizen Recognition of Local Government Performance." *Public Productivity and Management Review* 23(1):48–67.
- Healy, Andrew and Gabriel S. Lenz. 2014. "Substituting the End for the Whole: Why Voters Respond Primarily to the Election-Year Economy." *American Journal of Political Science* 58(January):31–47.

- Heckman, James J., Carolyn J. Heinrich, Pascal Courty, Gerald Marschke and Jeffrey Smith. 2011. *The Performance of Performance Standards*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- Heinrich, Carolyn J. and Gerald Marschke. 2010. "Incentives and Their Dynamics in Public Sector Performance Management Systems." *Journal of Policy Analysis and Management* 29(1):183–208.
- Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty*. Cambridge, MA: Harvard University Press.
- Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *Bell Journal of Economics* 10(1):74–91.
- Huber, Gregory A., Seth J. Hill and Gabriel S. Lenz. 2012. "Sources of Bias in Retrospective Decision Making: Experimental Evidence on Voters' Limitations in Controlling Incumbents." *American Political Science Review* 106:720–741.
- Jackson, C. Kirabo, Rucker Johnson and Claudia Persico. 2014. "The Effect of School Finance Reforms on the Distribution of Spending, Academic Achievement, and Adult Outcomes." National Bureau of Economic Research Working Paper No. 201183.
- Jacobsen, Rebecca, Andrew Saultz and Jeffrey W. Snyder. 2013. "When Accountability Strategies Collide: Do Policy Changes That Raise Accountability Standards Also Erode Public Satisfaction?" *Education Policy* 27(2):360–389.
- James, Oliver and Peter John. 2007. "Public Management at the Ballot Box: Performance Information and Electoral Support for Incumbent English Local Governments." *Journal of Public Administration, Research, and Theory* 17(4):567–580.
- Kramer, Gerald H. 1971. "Short-Term Fluctuations in U.S. Voting Behavior, 1896-1964." *American Political Science Review* 65(1):131–143.

- Ladd, Helen F. 2012. "Education and Poverty: Confronting the Evidence." *Journal of Policy Analysis and Management* 31(2):203–227.
- Lau, Richard R. and David P. Redlawsk. 2001. "Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making." *American Journal of Political Science* 45(3):951–971.
- Lorinkas, R.A., B.W. Hawkins and S.D. Edwards. 1969. "The Persistence of Ethnic Voting in Urban and Rural Areas: Results from a Controlled Election Method." *Social Science Quarterly* 49(3):491–499.
- Lowery, David and Lee Sigelman. 1981. "Understanding the Tax Revolt: Eight Explanations." *American Political Science Review* 75(4):963–974.
- Malhotra, Neil and Alexander G. Kuo. 2008. "Attributing Blame: The Public's Response to Hurricane Katrina." *Journal of Politics* 70(1):120–135.
- Manna, Paul. 2006. *School's In: Federalism and the National Education Act*. Washington, DC: Georgetown University Press.
- Manna, Paul. 2011. *Collision Course: Federal Education Policy Meets State and Local Realities*. Washington, DC: CQ Press.
- Matson, Marsha and Terri Susan Fine. 2006. "Gender, Ethnicity, and Ballot Information: Ballot Cues in Low-Information Elections." *State Politics and Policy Quarterly* 6(1):49–71.
- McDermott, Monika L. 2005. "Candidate Occupations and Voter Information Shortcuts." *Journal of Politics* 67(1):201–219.
- McGuire, Therese J. and Leslie E. Papke. 2008. The Local Funding of Schools: The Property Tax and Its Alternatives. In *Handbook of Research in Education Finance and Policy*, ed. Helen F. Ladd and Edward B. Fiske. New York: Routledge.

- Moynihan, Donald P. and Stéphane Lavertu. 2012. “Does Involvement in Performance Reforms Encourage Performance Information Use? Evaluating GPRA and PART.” *Public Administration Review* 72(4):592–602.
- Oates, Wallace. 1972. *Fiscal Federalism*. New York: Harcourt Brace Jovanovich.
- Peterson, Paul E. and Mark Paul. 1989. “Federalism, Welfare Policy, and Residential Choices.” *American Political Science Review* 83(3):711–728.
- Pollitt, Christopher and Geert Bouckaert. 2004. *Public Management Reform: A Comparative Analysis, Second Edition*. Oxford: Oxford University Press.
- Schaffner, Brian F., Matthew Streb and Gerald Wright. 2001. “Teams Without Uniforms: The Nonpartisan Ballot in State and Local Elections.” *Political Research Quarterly* 54(1):7–30.
- Scholz, John T. and Feng Heng Wei. 1986. “Regulatory Enforcement in a Federalist System.” *American Political Science Review* 80(4):1249–1270.
- Sears, David O. and Jack Citrin. 1982. *Tax Revolt: Something for Nothing in California*. Cambridge, MA: Harvard University Press.
- Simonsen, Bill and Mark D. Robbins. 2003. “Reasonableness, Satisfaction, and Willingness to Pay Property Taxes.” *Urban Affairs Review* 36(6):831–854.
- Volden, Craig. 2005. “Intergovernmental Political Competition in American Federalism.” *American Journal of Political Science* 49(2):327–342.
- Walker, David B. 1995. *The Rebirth of Federalism: Slouching Toward Washington*. New York: Chatham House Publishers.

Figure 1: Trends in Proficiency Rates, AYP Proficiency Targets, and AYP Failure, 2004-2011.



Note: The Ohio proficiency rate tracks the proportion of students meeting the state’s cutoff for proficiency in both math and reading. The NAEP proficiency rate is the average for both math and reading for 4th and 8th grade students. The AYP proficiency requirement is the aggregate required proficiency rate, which is created by weighting the proficiency targets across different grades and tests. The district AYP failure rate reflects the proportion of districts that failed AYP.

Figure 2: Comparison of AYP Performance with District Value-Added Designations, 2007-2011.

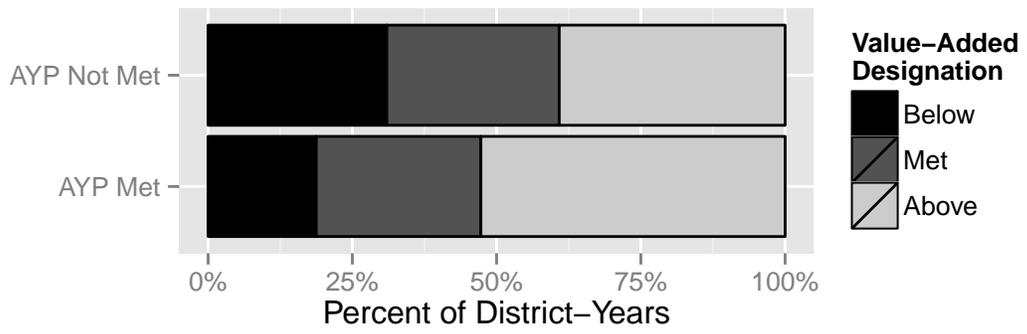


Table 1: Linear Probability Models for Levy Passage

	(1) Pass	(2) Pass	(3) Pass	(4) Pass	(5) Pass
AYP Failure	-0.0898*** (0.0173)	-0.0643*** (0.0179)	-0.0627*** (0.0176)	-0.0526** (0.0181)	-0.0487* (0.0197)
Current Millage		0.0150*** (0.00177)	0.0168*** (0.00181)	0.0169*** (0.00181)	0.0356*** (0.00428)
Property Values (per pupil)		0.00324 (0.00244)	0.00486 (0.00249)	0.00484 (0.00247)	0.0249* (0.0113)
Tot. Expenditures (per pupil)		-0.0672*** (0.0167)	-0.0600*** (0.0167)	-0.0607*** (0.0167)	-0.0667* (0.0306)
Instructional Expenditures (per pupil)		0.0915** (0.0295)	0.0701* (0.0294)	0.0655* (0.0292)	-0.0223 (0.0521)
Black (%)		0.000818 (0.000969)	0.000550 (0.00102)	0.000372 (0.00110)	-0.00380 (0.00515)
Hispanic(%)		0.00828** (0.00272)	0.00698** (0.00252)	0.00661* (0.00257)	0.0113 (0.0153)
English-Learners (%)		-0.000620 (0.00208)	0.000502 (0.00247)	0.000514 (0.00240)	-0.0118 (0.0202)
Poor (%)		-0.000439 (0.000854)	-0.000522 (0.000902)	-0.0000994 (0.000920)	0.00200 (0.00142)
Disabled (%)		0.000549 (0.00348)	-0.000919 (0.00343)	0.000311 (0.00343)	-0.00284 (0.00692)
District Enrollment (natural log)		-0.0579*** (0.0147)	-0.0565*** (0.0147)	-0.0630*** (0.0147)	-0.362 (0.222)
Mills			-0.0245*** (0.00320)	-0.0247*** (0.00320)	-0.0203*** (0.00380)
Private School Enrollment (%)			-0.00813 (0.00522)	-0.00832 (0.00514)	
Owner-Occupied Residence (%)			-0.00363* (0.00142)	-0.00380** (0.00141)	
Constant	0.510*** (0.0376)	0.226* (0.0882)	0.631*** (0.160)	0.819*** (0.201)	0.624 (0.493)
Observations	3990	3988	3955	3955	3986
R ²	0.016	0.054	0.079	0.082	0.066
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	No	No	No	No	Yes
Academic Designations	No	No	No	Yes	Yes

Heteroskedasticity robust standard errors clustered at the district level in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Lagged AYP Status

	(1) Pass	(2) Pass	(3) Pass	(4) Pass
Lagged AYP Failure	-0.00880 (0.0197)	0.00754 (0.0213)	0.0335 (0.0259)	0.0427 (0.0262)
AYP Failure			-0.0716** (0.0242)	-0.0690** (0.0246)
Current Millage	0.0171*** (0.00182)	0.0357*** (0.00428)	0.0170*** (0.00181)	0.0357*** (0.00426)
Property Values (per pupil)	0.00458 (0.00247)	0.0245* (0.0113)	0.00484* (0.00246)	0.0248* (0.0112)
Tot. Expenditures (per pupil)	-0.0581*** (0.0166)	-0.0672* (0.0306)	-0.0608*** (0.0167)	-0.0687* (0.0305)
Instructional Expenditures (per pupil)	0.0630* (0.0291)	-0.0179 (0.0520)	0.0662* (0.0292)	-0.0176 (0.0520)
Black (%)	0.000365 (0.00111)	-0.00439 (0.00500)	0.000368 (0.00110)	-0.00395 (0.00509)
Hispanic(%)	0.00665** (0.00254)	0.0105 (0.0154)	0.00663* (0.00258)	0.0105 (0.0153)
English-Learners (%)	0.000649 (0.00236)	-0.0111 (0.0196)	0.000467 (0.00239)	-0.0113 (0.0201)
Poor (%)	-0.000270 (0.000923)	0.00189 (0.00141)	-0.000176 (0.000921)	0.00188 (0.00142)
Disabled (%)	0.0000468 (0.00345)	-0.00321 (0.00696)	0.000210 (0.00344)	-0.00298 (0.00692)
District Enrollment (natural log)	-0.0696*** (0.0144)	-0.355 (0.221)	-0.0658*** (0.0146)	-0.364 (0.217)
Mills	-0.0248*** (0.00319)	-0.0203*** (0.00380)	-0.0245*** (0.00319)	-0.0203*** (0.00379)
Private School Enrollment (%)	-0.00804 (0.00515)		-0.00809 (0.00518)	
Owner-Occupied Residence (%)	-0.00378** (0.00141)		-0.00380** (0.00140)	
Constant	0.769*** (0.202)	0.566 (0.494)	0.829*** (0.202)	0.634 (0.488)
Observations	3955	3986	3955	3986
R^2	0.080	0.064	0.083	0.067
Year Fixed Effects	Yes	Yes	Yes	Yes
District Fixed Effects	No	Yes	No	Yes
Academic Designations	Yes	Yes	Yes	Yes

Heteroskedasticity robust standard errors clustered at the district level in parentheses
 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: November vs. Earlier Elections

	(1)	(2)	(3)	(4)
	Pre-November Elections	November Elections	Full	Full
AYP Failure	-0.0295 (0.0258)	-0.0624* (0.0259)	-0.0757** (0.0247)	-0.0714** (0.0253)
Pre-November Election*AYP Failure			0.0469 (0.0330)	0.0468 (0.0330)
Pre-November Election			-0.0427 (0.0238)	-0.0438 (0.0235)
Current Millage	0.0152*** (0.00239)	0.0190*** (0.00255)	0.0168*** (0.00181)	0.0348*** (0.00423)
Property Values (per pupil)	0.00517 (0.00305)	0.00472 (0.00346)	0.00511* (0.00247)	0.0247* (0.0113)
Tot. Expenditures (per pupil)	-0.0373 (0.0231)	-0.0827*** (0.0247)	-0.0614*** (0.0166)	-0.0693* (0.0307)
Instructional Expenditures (per pupil)	0.0201 (0.0415)	0.100* (0.0416)	0.0655* (0.0293)	-0.0212 (0.0525)
Black (%)	-0.00132 (0.00136)	0.00174 (0.00134)	0.000360 (0.00111)	-0.00416 (0.00508)
Hispanic(%)	0.00823 (0.00497)	0.00672 (0.00396)	0.00667* (0.00261)	0.0135 (0.0152)
English-Learners (%)	0.00214 (0.00349)	-0.00193 (0.00479)	0.000659 (0.00241)	-0.0111 (0.0199)
Poor (%)	0.000487 (0.00128)	-0.000717 (0.00124)	-0.0000501 (0.000914)	0.00219 (0.00142)
Disabled (%)	-0.00701 (0.00458)	0.00792 (0.00470)	0.000224 (0.00342)	-0.00327 (0.00695)
District Enrollment (natural log)	-0.0545** (0.0195)	-0.0725*** (0.0204)	-0.0641*** (0.0147)	-0.385 (0.225)
Mills	-0.0275*** (0.00407)	-0.0211*** (0.00413)	-0.0244*** (0.00314)	-0.0201*** (0.00375)
Private School Enrollment (%)	-0.00797 (0.00682)	-0.00769 (0.00736)	-0.00814 (0.00516)	
Owner-Occupied Residence (%)	-0.00545** (0.00175)	-0.00238 (0.00191)	-0.00385** (0.00140)	
Constant	1.129*** (0.266)	0.522 (0.297)	0.846*** (0.209)	0.673 (0.503)
Observations	1960	1995	3955	3986
R ²	0.085	0.099	0.082	0.066
Year Fixed Effects	Yes	Yes	Yes	Yes
District Fixed Effects	No	No	No	Yes
Academic Designations	Yes	Yes	Yes	Yes

Heteroskedasticity robust standard errors clustered at the district level in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: First AYP Failure

	(1) Pass	(2) Pass	(3) Pass	(4) Pass
First AYP Failure	-0.0740* (0.0341)	-0.0701 (0.0361)	-0.0950* (0.0442)	-0.0823 (0.0480)
Subsequent AYP Failure			-0.0487* (0.0207)	-0.0507* (0.0235)
Current Millage	0.0172*** (0.00180)	0.0360*** (0.00430)	0.0179*** (0.00199)	0.0448*** (0.00454)
Property Values (per pupil)	0.00479 (0.00245)	0.0246* (0.0113)	0.00432 (0.00257)	0.0438*** (0.0119)
Tot. Expenditures (per pupil)	-0.0588*** (0.0165)	-0.0697* (0.0306)	-0.0533** (0.0178)	-0.0506 (0.0348)
Instructional Expenditures (per pupil)	0.0634* (0.0291)	-0.0142 (0.0520)	0.0599 (0.0313)	-0.0577 (0.0621)
Black (%)	0.000384 (0.00110)	-0.00489 (0.00510)	-0.000187 (0.00117)	-0.0114 (0.00671)
Hispanic(%)	0.00671** (0.00256)	0.00942 (0.0154)	0.00343 (0.00294)	0.0151 (0.0183)
English-Learners (%)	0.000542 (0.00236)	-0.0114 (0.0197)	0.0000751 (0.00268)	-0.0152 (0.0219)
Poor (%)	-0.000347 (0.000917)	0.00190 (0.00142)	0.000175 (0.000979)	0.00133 (0.00160)
Disabled (%)	0.0000858 (0.00344)	-0.00280 (0.00696)	-0.00148 (0.00374)	-0.00696 (0.00844)
District Enrollment (natural log)	-0.0725*** (0.0142)	-0.369 (0.219)	-0.0752*** (0.0164)	-0.242 (0.206)
Mills	-0.0247*** (0.00319)	-0.0203*** (0.00379)	-0.0231*** (0.00353)	-0.0179*** (0.00435)
Private School Enrollment (%)	-0.00769 (0.00514)		-0.0136* (0.00585)	
Owner-Occupied Residence (%)	-0.00378** (0.00140)		-0.00382* (0.00156)	
Constant	0.816*** (0.202)	0.620 (0.491)	0.836** (0.303)	0.0903 (0.509)
Observations	3955	3986	3252	3279
R^2	0.081	0.066	0.081	0.072
Year Fixed Effects	Yes	Yes	Yes	Yes
District Fixed Effects	No	Yes	No	Yes
Academic Designations	Yes	Yes	Yes	Yes

Heteroskedasticity robust standard errors clustered at the district level in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Placebo Tests for Levy Passage

	(1) Pass	(2) Pass
Future Failure	-0.0198 (0.0227)	-0.0171 (0.0226)
AYP Failure		-0.0395 (0.0202)
Current Millage	0.0379*** (0.00413)	0.0378*** (0.00413)
Property Values (per pupil)	0.0285* (0.0111)	0.0288* (0.0111)
Tot. Expenditures (per pupil)	-0.0671* (0.0314)	-0.0669* (0.0313)
Instructional Expenditures (per pupil)	-0.0201 (0.0538)	-0.0222 (0.0538)
Black (%)	-0.00717 (0.00575)	-0.00678 (0.00589)
Hispanic(%)	0.00845 (0.0186)	0.00953 (0.0185)
English-Learners (%)	-0.0115 (0.0206)	-0.0121 (0.0211)
Poor (%)	0.00304* (0.00141)	0.00311* (0.00142)
Disabled (%)	-0.00662 (0.00703)	-0.00627 (0.00700)
District Enrollment (natural log)	-0.350 (0.222)	-0.354 (0.221)
Mills	-0.0219*** (0.00402)	-0.0219*** (0.00402)
Constant	0.344 (0.474)	0.384 (0.473)
Observations	3806	3806
R^2	0.072	0.074
Year Fixed Effects	Yes	Yes
District Fixed Effects	Yes	Yes
Academic Designations	Yes	Yes

Heteroskedasticity robust standard errors clustered at the district level in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$