JULIA A. PAYSON
Stanford University

When Are Local Incumbents Held Accountable for Government Performance? Evidence from US School Districts

Do voters hold local officials accountable for government performance? Using over a decade of panel data on school district elections and academic achievement in California, I causally identify the effect of test score changes on school board incumbent re-election rates and show that incumbents are more likely to win re-election when test scores improve in their districts—but only in presidential election years. This effect disappears in midterm and off-years, indicating that election timing might facilitate local government accountability.

Determining the conditions under which voters hold elected officials accountable for government performance is central to the study of politics. We know that at the national level voters evaluate political leaders on the basis of economic trends (Fiorina 1981; Key 1966; Lewis-Beck and Stegmaier 2000). However, the vast majority of elected officials in the United States serve at the local level and have little control over the national economy. Instead, over 90,000 local governments in the United States employ 11 million workers and deliver public services that directly impact the day-to-day life of citizens, including education, police protection, and water and sanitation services (US Census Bureau 2012). An important open question is the extent to which voters hold local politicians responsible for delivering these services. Only a handful of studies have examined whether voters punish and reward local government officials on the basis of job performance (Arnold and Carnes 2012; Berry and Howell 2007; Hopkins and Pettingill 2015), and as a result, we are only beginning to learn about how accountability works in local government.

Do voters hold local officials accountable for the quality of the public services they provide? If so, under what conditions? The majority of local governments in the United States are single-issue districts that

oversee specific policy areas. Within this context, voters should have a relatively easy time attributing responsibility to the incumbents who set district policies. If residents are satisfied with a service, we would expect to see incumbents re-elected at high rates. If quality deteriorates, or if voters are unhappy for any reason, one of the easiest ways to express this dissatisfaction is at the ballot box. School board elections present a unique opportunity to test this theory. Not only are school districts the most common type of local government in the United States, but performance data are readily available in the form of standardized test scores.

I use panel data on school board elections and school district performance in California to show that district achievement influences election results, but the strength and direction of this relationship varies based on when an election is held. In presidential election years, when turnout is at its highest, there is a strong, positive correlation between incumbent performance and district achievement on standardized tests in the year leading up to the election. However, this relationship is less pronounced in midterm years and completely flat in off-years. I argue that a likely explanation for this pattern is the fact that the people who vote in low-turnout elections are often a unique subset of voters with vested interests in the election results (Anzia 2011; Berry and Gersen 2011). While ordinary district residents may rely on district performance to guide their vote choice, the voters who are active in nonpresidential school board elections—like teachers and their unions, district staff, and other education-savvy voters—likely evaluate incumbents on a much broader range of criteria.

I also investigate whether changes in test scores lead incumbents to strategically exit school board races. I find no evidence that incumbents are less likely to run or that challengers are more likely to enter school board races when test scores fall, which suggests that my estimates of incumbent performance are not biased as a result of strategic candidate behavior. Finally, I assess whether low turnout in nonpresidential elections is a plausible mechanism driving the observed findings. I show that turnout in school board elections drops from roughly 32% in presidential years to 25% in midterm years and to under 15% in off-year elections. This finding holds across districts but is also true for districts that switch from off-year to even-year elections. When districts hold their elections in nonpresidential years, turnout drops, and academic performance matters less for incumbent re-election.

Voting and Accountability in the Local Context

Literature on government accountability typically claims that elected officials can be considered accountable if voters are able to

determine how well incumbents are performing their jobs and then periodically re-elect or vote them out of office on the basis of that assessment (Manin, Przeworski, and Stokes 1999). The standard view is that voters are primarily retrospective in their evaluations and can induce accountability by selecting candidates and then either sanctioning or rewarding them based on job performance (Ferejohn 1986; Fiorina 1981). A well-developed literature documents retrospective voting in national elections (Lewis-Beck and Stegmaier 2000; Tufte 1978). While much of this work has focused on voter reactions to the economy, newer research also shows that voters are willing to punish and reward national candidates for a wide variety of policy outcomes (Healy and Malhotra 2013).[1]

What is not yet known is how widespread retrospective voting is in local contexts. Compared to the complex task of evaluating the effectiveness of national policies, single-purpose governments at the local level present an "easy" opportunity for citizens to hold officials accountable for providing government services. Water boards, mosquito abatement districts, fire protection agencies, transportation districts, and other single-purpose governments typically perform jobs within a well-defined policy area—especially in comparison to the state and national government. Surveys show that citizens care deeply about the issues that are the domain of local government, particularly the areas of crime and education (Trounstine 2009). If voters are willing to attribute responsibility for policy outcomes to candidates at the national level, it seems reasonable that they should do the same for local incumbents who oversee the provision of concrete public services. After all, monthly water bills, public transportation options, and the quality of local public schools are often more tangible to the average person than the effects of national economic policies.[2]

School districts are the most common type of local government in the United States and present a particularly interesting case of single-purpose, democratically elected governments for which voters have unique access to performance data. When No Child Left Behind (NCLB) became a law in 2001, states were required to adopt new academic performance standards and to make this information publicly available. Given that most school board elections in the United States are nonpartisan, most voters have little to inform their vote choice other than how well the district is doing academically. Retrospective voting models developed at the national level provide a clear prediction for how voters should behave in this context: If a district is performing adequately, re-elect the school board incumbent. If test scores fall, or if poorly performing schools fail to improve, vote the incumbents out in the hope that new leadership will turn things around. However, several recent

articles have found only mixed evidence that voters punish and reward incumbent school board members on the basis of district test scores (Barrows 2015; Berry and Howell 2007; Kogan, Lavertu, and Peskowitz 2014). Given the theoretical prediction offered by the literature at the national level, why have scholars found so little evidence of retrospective voting in school board elections?

In this article, I offer new evidence that local political accountability varies based on the timing of elections. A growing body of literature shows that election timing has a major impact on a variety of outcomes, including voter turnout, bond passage rates, and public policy (Berry and Gersen 2010; Hajnal, Lewis, and Louch 2002; Meredith 2009). The predominant theoretical framework argues that voters turn out selectively depending on how much they have at stake in a given election and how costly it is to participate (Berry and Gersen 2011). Local elections held concurrently with major national or state races encourage broad participation because voters are already at the polls and attuned to political issues. On the other hand, off-year contests are less visible and attract a different pool of voters. In particular, well-organized groups whose members will be affected by the result of an election tend to vote regardless of when that election is held. Anzia (2014) demonstrates that off-cycle, low-turnout elections often enable these organized groups to achieve favorable policy outcomes. For example, teachers and their unions dominate school board elections because board members make decisions that directly affect their day-to-day jobs (Moe 2006, 2011). When school board contests are held off-cycle, unions are able to elect representatives who then negotiate contracts that increase the salaries of local teachers (Anzia 2012).

In terms of political accountability, election timing wouldn't matter if the general public evaluated local candidates using the same criteria as special interest voters. But if the preferences of these two groups diverged, incumbents would fare quite differently depending on who turned out at the polls. In the next section, I draw from the literature on local election dynamics and interest group politics to develop a theory about when we should expect to see public accountability in school board elections. I argue that ordinary voters likely view test scores as a valuable source of information about school board effectiveness. But teachers and the other voters who dominate off-year elections almost certainly evaluate board members on a different set of criteria that are more relevant to their personal interests. As a result, we would expect to see a positive relationship between academic achievement and school board incumbent performance in high-turnout election years, but not necessarily in other years when organized issue voters are able to exert greater electoral influence.

How Homevoters and Organized Interests
Shape Local Elections

As voters decide how to select their school board candidates, school district quality is a key consideration. In general, people care a lot about local education policies (Trounstine 2009). This is particularly true among the voters most likely to turn out in local elections: long-term residents who own homes and have school-aged children (Fischel 2001; Oliver 2012). In a study on voting in suburban elections, Oliver and Ha (2007) found that homeowners were the group that demonstrated the greatest interest in local politics, the highest voting rates, and the most knowledge about the local candidates. In addition, homeowners are particularly concerned about the quality of schools in their neighborhood, given that this is a primary determinant of housing values (Black 1999). As a result, we would expect these local homevoters to pay attention to trends in school district performance. Recent work by Holbein (2016) demonstrates just this. He shows that being labeled a failing district under NCLB increases turnout in school board elections and that residents vote with their feet by moving or transferring their children out of the district. Holbein (2016) highlights the propensity of voters to exit from a neighborhood when public goods deteriorate (Hirschman 1970). My approach is complementary and addresses how voters assert their voice at the ballot box, which is oftentimes a more viable response for homeowners who are well-established in a neighborhood.

In general, even ordinary citizens are aware of how the public schools in their area are doing academically. A survey by Peterson, Henderson, and West (2014) showed that respondents' predictions of their district's performance relative to national averages correlated strongly with actual performance. People know, at least relatively, the level of achievement in their schools. Moreover, people are receptive to receiving information about their school district's performance and are willing to update their priors when presented with actual data about test scores (Clinton and Grissom 2015). The bottom line is that people know and care about public education. Voters in local elections are even more well-informed and politically knowledgeable than average citizens, and they are invested in the quality of their local schools as a result of owning homes in the area.

At the same time, it is an open empirical question whether voters translate their concerns about local education into the ballots they cast for school board incumbents. Research on the characteristics of effective school boards indicates that board members may have some level of control over academic performance in their district (Dervarics and O'Brien

2011). However, a large body of work on the correlates of educational achievement finds that student-level characteristics and teacher quality are by far the most important predictors of achievement (Chetty et al. 2011; Chingos, Whitehurst, and Gallaher 2015). Voters may well be taking district performance into account when selecting their school board members, but it is not clear that this will actually lead to better outcomes for school districts. I return to this point in the conclusion.

If voters do hold school board members accountable for test scores, we would expect to see a positive correlation between student achievement and school board election results. However, existing studies on this topic have produced mixed results. Using a cross-sectional approach in South Carolina, Berry and Howell (2007) discover evidence of retrospective voting in some years but not others. Using panel data with fixed effects, Kogan, Lavertu, and Peskowitz (2014) find no evidence that district performance leads to school board turnover in Ohio. Finally, Barrows (2015) uses a regression discontinuity design to show that being narrowly labeled a top-performing district by the State Department of Education leads to an increase in incumbent support in Florida, although he finds that this effect does not occur at other performance thresholds.

Why is there such variation in these findings? One likely reason is that elections are held exclusively in odd years in Ohio and exclusively in even years in South Carolina and Florida. My study uses data from California, which has the largest and most diverse public school system in the country. Moreover, local districts have discretion over when to hold their elections, and as a result districts hold contests not only during presidential years but also during midterm and off-year elections.[3] This variation allows me to test a previously unexplored hypothesis about school board accountability: namely, that the degree to which the public punishes and rewards incumbents on the basis of standardized test performance depends on how much influence ordinary voters have at the ballot box. As established by Anzia (2014) and others (Berry and Gersen 2011), we can expect this influence to be significantly diminished in off-years, when turnout drops by up to half and organized groups like teachers unions are at their most active. Crucially, California allows candidates to designate their profession on the ballot. Over 95% of school board incumbents choose to indicate their incumbent status, which makes it easy for voters to identify district leadership when they cast their ballot.

There are good reasons to believe that teachers and other high-demanding education voters evaluate school board incumbents differently than the public at large. It is no secret that teachers are largely opposed to the standardized testing movement. Educators state that they

are concerned about having to "teach to the test" and claim that overtesting is harmful for student learning (Long and Robertson 2014). However, a greater concern is likely that their job security might be undermined by pay-for-performance schemes that would link their salaries to student achievement. Teachers often view the proliferation of standardized testing and the collection of achievement data as precursors for this type of evaluation. In fact, the National Education Association asserts that nearly half of teachers have considered leaving the profession due to standardized testing requirements (Walker 2014). It is therefore not surprising that members of the California Teachers Association were strong supporters of AB 484, a 2013 bill that replaced the existing academic performance measure with the new Smarter Balanced Assessment System, which evaluates district performance using a variety of other information in addition to standardized test scores (Fensterwald 2015).

At the same time, district leaders often see standardized tests as a critical component of measuring and improving student achievement. The California School Boards Association platform states that district leaders "should focus on standards, frameworks, accountability and data . . . . Schools must use accurate data and reliable assessment results to regularly measure student progress. *Assessment results effectively improve student performance*" [emphasis added].[4] In fact, school boards are under pressure to increase test scores in their district or risk facing a variety of negative consequences that decrease district revenue and autonomy, including the implementation of state mandates and restrictions on funds. If districts fail to make sufficient academic progress, they are in danger of losing both students and funding to higher-performing districts (Larsen 2009). This means that school board members and teachers are often in conflict when it comes to their views on the importance of standardized tests. As a result, teachers probably don't use test scores as the primary metric by which they evaluate board members. But there are many other issues that teachers and their unions care deeply about—and over which school boards have direct control. The California Teachers Association advocacy agenda is transparent about these priorities, which include less testing, smaller class sizes, safe school environments, higher salaries, and more time for professional development.[5]

In summary, we should expect ordinary home voters to reward (or punish) incumbents at the ballot box largely based on district performance. But if special interest voters with different priorities are able to exert enough electoral influence, this relationship should be flat. The data that I present are consistent with this story: In presidential-year elections, when turnout peaks, incumbents are estimated to receive 2% more of the vote share than they would have and are re-elected at a rate that is

8% higher if their district oversaw test score improvements that were one standard deviation above the mean rather than one standard deviation below the mean for that year. In midterm and off-years, however, there is no relationship between district achievement and incumbent electoral performance.

School District Achievement and Election Data in California

School Districts and Academic Performance

California's public K–12 school system comprises over 1,000 school districts and serves over 6.2 million students. These local education agencies vary widely, from isolated rural districts with fewer than 20 students to Los Angeles Unified, which serves over 600,000 students and is the second largest urban district in the country. When NCLB passed in 2001, states were required to implement data collection and reporting systems to track academic achievement. In California, the Academic Performance Index (API) was the existing state-level metric that was adapted to meet NCLB regulations.[6] District API scores ranged from 200 to 1,000, and students were tested both in the fall and in the spring to chart their growth over each academic year. Data on API scores are available at the California Department of Education (CDE) website. Figure 1 shows how API scores changed from 2003 to 2012. The data indicate that statewide test scores improved by just under 100 points over this period, although my research design exploits the fact that districts change at different rates over time.

School Board Elections

School boards in California are responsible for hiring the district superintendent, setting policies, adopting curricula, establishing the budget, and negotiating collective bargaining agreements with school employees. As mandated by state law, boards consist of either three, five, or seven members. Board members usually serve four-year terms, and there are no term limits. The majority of these school board elections take place in November of even-numbered years, with roughly 40% falling in presidential years and 40% taking place in midterm years. The remaining 20% of districts schedule their elections in odd-numbered years. An even smaller minority of districts (less than 4%) hold their elections in the spring or during other months. Because the treatment of interest is the release of district test scores in the late summer, I exclude these races from the analysis.

FIGURE 1

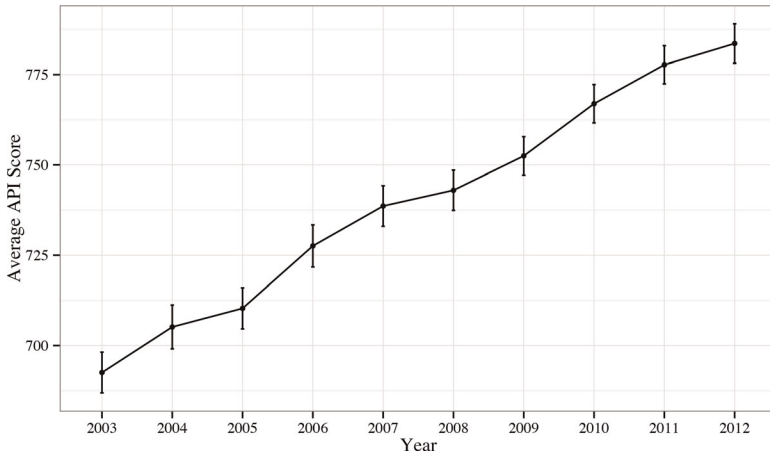Average District API Scores in California by Year

Table 1 presents descriptive statistics on November school board races and district characteristics in California from 2003 to 2012 broken down by election-year type. On average, contests consist of one or two incumbents and two or three challengers vying for two available seats on each school board. The incumbents who run perform well, winning re-election over 70% of the time. In these elections, voters may vote for as many candidates as there are seats available, and vote share is calculated by dividing the total number of votes for a particular candidate by the sum total of all votes. The average incumbent earned around 30% of the vote share, which is quite high given that most races had four candidates running.[7] On average, off-year and midterm elections look quite similar in terms of their electoral characteristics. There are slightly more seats at play in these races than in presidential contests, and a marginally higher percentage of incumbents are re-elected. However, these races look very different in terms of turnout. Districts with even-year elections are larger than districts with off-year elections, and turnout in presidential years is twice as high as off-years and roughly 25% higher than midterm years.

Methods and Empirical Results

How Do Voters React to Changes in School District Test Scores?

I created an original longitudinal data set by merging district-level election-return data with information about district characteristics and

TABLE 1
California School Board Election Descriptive Statistics, 2003–12

	Election-Year Type[a]		
	Presidential	Midterm	Off-Year
Election Variables			
Number of Elections per Year			
Mean	332	344	176
SD	4.6	33.2	33.0
Number of Seats per Race			
Mean	2.0	2.5	2.4
SD	0.7	0.8	0.7
Number of Incumbents per Race			
Mean	1.4	1.9	1.6
SD	0.9	0.9	1.0
Number of Challengers per Race			
Mean	2.6	2.7	2.8
SD	1.7	1.8	1.6
% Incumbents Running			
Mean	70.1	76.0	68.0
SD	37.7	31.0	34.9
Incumbent Vote Share			
Mean	31.7	27.9	28.2
SD	12.6	12.3	11.1
Challenger Vote Share			
Mean	26.7	22.7	23.4
SD	12.1	11.0	10.0
% of Incumbents Re-Elected			
Mean	73.5	78.4	75.9
SD	37.0	33.8	34.8
District Variables			
District Voting Age Population[b]			
Mean	50,405	49,125	36,263
SD	77,192	78,898	43,647
Votes Cast			
Mean	16,780	11,938	4,091
SD	31,173	20,691	5,504
% Turnout			
Mean	33.3	26.0	16.0
SD	23.2	17.4	27.7
District API			
Mean	748	758	738
SD	88	81	92
Made AYP[c]			
Mean	0.30	0.33	0.38
SD	0.46	0.47	0.49

[a]Presidential years are 2004, 2008, and 2012; midterm years are 2006 and 2010; off-years are 2003, 2005, 2007, 2009, and 2011.
[b]Voting age population at the district level is from the American Community Survey and available at: http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
[c]Variable equals 1 if a district made Adequate Yearly Progress and 0 otherwise.

student academic performance. The panel runs from 2003 to 2012. While API scores are available beginning in 1999, the new NCLB reporting standards were implemented in 2002 and provide additional district-level information, including whether a district made Annual Yearly Progress.[8] The panel ends in 2012 because the following year saw changes in statewide testing protocol as California shifted to the new set of Common Core academic standards, culminating in an over-haul of the existing API framework with the passage of AB 484 in 2013. Candidate-level data are aggregated into district averages, and the result-ing data set is an unbalanced panel that includes 2,562 district-year elections. On average, districts held three elections during this time peri-od. I use the district as the unit of analysis rather than the individual incumbent because this allows me to compare the same units over time and employ fixed effects to control for time-invariant district characteris-tics. Such contextual effects would be unidentified in models featuring individual incumbent observations as a result of multicollinearity.

Following Berry and Howell (2007), the treatment that I consider is the change in a district's test scores in the year leading up to the elec-tion. Each year, the California Department of Education releases two API scores for each school district: a base score and a growth score. The growth score is specifically designed to be comparable to the previous year's base score while accounting for differences in test questions, and scores are weighted to ensure consistency.[9] The one-year change in API base and growth scores is one of the metrics by which the CDE assesses whether districts are making Adequate Yearly Progress, and it is also a good proxy for whether districts are generally improving or not.[10] Litera-ture on retrospective voting has established that voters are consistently myopic in their assessments of government performance, overweighting recent economic and policy developments (Achen and Bartels 2004; Healy and Malhotra 2009). Updated API growth scores are released just a few months prior to the November school board elections in California and provide voters with a salient reminder of how their local schools are performing academically. Moreover, local news sources report widely on trends in student achievement. Typical headlines read: "Local test scores near best in state; Adjusted API results show Walnut Creek ele-mentaries far exceed California average" (*Contra Costa Times*), "Schools show API gains" (*San Bernardino Sun*), and "Test scores drop in La Habra, Brea schools" (*Orange County Register*).

The intuition for why changes in test scores should matter more than API levels in school board elections can be demonstrated in Figure 2. On average, district test scores increased in California in every year from 2003 to 2012. However, some districts changed more rapidly

FIGURE 2
California District API Scores by Year: Motivating Cases and
Demonstration of Treatment
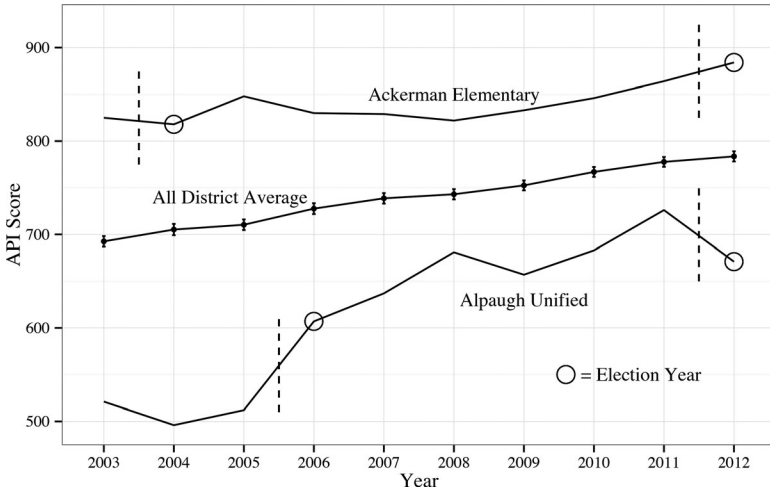
than others, and other districts saw declining scores in certain years. For example, Alpaugh Unified saw a dramatic improvement in scores right before the 2006 election, while it experienced a drop before the election in 2012. Despite the fact that the district's 2012 scores were higher than those in 2006, we know from the literature on economic voting that voters respond primarily to changes in economic conditions in the year and even quarter leading up to the election. As a result, we would expect the incumbents in Alpaugh Unified to fare worse in 2012 than they would have had scores risen—even though the overall API levels were higher than in the previous election.

The following models use both raw data on one-year test score changes as well as standardized scores. The benefit of the standard scores is that they are more easily interpretable and show how a district performed relative to the mean for that year; however, both the raw data and the standardized measures produce an identical pattern of results. Because the fixed effects approach compares a district's performance to its own time-averaged value, each district serves as its own control. Voters respond to overall changes in raw API scores, but they also appear to respond to relative changes, which are captured by the scores. Determining whether voters place greater value on absolute or relative academic performance is beyond the scope of this article: What is clear is that

incumbent performance in presidential years is linked to at least two measures of district achievement.

The setup for most of the following analyses is a generalized difference-in-differences approach for individual-level panel data with district and year fixed effects. This allows me to estimate the *within-district* effect of a change in test scores on a variety of district-level outcomes. The benefit of this approach is that it controls for time-invariant district characteristics that may be correlated with the observed regressors. I then allow the individual district intercepts to be estimated with random effects and employ a mixed model to more precisely identify the effect of changes in API; the results are robust to both approaches.

Academic Performance Does Not Lead to Strategic Candidate Exit

I begin by addressing the first-order question of whether incumbents are strategically deciding to rerun for office depending on how well their district performed over the previous year. If high-quality candidates only run for re-election when test scores in their district improve, then this would bias the observed relationship between test scores and incumbent performance. The raw data show a completely flat relationship between standardized one-year changes in API scores and the number of incumbents who choose to seek re-election in each district (Figure 1A in the online supporting information).

The fixed effects model confirms this visual trend. I estimate the following model via OLS:

$$V_{it} = \Delta A_{it}\beta_1 + X'_{it} + \beta_2 + \eta_t + \omega_i + \epsilon_{it},$$

where V_{it} is the percent of eligible incumbents in district i who sought re-election in year t. ΔA_{it} is either the raw or standardized change in API score in the year leading up to the election, X_{it} are time-varying district-level covariates, η are the year-specific fixed effects, and ω are the district-specific fixed effects. In this case, the X variables include the number of seats in each election, a control variable for the API level of each district, and an indicator variable that equals 1 if a district met its Annual Yearly Progress goal for that year. The standard errors presented in the following models are always clustered at the district level to allow for correlation across within-district error terms over time.[11] I then reestimate the model above and include an interaction between election-year type and change in API to check whether the results vary depending on when an election is held. Table 2 shows the lack of effect of either change in API scores or API levels on the percent of eligible incumbents who seek re-election. This is also true when the effect is allowed to vary

TABLE 2
API Scores and Percent of Incumbents Seeking Re-Election

	Dependent Variable			
	% Eligible Incumbents Seeking Re-Election (0–100)			
	(1)	(2)	(3)	(4)
One-Year API Change (Raw)	0.059 (0.065)	−0.014 (0.108)		
One-Year API Change (Standardized)			0.878 (1.136)	−0.348 (1.864)
Midterm Election	−2.081 (2.962)	−3.649 (3.253)	−2.494 (2.736)	−2.487 (2.737)
Off-Year Election	−0.760 (6.152)	−1.611 (6.469)	0.684 (5.934)	0.703 (5.952)
API Change × Midterm		0.157 (0.130)		2.645 (2.338)
API Change × Off-Year		0.083 (0.126)		1.424 (2.264)
API Level (Raw)	−0.024 (0.043)	−0.026 (0.043)		
API Level (Standardized)			0.716 (3.537)	0.611 (3.551)
Made AYP	−1.323 (2.319)	−1.175 (2.315)	−1.581 (2.309)	−1.447 (2.301)
Number of Open Seats	1.465 (1.576)	1.485 (1.580)	1.496 (1.577)	1.510 (1.581)
Year FE	Y	Y	Y	Y
District FE	Y	Y	Y	Y
	Districts = 836	Districts = 836	Districts = 836	Districts = 836
	$N = 2{,}544$	$N = 2{,}544$	$N = 2{,}544$	$N = 2{,}544$

Note: OLS estimates. MacKinnon and White (1985) robust standard errors clustered by district. Coefficients on both API changes and levels are small and imprecise for both the raw and standardized measures, showing that district performance does not cause incumbents to strategically exit from school board races. The result holds if election types are considered separately. $*p<0.1$; $**p<0.05$; $***p<0.01$.

across presidential, midterm, and off-year elections. The findings are robust to using either raw API data or standardized scores.

Across all models, the coefficients on change in API and API levels are small and imprecisely estimated, demonstrating that incumbents are no more likely to seek re-election when scores increase in their district. As a robustness check, I also run a Poisson model to estimate the raw number of incumbents who seek re-election, using the same set of controls. The results again indicate that changes in API scores have no effect on how many eligible incumbents run for office (see Table A1 in the online supporting information). The trend is similar for the number of challengers

who choose to enter school board races: After including year and district fixed effects, the same number of challengers are likely to run whether test scores drop or improve. Overall, the data do not indicate that incumbents and challengers are strategically entering the race on the basis of one-year changes in test scores. This might be because the year-end test scores are often released late in the summer, just after the filing period for declaring candidacy, or it may be that other factors are influencing the candidates' decisions to run. Regardless of the reason, if incumbents are not strategically exiting their races, then my estimates of how they fare electorally conditional on running are a better approximation of the overall effect of test score changes on incumbent success than they would be if there were widespread incumbent dropout in poorly performing districts.

*How Election-Year Timing Shapes School Board Incumbent Performance*

To test how changes in API scores affect incumbents at the ballot box, I operationalize incumbent performance in two different ways: the average incumbent vote share in a district and the percentage of incumbents who are re-elected. The data do not show an overall effect of one-year changes in API scores on either of the variables above when all years are considered together (see Table A2 in the online supporting information). This is not surprising, given my theoretical prediction that election timing leads to heterogeneous treatment effects. When I introduce an interaction effect between the type of election (presidential, midterm, or off-year) and one-year changes in API, I find strong evidence that API changes have a positive impact on both measures of incumbent performance in presidential years. In midterm years, this relationship is almost completely flat, and in off-years this relationship is just barely negative. The raw data provide crude intuition for the findings, showing the trends for average incumbent vote share when election types are either pooled or considered individually (Figure 3).

To estimate the effect of changes in test scores on incumbent performance in different types of elections, I employ a generalized difference-in-differences setup similar to the model presented in the previous section:

$$V_{it} = \Delta A_{it}\beta_1 + E_{it}\beta_2 + (\Delta A_{it} \times E'_{it})\beta_3 + X'_{it}\beta_4 + \eta_t + \omega_i + \epsilon_{it}.$$

V_{it} is either average vote share or the percentage of incumbents who won re-election in district i in year t. Both of these dependent variables range from 0 to 100. ΔA_{it} is either the raw or standardized one-year change in

FIGURE 3
Incumbent Vote Share Across Changes in API Scores
by Election Type
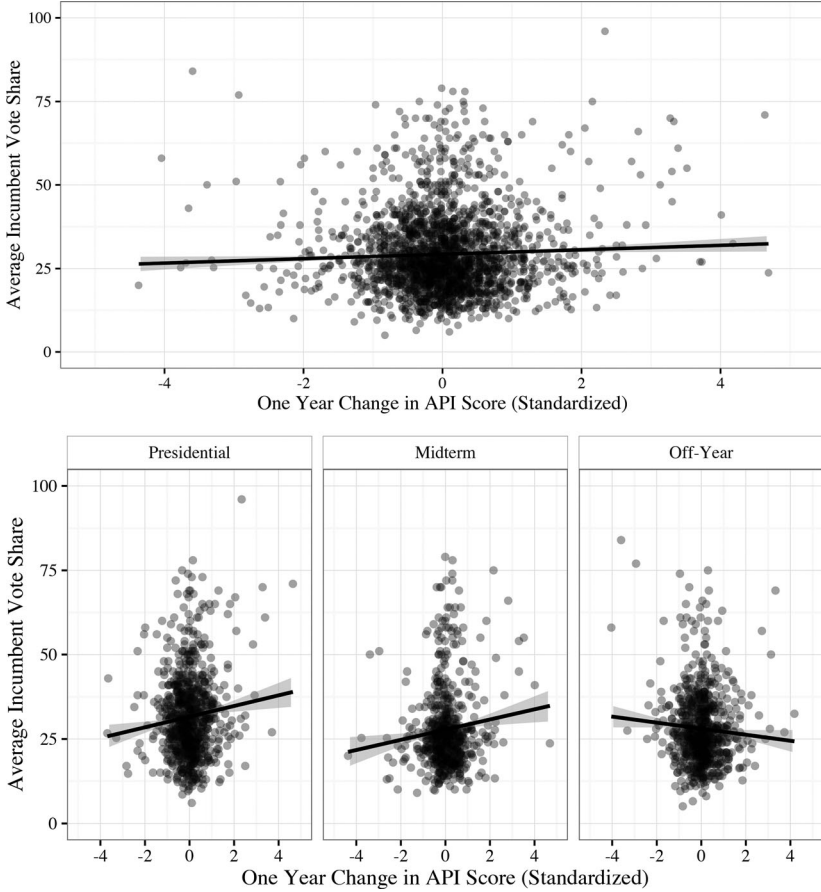
Note: Raw data demonstrate that when all elections are pooled, there is no relationship between change in API scores and average incumbent vote share. However, when elections are broken down by type, there is a modest positive correlation between API changes and incumbent performance in presidential years and a negative correlation in off-years.

API scores, and E_{it} is a categorical variable for the election type. X represents district-level covariates that include the API level, an indicator for whether the district made Annual Yearly Progress, the number of candidates running, and the number of seats. The coefficients of interest are β_1 and β_3, which tell us how changes in API affect the dependent variables

TABLE 3
One-Year Change in Test Scores and Incumbent Performance

| | Dependent Variable | | | |
| | Avg. Incumbent Vote Share (0–100) | | % Incumbents Re-Elected (0–100) | |
	(1)	(2)	(3)	(4)
One-Year API Change (Raw)	0.050**		0.218**	
	(0.022)		(0.096)	
One-Year API Change (Standardized)		0.912**		3.814**
		(0.383)		(1.656)
Midterm Election	0.826	0.308	2.531	0.117
	(0.630)	(0.536)	(3.588)	(3.058)
Off-Year Election	−0.095	−0.545	3.698	1.095
	(1.111)	(1.032)	(6.536)	(6.045)
API Change × Midterm	−0.047*	−0.923*	−0.214	−3.889
	(0.027)	(0.484)	(0.152)	(2.707)
API Change × Off-Year	−0.053**	−1.031**	−0.253**	−4.451**
	(0.026)	(0.481)	(0.120)	(2.149)
API Level (Raw)	−0.005		−0.026	
	(0.011)		(0.048)	
API Level (Standardized)		−0.057		−0.916
		(0.915)		(4.038)
Made AYP	0.296	0.252	−0.076	−0.257
	(0.455)	(0.447)	(2.639)	(2.631)
Number of Open Seats	−5.050***	−5.037***	9.800***	9.846***
	(0.390)	(0.390)	(2.079)	(2.083)
Year FE	Y	Y	Y	Y
District FE	Y	Y	Y	Y
	Districts = 800	Districts = 800	Districts = 800	Districts = 800
	N = 2,229	N = 2,229	N = 2,229	N = 2,229

Note: OLS estimates. MacKinnon and White (1985) robust standard errors clustered by district. In presidential election years, both raw and standardized one-year changes in API scores have strong, positive effects on incumbent vote share and the percentage of incumbents who are re-elected. This relationship is almost completely flat in midterm years and even slightly negative in off-years. *$p<0.1$; **$p<0.05$; ***$p<0.01$.

in presidential, midterm, and off-year elections. Because standardized change in API is a continuous variable, this strategy allows us to observe how the intensity of the treatment affects within-district incumbent outcomes. The results are presented in Table 3.

The results show that API changes have a strong, positive effect on incumbent performance in presidential election years across specifications. The models predicting incumbent vote share produce slightly

smaller estimates in terms of magnitude because so much of the variation in vote share is mechanically determined by the number of seats and the number of candidates running. Nevertheless, a 10-point increase in a district's raw one-year change in API is expected to increase the vote share of the average incumbent by half a percentage point in presidential election years. The standardized API scores present even more striking evidence: Each standard deviation increase in change in API score is expected to increase the within-district incumbent vote share by around 1 point in presidential election years. These school board races are typically quite close, with the average incumbent earning only 4.5 points more than the average challenger, so this effect is substantively quite large. The average incumbent, however, experiences virtually no electoral gain or punishment in midterm years: Adding the coefficient on API Change × Midterm to the coefficient on One-Year API Change yields an estimate of zero. In off-year elections, the effect of test scores on vote share is again almost flat and even barely negative (in Model 2, for example, the effect is $0.9–1.0 = −0.1$). To reemphasize, a benefit of the fixed effects within estimator is that it compares district-specific deviations of the regressors and dependent variables to their time-averaged values. The coefficients above estimate how incumbent vote share would change *in the same district* if that district oversaw test score changes that were higher or lower relative to their own yearly average.

Perhaps an even more substantively important outcome is whether incumbents are any more likely to win or lose based on changes in their district's performance. The fixed effects models above indicate that if a district oversaw API changes that were one standard deviation above the average for that year rather than one standard deviation below the average, the percentage of incumbents expected to win increases by over 7% in presidential years.[12] In midterm elections, incumbents again are virtually immune to changes in test score, and in off-years each standard deviation increase in test scores leads to a minor drop of around half a point in the percentage of incumbents expected to win. Figure 4 shows the marginal effect of one-year change in standardized API scores in each election type, holding other variables at their mean. The effect plot provides further visual confirmation that changes in API score influence the electoral outcomes of incumbents in presidential years but not in other years.

Finally, I ran the models above with interactions between API levels (rather than changes) and election-year type to check whether voters assess the overall performance of districts differently across years. I find no significant interaction effects, which makes sense given that the district fixed effects should largely capture any systematic differences in a

FIGURE 4

Effect of Test Score Changes on Percent of Incumbents Re-Elected
Across Election Types
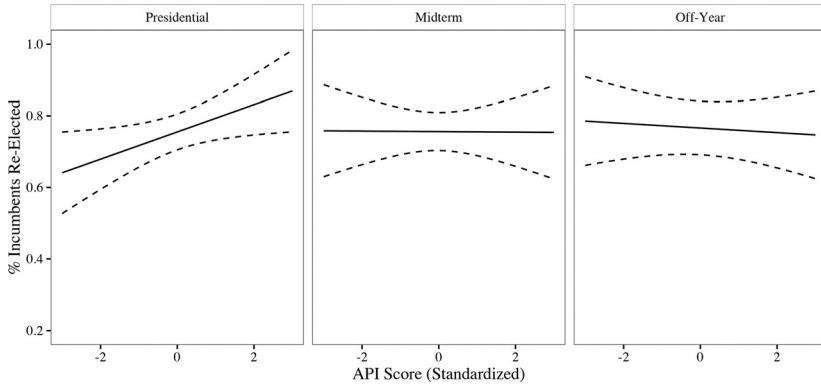
Note: This marginal effect plot shows that in presidential years, standardized one-year changes in API scores strongly predict the percentage of school board incumbents re-elected to office. This relationship is almost completely flat in midterm and off-year elections.

district's average incumbent performance that correlates with its underlying level of academic achievement. Rather, what seems to matter to voters is *changes* in a district's performance—at least in presidential years (see Table A5 in the online supporting information).[13]

Robustness

The within-district models specified in the previous section provide transparent and easy to interpret estimates of the effect of API changes on incumbent performance. However, there are drawbacks to using fixed effects. The fixed effects approach does not allow out-of-sample predictions because the unit effects of unobserved districts are not known. Furthermore, these estimates rely only on within-district variation and thus fail to account for potentially relevant information about variation across districts. Because there are only a small number of elections for each district over the course of the panel, the estimates are also sensitive to random error and prone to high variance, despite being unbiased. In fact, fixed effects models often display such high variance in small samples that a mixed or random effects approach is preferable (Clark and Linzer 2015; Greene 2012). This is because although a random effects specification will be biased if the individual unit effects are too highly

FIGURE 5

Effect of Change in API Scores by Election Type: Fixed vs.
Random District Effects
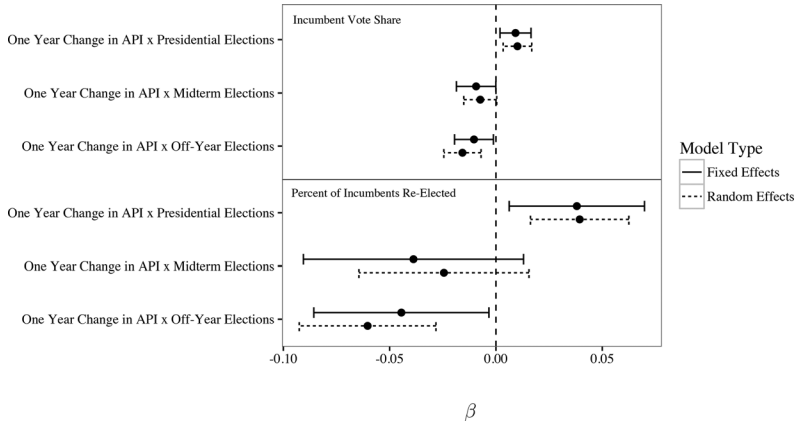
Note: Fixed and random effect models produce similar estimates of the effect of API changes on incumbent performance, with the random effects models yielding smaller standard errors.

correlated with the other explanatory variables, the associated reduction in variance can often lead to better estimates, on average.

I reestimate the models above allowing the unit effects for the districts to vary randomly. This approach assumes that the district effects, ω_i, are drawn from a normal probability distribution with an average unit effect μ_ω and variance σ^2. The benefit of this strategy is that it reduces the variability of the estimates by partially pooling information across districts (Gelman and Hill 2007). While the fixed effects models compare within-district incumbent performance to within-district changes in API scores, a model with random district effects also measures the difference *across* districts in a given election type. Although this approach will be biased if the changes in API scores are too closely correlated with the individual district effects, it also improves the precision of the estimates.

In fact, I find that the magnitude and direction of the coefficients on change in API remain remarkably similar across both sets of models. The standard errors are smaller in the mixed effects models (with random district effects and fixed election-type effects), and any bias that is introduced appears to be small. The coefficient plot in Figure 5 compares the coefficients from the fixed and random effects models. The pattern remains identical, with incumbents performing significantly better in presidential years when API scores increase; the negative coefficients on the interactions of midterm and off-year elections yield an estimate of

TABLE 4
Turnout by Election Type

	Dependent Variable	
	Voting Age Turnout	
	(Pooled)	(Fixed Effects)
Midterm Election	−7.039*** (0.534)	−7.567*** (0.382)
Off-Year Election	−17.501*** (0.829)	−11.195*** (1.268)
(Intercept)	32.296*** (0.676)	
District FE		Y
		Districts = 736
	N = 1,897	N = 1,897

Note: Voting Age Turnout is equal to the number of voters casting a ballot in a district election divided by the voting age population in that district (American Community Survey Five-Year Estimates). MacKinnon and White (1985) standard errors clustered by district. The pooled model shows clear differences in school board election turnout across election types. Turnout drops by 7 points in midterm years and by 17 points in off-years compared to presidential years. Fixed effects models that estimate within-district turnout similarly show that turnout drops when a district moves from presidential to midterm or off-year contests. $*p<0.1; **p<0.05; ***p<0.01$.

almost zero after being added to the baseline coefficient on presidential years. The similarity of the estimates provided by both the fixed and random effects models demonstrates that the findings are robust to various modeling specifications.

Election Timing and Turnout

A likely reason that voters appear to hold school board incumbents accountable for district test scores in presidential years is that these elections experience substantially higher turnout than contests held in midterm and off-years. Results from a simple pooled OLS model reveal that turnout drops by around 7 points in midterm years and 17 points in off-years (Table 4). However, districts that choose to hold their contests in odd years may differ in unobservable ways from districts that hold elections in even years. To address this concern, I also estimate a district fixed effects model to predict differences in turnout *within* districts that hold elections in different years. In fact, over 60 districts switched from odd-year to even-year contests over the course of the panel, and over 400 held elections in at least two different types of years.

As with the pooled model, the within-district fixed effects estimate reveals a dramatic drop in voter turnout when a district holds its school board race in an odd or midterm year compared to a presidential year. These results demonstrate that low turnout is a plausible mechanism linking incumbent performance to district achievement in some years but not in others. When turnout among ordinary voters is high—in presidential-election years— incumbents pay an electoral penalty when test scores in their district drop. When turnout is lower and favors well-organized groups with a vested interest in local education, voters appear to evaluate incumbents on criterion other than changes in test scores.

Discussion

The literature on government accountability and retrospective voting at the national level provides a clear theoretical prediction that voters should punish and reward local incumbents on the basis of public service provision outcomes. At the same time, local political dynamics are often quite different from national elections in terms of who votes and how organized groups are able to exert influence. Relatively few studies have tested how responsive voters are to government performance in the local context, and those that have have often produced mixed results. School districts are an ideal case to evaluate competing theories about the relationship between government performance and electoral accountability. These local education agencies are democratically elected governments that operate within a well-defined policy domain, and a clear metric of performance data is readily available in the form of standardized test scores. School board elections in California provide the added bonus of representing districts from the largest public school system in the nation and include contests that fall in presidential, midterm, and odd years.

This article provides new evidence that school board election outcomes are tied to district academic achievement, but this effect varies by the type of year in which elections are held. In presidential election years, there is a robust, positive correlation between district performance on standardized tests and incumbent performance at the polls. In midterm elections, this relationship is flat, and in off-years—when turnout is up to 50% lower—the effect is even slightly negative. There is no reason to believe that ordinary voters are evaluating incumbents differently across election types; rather, the most likely explanation for this observed finding is that special interest voters are able to exert substantially more influence at the ballot box in nonpresidential-year elections. Given that teachers care more about issues like classroom size, teacher salaries, and curricular decisions, it is not surprising that they would not

necessarily reward incumbents who oversee large gains in test scores. These findings build on work by Anzia (2011, 2012) and show that the timing of elections may have implications for political accountability in addition to influencing public policy outcomes.

Of course, there may be other reasons why presidential-year elections seem to facilitate accountability. Presidential elections receive higher levels of media coverage, so it is possible that voters are simply less attuned to issues of student achievement in other years. Berry and Howell (2007) propose this as an explanation for their findings that there was an association between test scores and incumbent electoral performance in 2000 but not in 2002 or 2004. However, I find that local sources are just as likely to report on student achievement in off-years as in even-numbered years in the three months leading up to the election.[14] The overall salience of elections held off-cycle may be lower, but the level of substantive information available about school board contests remains similar. Unless voters are using some different metric to evaluate their school boards in even and odd years, a more plausible explanation for the observed disparity is that the composition of the electorate is different in nonpresidential years—and favors well-organized groups that vote according to their personal interests.

This does not rule out the possibility that districts choosing to hold their elections in odd years may vary systematically from other districts in a way that also reduces political accountability. This article estimates the effect of test score performance on incumbent re-election chances for districts with elections at different times and does not causally identify the effect of timing itself. However, the finding that voter turnout increases dramatically in districts that switch from off-year to even-year elections over the course of the panel lends support to the idea that voters who participate in off-year elections are a unique subset of the voting population. My findings are consistent with a growing body of literature showing that election timing matters for a wide variety of political outcomes, and the next step for this research agenda will be to uncover the specific mechanisms that lead to greater public accountability in presidential years.

Overall, my results indicate that voters hold school board incumbents accountable for changes in district test scores in certain electoral contexts. However, while voting school board members out of office when test scores drop might be a rational response by voters seeking to induce local accountability, it is not clear that this is a normatively desirable result. For one thing, school boards likely have only a modest ability to shape academic performance in their districts. If this is true, then replacing school board incumbents will not necessarily increase student achievement. Moreover, increased turnover among district leaders may

actually harm districts. Cain and Kousser (2004) found that term limits in California led to a decrease in the policy expertise of state legislators who were constantly entering and exiting office. Similarly, school board members may gain valuable experience from their time serving in office, in which case voting incumbents out might not be the best way to improve failing schools. This article demonstrates that whether incumbents are electorally punished or rewarded for changes in district test scores varies by election type, but future research will be required to determine whether these voting patterns ultimately help or hurt school districts.

Julia A. Payson <jpayson@stanford.edu> *is a Ph.D. candidate in the Department of Political Science at Stanford University and an Affiliated Researcher at the Bill Lane Center for the Study of the American West. She studies representation, accountability, and public service provision in state and local governments in the United States. Her work has appeared in* Legislative Studies Quarterly.

Notes

1. Other scholars have rightly noted that voters also use prospective considerations in choosing candidates (Erikson, MacKuen, and Stimson 2000; Fearon 1999). Ashworth (2012) provides an excellent review of the modern perspective, which holds that voters use past incumbent actions as a way to form prospective assessments. Similarly, I use the language of retrospective voting in this article simply to indicate that people are reacting at least in part to observed government performance when they cast their votes and do not mean to imply that this decision is purely retrospective.

2. To be clear, voters may not be entirely accurate in their attributions. Healy, Malhotra, and Mo (2010), for example, find that irrelevant events can effect voter evaluations of government incumbents. Voters do not need to know everything about how their local school board members affect district achievement in order to engage in retrospective voting: They simply need to associate district leadership with academic performance. This task is at least as plausible as evaluating national candidates on economic policies.

3. See California School Boards Association (2007).

4. See https://www.csba.org/Advocacy/CSBAPositions/~/media/CSBA/Files/Advocacy/Positions/CSBAPolicyPlatform.ashx.

5. See http://www.cta.org/en/Issues-and-Action/Education-Improvement/Advocacy-Agenda.aspx.

6. The API was a measure of academic performance that had been in place since the Public Schools Accountability Act of 1999. In 2014, it was replaced by the California Assessment of Student Performance and Progress (CAASPP).

7. Candidate-level election results are available online at the California Election Data Archive. Additional details about the construction of the district-level data set used in this study are in the online supporting information, and replication data are available at the author's website.

8. These data are available at http://www.cde.ca.gov/ta/ac/ay/aypdatafiles.asp.

9. See http://www.cde.ca.gov/ta/ac/ap/documents/apiexecsummary.pdf.

10. Adequate Yearly Progress (AYP) is based on four sets of requirements: student participation in testing, percentage of students proficient in English-language arts and math, high school graduation rate, and the API score.

11. Specifically, I employ MacKinnon and White (1985) HC3 standard errors, which were found by Long and Ervin (2000) to perform best among small-cluster samples.

12. The results are substantively similar when I use different combinations of covariates. I present the full models here; additional results from other specifications are available in the online supporting information.

13. I also ran models with interactions between API changes and API levels to check whether voters in highly performing districts react differently to changes in test score than voters in low-performing districts. Again, I find no significant interaction effects.

14. I searched through over 300 California news publications in the NewsBank database using the keywords "school board election" and "test scores" or "achievement" in each year from 2003 to 2012. In presidential-election years, an average of 370 stories appeared between August 1 and Election Day, and in odd years, there were 206 stories on average. Given that there are usually 332 school board races in presidential-election years and only 176 contests in odd years, the story per race ratio is nearly identical (1.1 stories per contest in presidential years, and 1.2 stories per contest in off-years).

REFERENCES

Achen, Christopher H., and Larry M. Bartels. 2004. "Musical Chairs: Pocketbook Voting and the Limits of Democratic Accountability." Presented at the Annual Meeting of the American Political Science Association, Chicago.

Anzia, Sarah F. 2011. "Election Timing and the Electoral Influence of Interest Groups." *Journal of Politics* 73: 412–27.

Anzia, Sarah F. 2012. "The Election Timing Effect: Evidence from a Policy Intervention in Texas." *Quarterly Journal of Political Science* 7: 209–48.

Anzia, Sarah F. 2014. *Timing and Turnout: How Off-Cycle Elections Favor Organized Groups*. Chicago: University of Chicago Press.

Arnold, R. Douglas, and Nicholas Carnes. 2012. "Holding Mayors Accountable: New York's Executives from Koch to Bloomberg." *American Journal of Political Science* 56: 949–63.

Ashworth, Scott. 2012. "Electoral Accountability: Recent Theoretical and Empirical Work." *Annual Review of Political Science* 15: 183–201.

Barrows, Samuel. 2015. "Performance Information and Retrospective Voting: Evidence from a School Accountability Regime." Harvard University. Program on Education Policy and Governance, Working Paper 15-03. http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG15_03.pdf.

Berry, Christopher R., and Jacob E. Gersen. 2010. "Voters, Non-Voters and the Implications of Election Timing for Public Policy." University of Chicago Law School Law and Economics Working Paper Series, Working Paper 536. http://www.law.uchicago.edu/files/file/536-324-jg-voters.pdf (accessed September 6, 2016).

Berry, Christopher R., and Jacob E. Gersen. 2011. "Election Timing and Public Policy." *Quarterly Journal of Political Science* 6: 103–35.

Berry, Christopher R., and William G. Howell. 2007. "Accountability and Local Elections: Rethinking Retrospective Voting." *Journal of Politics* 69: 844–58.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics* 114: 577–99.

Cain, Bruce E., and Thad Kousser. 2004. *Adapting to Term Limits: Recent Experiences and New Directions*. San Francisco: Public Policy Institute of California.

California School Board Association. 2007. "School Board Leadership: The Role and Function of California's School Boards." https://www.csba.org/~/media/51E3FBB839504700825CB16B7265F3C4.ashx (accessed September 6, 2016).

Chetty, Raj, John Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126: 1592–1660.

Chingos, Matthew M., Grover J. Whitehurst, and Michael R. Gallaher. 2015. "School Districts and Student Achievement." *Education Finance and Policy* 10: 378–98.

Clark, Tom S., and Drew A Linzer. 2015. "Should I Use Fixed or Random Effects?" *Political Science Research and Methods* 3: 399–408.

Clinton, Joshua D., and Jason A. Grissom. 2015. "Public Information, Public Learning and Public Opinion: Democratic Accountability in Education Policy." *Journal of Public Policy* 35: 355–85. doi:10.1017/S0143814X14000312.

Dervarics, Chuck, and Eileen O'Brien. 2011. "Eight Characteristics of Effective School Boards." Center for Public Education. http://www.centerforpubliceducation.org/ (accessed September 6, 2016).

Erikson, Robert S., Michael B. MacKuen, and James A. Stimson. 2000 "Bankers or Peasants Revisited: Economic Expectations and Presidential Approval." *Electoral Studies* 19: 295–312.

Fearon, James D. 1999. "Electoral Accountability and the Control of Politicians: Selecting Good Type versus Sanctioning Poor Performance." In *Democracy, Accountability, and Representation*, ed. A. Przeworski, S. Stokes, and B. Manin. Cambridge: Cambridge University Press, 55–97.

Fensterwald, John. 2015. "Test Scores to Be Only One Factor in Measuring School Progress." *EdSource*. http://edsource.org/2015/new-school-accountability-system-will-de-emphasize-smarter-balanced/86317 (accessed September 6, 2016).

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50: 159–73.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.

Fischel, William A. 2001. *The Homevoter Hypothesis: How Home Values Influence Local Government Taxation, School Finance, and Land-Use Policies*. Cambridge, MA: Harvard University Press.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Greene, William H. 2012. *Econometric Analysis*, 7th ed. Upper Saddle River, NJ: Prentice Hall.

Hajnal, Zoltan L., Paul G. Lewis, and Hugh Louch. 2002. *Municipal Elections in California: Turnout, Timing, and Competition.* San Francisco, California: Public Policy Institute of California http://www.ppic.org/content/pubs/report/R_302ZHR.pdf (accessed September 6, 2016)

Healy, Andrew J., and Neil Malhotra. 2009. "Myopic Voters and Natural Disaster Policy." *American Political Science Review* 103: 387–406.

Healy, Andrew J., and Neil Malhotra. 2013. "Retrospective Voting Reconsidered." *Annual Review of Political Science* 16: 285–306.

Healy, Andrew J., Neil Malhotra, and Cecilia Hyunjung Mo. 2010. "Irrelevant Events Affect Voters' Evaluations of Government Performance." *Proceedings of the National Academy of Sciences* 107: 12804–09.

Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Cambridge, MA: Harvard University Press.

Holbein, John B. 2016. "Left Behind? Citizen Responsiveness to Government Performance Information." *American Political Science Review* 110: 353–68.

Hopkins, Daniel J., and Lindsay M. Pettingill. 2015. "Economic Voting in Big-City U.S. Mayoral Elections." http://ssrn.com/abstract=2567294 (accessed September 6, 2016).

Key, V. O. 1966. *The Responsible Electorate*. New York: Vintage Books.

Kogan, Vladimir, Stephane Lavertu, and Zachary Peskowitz. 2014. "Performance Measurement and School Board Elections: Do School Report Cards Produce Accountability at the Ballot Box?" http://glenn.osu.edu/educational-governance/research/research-attributes/KLP_Turnover_20160708.pdf (accessed September 6, 2016).

Larsen, Eric S. 2009. *School Accountability and Administrator Incentives in California.* San Francisco: Public Policy Institute of California. http://www.ppic.org/content/pubs/report/R_709EL2R.pdf (accessed September 6, 2016).

Lewis-Beck, Michael S., and Mary Stegmaier. 2000. "Economic Determinants of Electoral Outcomes." *Annual Review of Political Science* 3: 183–219.

Long, Cindy, and Sara Robertson. 2014. "NEA Launches Campaign to End 'Toxic Testing'." National Education Association. http://www.nea.org/home/59747.html (accessed September 6, 2016).

Long, J. Scott, and Ervin, Laurie H. 2000. "Using Heteroskedasticity Consistent Standard Errors in the Linear Regression Model." *American Statistician* 54: 217–24.

MacKinnon, James G., and Halbert White. 1985. "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29: 305–25.

Manin, Bernard, Adam Przeworski, and Susan C. Stokes. 1999. "Elections and Repre-
sentation." In *Democracy, Accountability, and Representation*, ed. A.
Przeworski, S. Stokes, and B. Manin. Cambridge: Cambridge University Press,
29–54.
Meredith, Marc. 2009. "The Strategic Timing of Direct Democracy." *Economics and
Politics* 21: 159–77.
Moe, Terry M. 2006. "Political Control and the Power of the Agent." *Journal of Law,
Economics, and Organization* 22: 1–29.
Moe, Terry M. 2011. *Special Interest: Teachers Unions and America's Public Schools*.
Washington, DC: Brookings Institution Press.
Oliver, J. Eric. 2012. *Local Elections and the Politics of Small-Scale Democracy*.
Princeton, NJ: Princeton University Press.
Oliver, J. Eric, and Shang E. Ha. 2007. "Vote Choice in Suburban Elections." *American
Political Science Review* 101: 393–408.
Peterson, Paul E., Michael Henderson, and Martin R. West. 2014. *Teachers versus the
Public: What Americans Think about Their Schools and How to Fix Them*.
Washington, DC: Brookings Institution Press.
Trounstine, Jessica L. 2009. "All Politics Is Local: The Reemergence of the Study of
City Politics." *Perspectives on Politics* 7: 105–33.
Tufte, Edward R. 1978. *Political Control of the Economy.* Princeton, NJ: Princeton Uni-
versity Press.
US Census Bureau. 2012. "U.S. Census of Governments." https://www.census.gov/
newsroom/releases/archives/governments/cb12-161.html (accessed September 6,
2016).

Supporting Information

Additional Supporting Information may be found in the online version
of this article.

Data Set Construction
Figure A1: Change in District API Scores and Number of Incumbents
Seeking Election
Table A1: API Scores and Number of Incumbents Seeking Re-
Election
Table A2: One-Year Change in Test Scores and Incumbent Perfor-
mance (All Years)
Table A3: One-Year Change in Test Scores and Incumbent
Performance
Table A4: One-Year Change in Test Scores and Incumbent Perfor-
mance: Random District Effects
Table A5: API Levels and Incumbent Performance