

Contents

Contents	A1
A1 Robustness to Different Measures of Election Timing	A2
A2 Calculating the Fiscal Cost of Absence	A4
A3 Consistency of Absence Measures Between Data Sources	A5
A4 “Cooking the Books” or Reduced Absenteeism?	A7
A5 Matching Schools to Assembly Constituencies	A18
A6 Systematic Measurement Error	A22
A7 Full Results Tables	A25

Table A1: Absence by Years to Election in Government Schools

	Absence					Log Average Absence				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Years to Election	-0.003 (0.006)	-0.004 (0.005)	-0.006 (0.005)	-0.009* (0.004)	-0.006 (0.005)	-0.019 (0.029)	-0.025 (0.026)	-0.032 (0.026)	-0.049** (0.022)	-0.006 (0.005)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Number of Schools	1,156,790	1,156,790	1,156,790	1,156,790	1,233,325	1,156,790	1,156,790	1,156,790	1,156,790	1,233,325
Observations	9,744,359	9,744,359	9,744,359	9,744,359	11,013,448	9,744,359	9,744,359	9,744,359	9,744,359	11,013,448

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors clustered at the constituency-year level in parentheses. The dependent variable in columns 1-5 is a dummy variable that takes the value of one if the school reports any teacher absenteeism in that year and the log number of absences per teacher in columns 6-10. Each specification includes controls for the number of teachers in each school, a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 include a lagged dependent variable. “Years to Election” is a variable equal to the number of years that the school is from an election.

A1 Robustness to Different Measures of Election Timing

A potential concern with my results is that the results might be sensitive to the specification of the functional form of election timing. In this appendix, I re-estimate the results from Figures 2 to 3 using the number of years to the election instead of individual electoral cycle dummies.

A1.1 Using A Continuous Measure of Distance to Elections

I run a regression using the number of years to the next election. The formal equation takes the form:

$$Y_{i,t} = \beta_1 \text{Years to Next Election}_{i,t} + \beta_2 y_{i,t-1} + Z_{i,t} + \gamma_i + \zeta_t + \mu_{i,t,d}, \tag{A1}$$

where $\text{Years to Next Election}_{i,t}$ is a continuous variable that indicates the number of years school i is to the next legislative election, and the rest of the equation is as in Equations 1.

I present results for government schools in Table A1 and for private schools in Table A2. The results using a continuous measure of years to election are largely the same as other specifications, but it is important to note that the results for private schools are significant in all specifications that include lagged dependent variables (columns 1-4 and 6-9). The point estimates are between one-third and one-half smaller than those for government schools, suggesting that if any effect in this specification, it is smaller and weaker.

Table A2: Absence by Years to Election in Private Schools

	Absence					Log Average Absence				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Years to Election	0.000 (0.003)	0.000 (0.002)	0.000 (0.002)	-0.001 (0.002)	0.000 (0.002)	0.003 (0.014)	0.001 (0.013)	0.000 (0.010)	-0.005 (0.009)	0.000 (0.002)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Number of Schools	503,330	503,330	503,330	503,330	579,172	503,330	503,330	503,330	503,330	579,172
Observations	2,790,985	2,790,985	2,790,985	2,790,985	3,331,054	2,790,985	2,790,985	2,790,985	2,790,985	3,331,054

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the constituency-year level in parentheses. The dependent variable in columns 1-5 is a dummy variable that takes the value of one if the school reports any teacher absenteeism in that year and the log number of absences per teacher in columns 6-10. Each specification includes controls for the number of teachers in each school, a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 include a lagged dependent variable. "Years to Election" is a variable equal to the number of years that the school is from an election.

A2 Calculating the Fiscal Cost of Absence

In this section, I calculate the total amount of money lost to absenteeism as a function of teachers' wages, and the total amount of money recovered in election years from the highest level of absenteeism. Table A3 provides calculations of the estimate of how much money is lost and recovered between election and non-election years.

Table A3: Calculating the Fiscal Recovery of Reduced Absenteeism

	Highest Wage Estimate	Lowest Wage Estimate
Average Monthly Wage (₹)	11,368	5,418
Average Daily Wage (₹)	454.72	216.72
Average Absence per School (Days)	5.16	
Average Number of Teachers per Year	4,274,454	
Wages Lost in Average Year (₹)	10,038,743,354	4,784,474,973
Wages Lost in Average Year (\$)	218,233,551	104,010,325
Reduction in Absenteeism per School (Days)	1.76	
Wage Recovery in Year Before Election (₹)	3,413,443,758	1,626,850,658
Wage Recovery in Year Before Election (\$)	74,205,299	35,366,319
Share of Wages Lost Recovered (%)	34	

All costs in 2010 prices. The high wage estimate is taken from [Muralidharan et al. \(2017\)](#) and the low wage estimate is taken from [Kingdon and Sipahimalani-Rao \(2010\)](#). The mean level of absence is the mean government school absence in the entire DISE dataset, and the average number of teachers per year is the mean number of teachers in the DISE dataset across all years. The wages lost due to absenteeism in the average year in Rupees (₹) is the average daily wage multiplied by the mean number of absences and by the mean number of teachers. The wages lost in USD Dollars (\$) divides the number by the exchange rate in 2010. The reduction in absenteeism is calculated by taking the difference in the point estimates from “1 Year After Election” and “1 Year from Election” in Column 4 in Table A11. The wage recovery multiplies the reduction in absenteeism by the average daily wage and the average number of teachers, while the share of wages lost recovered divides the wage recovery by the wages lost.

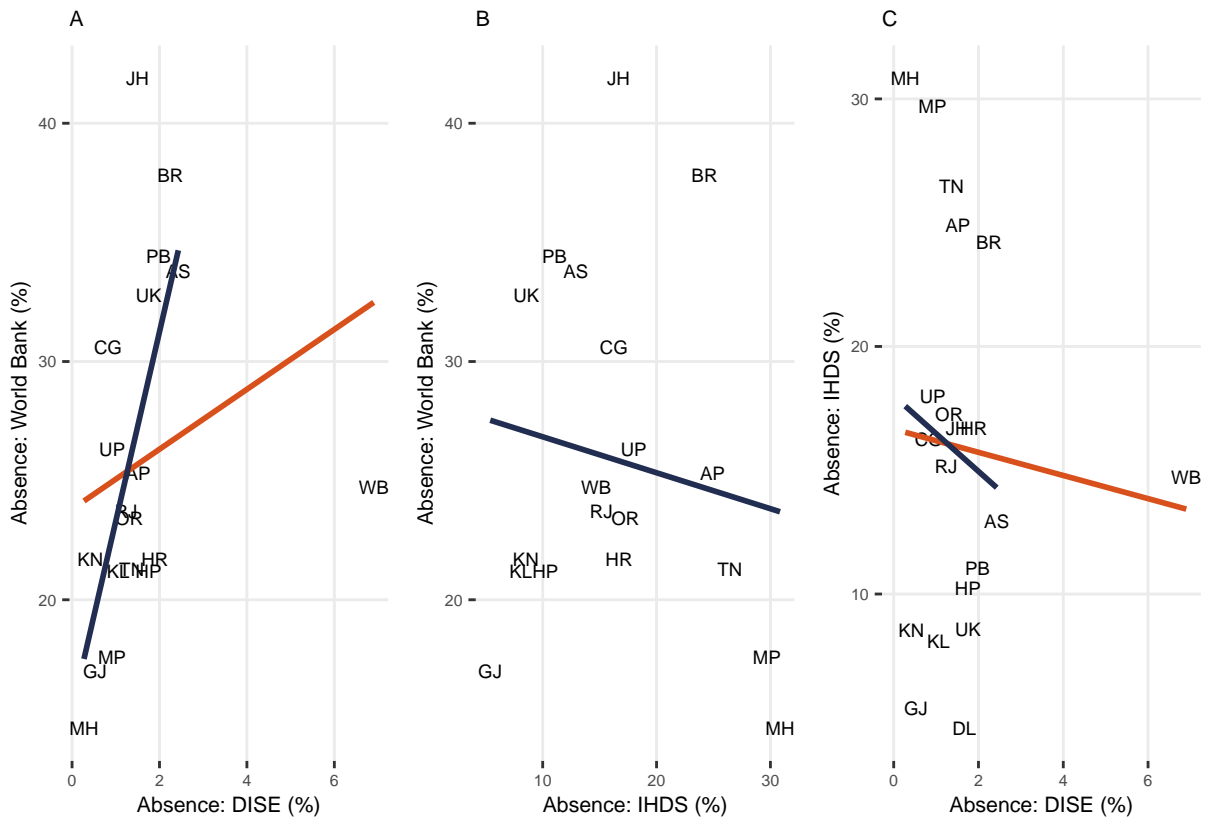
A3 Consistency of Absence Measures Between Data Sources

I rely on two different measures of teacher absenteeism in the paper, the DISE School Report Cards that are self-reported by head teachers in schools, and the Indian Human Development Survey (IHDS) from 2011-12 that is independently collected by the National Council for Applied Economic Research through unannounced visits to schools.

I compare these two measures to the measures from a third independently collected source, the data from [Kremer et al. \(2005\)](#). In Figure [A1](#), I provide state-level scatter plots of the levels of absenteeism. There is a strong positive relationship between measures in the DISE School Report Cards and the World Bank's from [Kremer et al. \(2005\)](#) (Panel A). This relationship is stronger removing the outlier state of West Bengal from the DISE School Report Cards data.

The relationship is weaker between the IHDS and the other two measures, with a negative relationship for both measures (Panels B & C), and null when removing West Bengal in the DISE School Report Cards (Panel C).

Figure A1: Consistency of Absence Measures Between Data Sources



Notes: Panel A plots absence in DISE against absence in Table 2 of Kremer et al. (2005). Panel B plots absence in the 2011-12 round of the IHDS against absence in Table 2 of Kremer et al. (2005). Panel C plots absence in DISE against absence in the 2011-12 round of the IHDS. Each point represents the state-level average of the probability that an individual teacher was absent on the day of the school visit. For the DISE SRC data, this represents the teacher-level absence per working day.

A4 “Cooking the Books” or Reduced Absenteeism?

This paper uses rich administrative data to answer an important political problem. While the use of administrative data is increasingly common (Lindgren, Oskarsson and Dawes, 2016; Gulzar and Pasquale, 2017), I do not take the quality of this data at face value. Instead, I verify the quality of the data by triangulating against independently collected sources (Herrera and Kapur, 2007). Administrative data suffers from the additional concern that bureaucrats have an incentive to misreport in ways that make their performance look better (Martínez, 2022), on top of all the data quality concerns of other sources of data. Verified against other sources of data, however, administrative data provides great potential as it allows us to answer big questions at scale, especially as the data gathering capacity of states improves (Jerven, 2013; Jensenius and Verniers, 2017).

A concern with the self-reported nature of the DISE school report cards is that head teachers and district-level officers might be “cooking the books”, or falsely reporting lower rates of absenteeism, around elections to make bureaucratic effort look greater with no underlying change in behavior. The empirical evidence of “cooking the books” would be substantively similar – lower absenteeism in election years – although the mechanisms would be different – perceived pressure to modify government data rather than hold providers accountable. Reported absenteeism of 12 percent in the DISE data and 14 percent in the IHDS data in Table 1 is lower than absenteeism in independent audits conducted in 2005 (Chaudhury et al., 2006; Muralidharan et al., 2017), suggesting either a secular decline in absenteeism or in the *reporting* of absenteeism, suggestive evidence of “cooking the books”.

To separate whether schools are “cooking the books” or there is electoral pressure on teachers, I use the IHDS school surveys, an independent survey of schools conducted in 2011-2012 and contemporaneously to the DISE school report cards data collection to test whether I find similar patterns in this data. The IHDS data collects a broad range of demographic and socioeconomic data, and also conducts a school survey of the largest private and government school in each village they survey households. Vis-

its to schools are otherwise unannounced and unexpected by school staff and replicate the randomized audits in [Chaudhury et al. \(2006\)](#).

I use the second wave of the Indian Human Development Survey (IHDS) for data on teacher absenteeism and reasons for absenteeism. The IHDS survey is a nationally representative survey of 1,503 villages and 971 urban neighborhoods across India ([Desai, Vanneman and National Council of Applied Economic Research, New Delhi, 2015](#)). The 2011-2012 round was the second round of a panel survey that surveyed the same villages in 2004-2005. The survey asked a small number of questions about the largest government and private school in each surveyed village, and the next two largest schools, irrespective of whether they were public or private. For every teacher in the school, the surveyors checked if the teacher was present on the day of the survey, and, if they were absent, whether they were absent on officially sanctioned government work. In this sense, the survey replicates the data collection in [Chaudhury et al. \(2006\)](#), randomly auditing schools on staff absenteeism.

In [Table A4](#), I look at the probability that a teacher is absent from the school on the day of the survey, absent from the school on the day of the survey to conduct official work, or present at the school on the day of the interview and present for the interview as a function of whether their school is a government or private school. As expected, teachers are four percentage points more likely to be absent from a government school (columns 1-2), and about two percentage points more likely to be absent for officially sanctioned duty (columns 3-4). Given that government schools have fewer teachers ([Panel C of Table 1](#)), conditional on being at the school teachers in government schools are also more likely to be present at the IHDS interview as there are likely fewer teachers to answer the survey and schools are smaller (columns 5-6). These results confirm common sense expectations of what we think differences in teacher absenteeism should look like between government and private schools, with government schools showing consistently higher levels of absenteeism.

Taking the DISE and IHDS data together, the two sources allow me to triangulate between two

Table A4: Absence by School Type

	Absent		On Official Duty		Absent from Interview	
	(1)	(2)	(3)	(4)	(5)	(6)
Government School	0.043*** (0.004)	0.046*** (0.004)	0.014*** (0.002)	0.016*** (0.002)	-0.036*** (0.007)	-0.038*** (0.006)
Village Fixed Effects	N	Y	N	Y	N	Y
N	30,878	30,878	30,878	30,878	26,448	26,448

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. Linear models of the effects of working in a government school on either absence from school on the day of the interview, absent from the school for official government-sanctioned work, or absent from the interview conditional on being present at the school on the day of the interview. All models control for gender, age, religion, and caste, while even-numbered columns also include village-fixed effects.

otherwise imperfect data sources (Herrera and Kapur, 2007), and draw broader conclusions on the drivers of absence in Indian public service. While the IHDS data effectively provides a randomized audit of a select number of government and private schools in the country, designed to be representative of the country as a whole, the DISE data allows me to generalize the results to *all* schools in India. The consistency in the results between the two sources provides support that the electoral cycles I discover are real and not the product of data quality, manipulation by teachers or school officials, or a result of self-reporting.

A key question surrounding data quality is how self-reported data provided by organizations like DISE compare to independent evaluations of absence from random audits such as in Banerjee and Duflo (2006); Chaudhury et al. (2006). The levels of absence found in this paper are much lower than absence found by independent evaluations of service worker absenteeism from other papers in India. Average levels of absence self-reported in the DISE dataset reach 13 percent for the *year*, far shorter than the levels of absence recorded on random spot checks in Chaudhury et al. (2006) of 25 percent on any given day. For education in India, the DISE data serves as the only comparable source of data available to the government and broader public, and is used by the former to assess the state of schools. While the data is almost certainly biased downwards, it does have important implications for decision-making as

this is the dataset used by policymakers. Finding similar results in the IHDS data adds confidence that the self-reported data is not being systematically manipulated in election years and that we are seeing real decreases in absenteeism. Across all specifications, results are comparable in direction between both sources of data.

This is a larger problem for any study in the social sciences that relies on administrative data. In contexts of low capacity, low attention, or poor measurement, the quality of this data may deviate from common sense understandings or reality. This paper provides one path forward – triangulate administrative data with other sources of high-quality data collected from other sources (Herrera and Kapur, 2007). The benefits of using administrative data are too great to simply ignore them, but we should be cautious in how we employ them and the conclusions we draw from them, ensuring that they are verified through other means. I provide one way forward in this paper by independently verifying results from an administrative dataset against a second source of data collected by a different organization for a different purpose, with different incentives in data collection. It provides one way forward for students of political science interested in employing large data moving forward.

A4.1 Absenteeism in By-Elections

Another concern in cooking the books is that politicians and teachers could have time to coordinate around either actions or reporting on absenteeism before an election. For example, politicians and district-level officers could coordinate to report lower levels of absenteeism in an election year ex-post by reporting lower levels of absenteeism after the election.

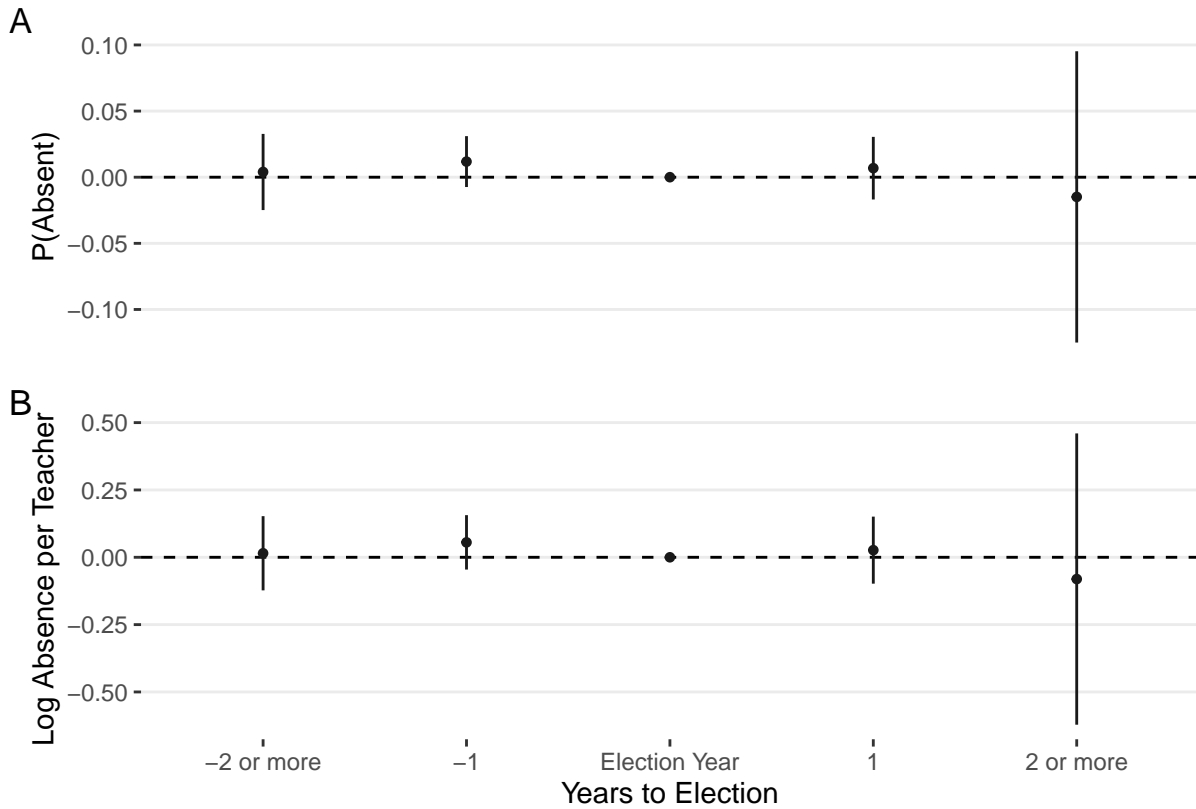
If this were the case, we would see similar effects in elections on a regular five year cycle as well as by-elections, elections held because an MLA has had to step down, often because of a sudden death. By-elections are held quickly often with only months of campaigning, but the winner serves the remainder of the term until the next election. I can leverage by-elections to see if politicians are “cooking the books”

with education officials as after the election they could alter levels of absenteeism. Given the short campaign period and unexpected election timing, we should *not* see an electoral cycle in absenteeism around by-elections if politicians are not cooking the books, but we should see one if they are cooking the books ex-post.

Below, I also provide the same tables and figures I show in the main manuscript of all the by-election elections using data from DISE, and the election immediately preceding and following the by-election. There are a much smaller number of by-elections and schools in constituencies holding by-elections (see the Footers to Tables A5, A6, A7, and A8). The standard errors for “2 or more” years to the next election are large as by-elections cut the time to the next election often to one year or less.

If politicians were “cooking the books” ex-post, we would see similar effects for by-elections as we do in the main set of results using the full set of elections. Politicians would be able to coordinate with education officials at the state or national level to change data, which we see no evidence of here. At the same time, given the constrained campaigning window in by-elections, which are often held within a couple of months after the election is declared, it adds confidence in the results as we should not expect politicians to be able to exert pressure over the entirety of a constituency within such a short time period.

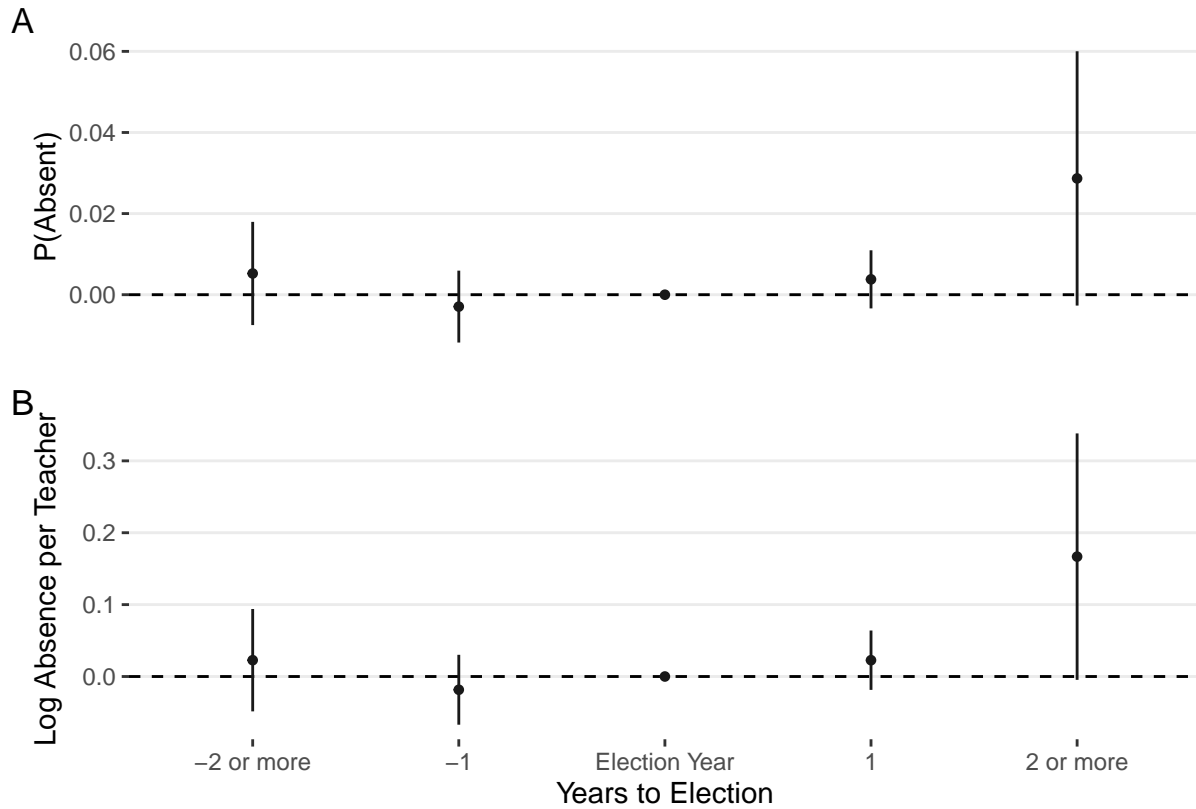
Figure A2: Absence in a School Year over the Electoral Cycle in Government Schools During By-Elections



Notes: In Panel A, the dependent variable is a dummy variable that takes the value of one if a school reports any teacher absenteeism in that year. In Panel B, the dependent variable is the log number of absences per teacher. The regression includes controls for the number of teachers in a school, a dummy for whether the school is in a rural area, and year and school fixed effects. The line represents 95% confidence intervals with standard errors clustered at the constituency-year level. There are 345,858 school-year observations and 107,329 in Panel A, and 345,858 and 107,329 total schools in Panel B. The election year mean is 0.134 in Panel A and 0.005 in Panel B. Panel A corresponds to Column 9 in Table A5 and Panel B corresponds to Column 9 in Table A6.

Data Source: District Information System for Education (DISE) School Report Cards.

Figure A3: Absence in a School Year over the Electoral Cycle in Private Schools During By-Elections



Notes: In Panel A, the dependent variable is a dummy variable that takes the value of one if a school reports any teacher absenteeism in that year. In Panel B, the dependent variable is the log number of absences per teacher. The regression includes controls for the number of teachers in a school, a dummy for whether the school is in a rural area, and year and school fixed effects. The line represents 95% confidence intervals with standard errors clustered at the constituency-year level. There are 109,511 school-year observations and 38,012 schools in Panel A, and 109,511 and 38,012 total schools in Panel B. The election year mean is 0.031 in Panel A and 0.005 in Panel B. Panel A corresponds to Column 9 in Table A7 and Panel B corresponds to Column 9 in Table A8.

Data Source: District Information System for Education (DISE) School Report Cards.

Table A5: Any Absence over the Electoral Cycle in Government Schools Using DISE Data For By-Elections

	Absent									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2 or More Years Before Election	0.007 (0.043)	0.015 (0.036)	0.006 (0.023)	-0.001 (0.014)	0.004 (0.015)	0.010 (0.044)	0.019 (0.036)	0.009 (0.023)	0.001 (0.014)	0.006 (0.014)
1 Year Before Election	0.011 (0.018)	0.004 (0.007)	-0.003 (0.019)	0.011 (0.011)	0.012 (0.010)	0.013 (0.019)	0.007 (0.008)	-0.003 (0.018)	0.010 (0.011)	0.011 (0.010)
1 Year After Election	0.021 (0.019)	0.019 (0.014)	0.021 (0.018)	0.004 (0.015)	0.007 (0.012)	0.023 (0.020)	0.017 (0.014)	0.021 (0.018)	-0.001 (0.014)	0.002 (0.012)
2 or More Years After Election	-0.024 (0.033)	0.034 (0.023)	-0.171* (0.091)	-0.024 (0.059)	-0.015 (0.056)	-0.023 (0.034)	0.036 (0.023)	-0.170* (0.090)	-0.021 (0.059)	-0.012 (0.057)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Teacher Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of Schools	107,329	107,329	107,329	107,329	118,267	106,909	106,909	106,909	106,909	117,520
Observations	351,609	351,609	351,609	351,609	400,689	345,858	345,858	345,858	345,858	393,540

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the constituency-year level in parentheses. The dependent variable is a dummy variable that takes the value of one if the school reports any teacher absenteeism in that year. Each specification includes a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 also include a lagged dependent variable. The election year mean is 0.134. Column 9 corresponds to Panel A of Figure A2.

Table A6: Log Average Absence in Government Schools over the Electoral Cycle Using DISE Data For By-Elections

	Log Average Absence									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2 or More Years Before Election	0.038 (0.202)	0.070 (0.168)	0.030 (0.115)	-0.007 (0.071)	0.015 (0.070)	0.047 (0.206)	0.085 (0.169)	0.042 (0.114)	0.005 (0.070)	0.025 (0.067)
1 Year Before Election	0.069 (0.093)	0.031 (0.042)	-0.008 (0.097)	0.053 (0.058)	0.055 (0.052)	0.075 (0.098)	0.040 (0.049)	-0.005 (0.094)	0.046 (0.058)	0.049 (0.053)
1 Year After Election	0.096 (0.093)	0.084 (0.076)	0.100 (0.089)	0.013 (0.077)	0.027 (0.063)	0.101 (0.097)	0.076 (0.076)	0.097 (0.089)	-0.011 (0.072)	-0.002 (0.061)
2 or More Years After Election	-0.160 (0.162)	0.128 (0.110)	-0.863* (0.453)	-0.128 (0.290)	-0.081 (0.276)	-0.156 (0.165)	0.142 (0.113)	-0.860* (0.446)	-0.111 (0.290)	-0.067 (0.279)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Teacher Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of Schools	107,329	107,329	107,329	107,329	118,267	106,909	106,909	106,909	106,909	117,520
Observations	351,609	351,609	351,609	351,609	400,689	345,858	345,858	345,858	345,858	393,540

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the constituency-year level in parentheses. The dependent variable is the log number off average absences per school. Each specification includes a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 also include a lagged dependent variable. The election year mean is 0.005. Column 9 corresponds to Panel B of Figure A2.

Table A7: Any Absence over the Electoral Cycle in Private Schools Using DISE Data For By-Elections

	Absent									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2 or More Years Before Election	0.004 (0.008)	0.001 (0.008)	0.007 (0.011)	0.004 (0.008)	0.005 (0.006)	0.003 (0.008)	0.001 (0.008)	0.007 (0.011)	0.003 (0.008)	0.005 (0.007)
1 Year Before Election	-0.002 (0.007)	-0.003 (0.006)	-0.002 (0.005)	-0.002 (0.005)	-0.003 (0.005)	-0.002 (0.007)	-0.003 (0.006)	-0.002 (0.005)	-0.003 (0.005)	-0.003 (0.004)
1 Year After Election	0.000 (0.006)	0.002 (0.003)	0.003 (0.003)	0.001 (0.003)	0.004 (0.004)	0.000 (0.006)	0.003 (0.003)	0.003 (0.003)	0.001 (0.003)	0.003 (0.003)
2 or More Years After Election	0.007 (0.007)	0.014** (0.005)	-0.005 (0.016)	0.026 (0.018)	0.029* (0.016)	0.010 (0.007)	0.017** (0.006)	-0.003 (0.016)	0.030 (0.018)	0.032* (0.017)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Teacher Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of Schools	38,012	38,012	38,012	38,012	45,067	37,452	37,452	37,452	37,452	44,256
Observations	112,446	112,446	112,446	112,446	135,241	109,511	109,511	109,511	109,511	131,092

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the constituency-year level in parentheses. The dependent variable is a dummy variable that takes the value of one if the school reports any teacher absenteeism in that year. Each specification includes a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 also include a lagged dependent variable. The election year mean is 0.031. Column 9 corresponds to Panel A of Figure A3.

Table A8: Log Average Absence in Private Schools over the Electoral Cycle Using DISE Data For By-Elections

	Log Average Absence									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2 or More Years Before Election	0.009 (0.046)	-0.004 (0.047)	0.035 (0.058)	0.015 (0.046)	0.023 (0.036)	0.006 (0.048)	-0.005 (0.049)	0.032 (0.058)	0.013 (0.046)	0.022 (0.037)
1 Year Before Election	-0.012 (0.039)	-0.015 (0.035)	-0.015 (0.028)	-0.015 (0.025)	-0.018 (0.025)	-0.011 (0.041)	-0.014 (0.036)	-0.018 (0.029)	-0.020 (0.026)	-0.020 (0.023)
1 Year After Election	-0.004 (0.032)	0.013 (0.018)	0.009 (0.017)	0.007 (0.017)	0.023 (0.021)	-0.004 (0.034)	0.017 (0.020)	0.009 (0.017)	0.006 (0.017)	0.016 (0.019)
2 or More Years After Election	0.025 (0.036)	0.063** (0.025)	-0.030 (0.078)	0.146 (0.091)	0.167* (0.087)	0.041 (0.038)	0.084** (0.029)	-0.019 (0.080)	0.164 (0.094)	0.184* (0.091)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Teacher Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of Schools	38,012	38,012	38,012	38,012	45,067	37,452	37,452	37,452	37,452	44,256
Observations	112,446	112,446	112,446	112,446	135,241	109,511	109,511	109,511	109,511	131,092

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the constituency-year level in parentheses. The dependent variable is the log number of average absences per school. Each specification includes a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 also include a lagged dependent variable. The election year mean is 0.005. Column 9 corresponds to Panel B of Figure A3.

A5 Matching Schools to Assembly Constituencies

As the DISE school report card data does not identify which Assembly Constituency a school is located in, I use geographic information on the school to place the school in an Assembly Constituency. This matching proceeded in four steps:

1. Using the precise location of the school
2. Using the location of the village in which the school is located
3. Using data from [Adukia, Asher and Novosad \(2019\)](#) (AAN) to cross-reference unmatched schools to geographic locations
4. Using the postal pincode of the school to match the school to the Assembly Constituency

I provide a further description of each step below, and a matching rate table in [Table A9](#). The overall match rate for both public and private schools was 85.5 %

Table A9: Matching Rate by Matching Strategy

Matching Strategy	Government Schools				Private Schools			
	Remaining Unmatched	Number Matched	Match Rate (%)	Overall Match Rate (%)	Remaining Unmatched	Number Matched	Match Rate (%)	Overall Match Rate (%)
GIS	1,494,885	979,532	65.5	65.5	682,771	271,953	39.8	39.8
Village Match	515,353	179,373	34.8	12.0	410,818	265,354	63.6	38.9
AAN	335,980	81,536	24.3	5.5	145,464	17,388	12.0	2.5
Pincodes	254,444	40,324	15.8	2.7	128,076	27,497	21.5	4.0
Total		1,280,765		85.7		582,192		85.3

Notes: The Matching Strategy column identifies the strategy used to match schools to assembly constituencies. The Remaining Unmatched columns report the number of schools left to match after all previous matching strategies, the Number Matched columns report how many schools were matched using that particular matching strategy, the Match Rate columns report the percentage of schools matched of all remaining schools left to match, and the Overall Match Rate columns report the percentage of schools matched by that strategy out of all the schools in the data set for government and private schools respectively

A5.1 School GIS

The Government of India provides georeferenced information on many schools in India at <https://schoolgis.nic.in/>. I scraped the site and merged the locations with the DISE school report cards using the school code provided in each data set. I then used a spatial join with Assembly Constituency shapefiles to identify the Assembly Constituency in which the school was located.

A5.2 Village Codes

Next, the first nine digits of each school's school code identifies the village in which a school is located. For unmatched schools located in a village with a matched school, I coded that school as located in the same assembly constituency.

A5.3 Adukia, Asher and Novosad (2019)

Adukia, Asher and Novosad (2019) provide a crosswalk between DISE village codes and Census of India village codes. For any remaining unmatched schools, I use the nine-digit DISE village code to match schools to Census villages. Then, I use village-level shapefiles to spatially join villages to Assembly Constituencies, and code schools in the Assembly Constituency they are located.

A5.4 Postal Pincodes

Finally, for the remaining unmatched villages, each school observation in the school report cards data reports the postal pincode in which the school is located. I geo-reference these pincodes using Google Maps and take the centroid of the pincode. Using the latitude and longitude of the centroid, I place this in Assembly Constituencies and code the school as being in that Assembly Constituency.

A5.5 Differences Between Matched and Unmatched Schools

I then test for differences in matched and unmatched schools. Figure [A4](#) plots the differences in means. Given the large sample size and high match rate, most variables have significant differences, although their substantive sizes are small. For example, rural schools are eight percent more likely to be matched, but given that 85.5 % of schools are matched and of those, 86% are rural, this means that 83% of schools in the population are rural. These differences are unlikely to lead to systematic bias in the results.

Figure A4: Difference in Means Between Schools Matched to their Assembly Constituency and Unmatched Schools



Notes: Each point estimate is a t-test of the difference between schools I was able to place in an assembly constituency and unmatched schools. For continuous variables, variables are standardized to range from 0-1 by subtracting the mean value and dividing by two standard deviations (Gelman, 2008).

The plot is ordered from largest to smallest value.

A6 Systematic Measurement Error

Given the lower levels of absenteeism reported in the DISE data relative to data collected independently by the World Bank in [Chaudhury et al. \(2006\)](#) and in the IHDS ([Desai, Vanneman and National Council of Applied Economic Research, New Delhi, 2015](#)), the DISE data is likely to contain some level of measurement error. A greater concern for the identification strategy in this paper is if this measurement error and underreporting is systematic to the treatment and only occurs in pre-election years where there are greater incentives for politicians and education administrators to underreport absenteeism.

Here, I explore how large the systematic measurement error would have to be to overturn the results in the paper. To do so, I conduct a placebo test where I replace the value of pre-election year, and only pre-election year, absenteeism to either the maximum or median observed at that school for a randomly sampled number of constituencies. To do so, I randomly sample between 10 to 100 constituencies in increments of 10 200 times each. For every school within {10, 20, 30, 40, 50, 60, 70, 80, 90, 100} randomly sampled constituencies, I replace their pre-election absenteeism to either be 1 (for any absence reported), the median value observed in that school, or the maximum value observed in that school (Panels A, B, and C of [Figure A5](#) respectively). I randomly sample the {10, 20, 30, 40, 50, 60, 70, 80, 90, 100} constituencies 200 times each and take the mean of the point estimate and 95% confidence interval in [Figure A5](#).

For example if a school within a randomly sampled constituency reported absenteeism rates of {12, 8, 6, 0, 10, 9, 8, 10, 9, 5, 7, 9} (pre-election years in red), in the placebo test, the resulting values would be:

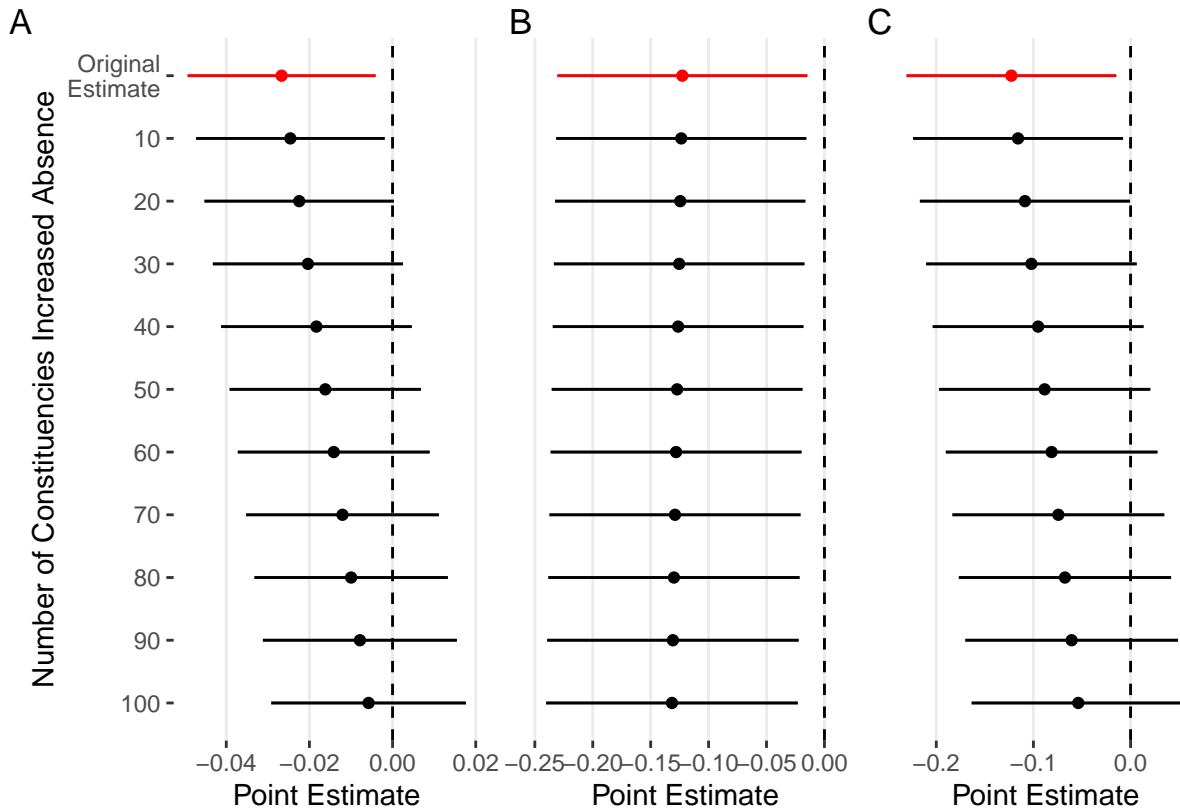
Panel A: 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

Panel B: 12, 8, 6, 8.5, 10, 9, 8, 10, 8.5, 5, 7, 9

Panel C: 12, 8, 6, 12, 10, 9, 8, 10, 12, 5, 7, 9

There are between 800 to 1,000 constituencies that hold elections every year, so this represents

Figure A5: Placebo Test Increasing Schools within an Increasing Share of Constituencies



Notes: The placebo test increases the pre-election year absenteeism for all schools in a given number of constituencies. The y-axis reports the number of randomly sampled constituencies in which all the schools have their pre-election year absenteeism replaced, increasing from 10 to 100 constituencies. I estimate Equation 1 200 times for each level of constituencies, randomly sampling the number of constituencies that have their absenteeism increased without replacement. I present the mean value of the pre-election year point estimate from these 200 estimations. The dependent variable in Panel A is a dummy that takes the value of 1 if the school reports any absenteeism in a year and 0 otherwise. For the randomly sampled number of constituencies I set the pre-election year dependent variable to 1 for all schools. The dependent variable in Panels B and C is the logged number of absences in a school-year. In Panel B, for the randomly sampled number of constituencies, I set the pre-election year dependent variable to the within-school median. In Panel C, for the randomly sampled number of constituencies, I set the pre-election year dependent variable to the within-school maximum.

Data Source: District Information System for Education.

between 1 to 12.5 percent of all constituencies in an election year. The strong assumption here is that if a politician is working with education officials to misreport absenteeism in the year before an election, they would misreport absenteeism for *all* the schools in their constituency and the observed value was misreported from the highest possible value ever seen in that school.¹ As constituencies hold an average of three elections in the data, I am setting the pre-election absenteeism rate for three sets of elections in these models. Panels A and C represent hard tests of the measurement error, assuming that *all* the schools within the randomly sampled constituencies misreport absenteeism from the maximum possible value observed in the data, while Panel B presents an upper bound bounded by the median level reported in the school.

In Panels A and C, between 20 and 30 constituencies have to misreport absenteeism (or between 2 to 4 percent of the data) and the misreported absenteeism has to be the maximum possible level of observed absenteeism in a school to make the results insignificant. There is no number of constituencies that misreport absenteeism in Panel B that overturn the results. This represents a strict test and suggests that the level of measurement error would have to be high to overturn the results.

¹This is a similar bounding exercise to Mansky bounds ([Horowitz and Manski, 2000](#)).

Table A10: Any Absence over the Electoral Cycle in Government Schools Using DISE Data

	Absent									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2 or More Years Before Election	-0.033 (0.021)	-0.033 (0.023)	-0.022 (0.015)	-0.018 (0.020)	-0.022 (0.019)	-0.033 (0.020)	-0.034 (0.023)	-0.022 (0.016)	-0.019 (0.020)	-0.025 (0.019)
1 Year Before Election	-0.034* (0.018)	-0.034* (0.018)	-0.028* (0.015)	-0.026* (0.012)	-0.025* (0.012)	-0.034* (0.018)	-0.034* (0.017)	-0.028* (0.014)	-0.027** (0.012)	-0.028** (0.011)
1 Year After Election	0.008 (0.027)	0.015 (0.025)	0.009 (0.022)	0.021 (0.019)	0.011 (0.020)	0.008 (0.027)	0.014 (0.024)	0.009 (0.023)	0.020 (0.019)	0.008 (0.021)
2 or More Years After Election	-0.022 (0.014)	-0.013 (0.014)	-0.023 (0.014)	-0.007 (0.015)	-0.005 (0.014)	-0.015 (0.015)	-0.005 (0.014)	-0.019 (0.016)	0.000 (0.017)	-0.001 (0.016)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Teacher Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of Schools	1,167,685	1,167,685	1,167,685	1,167,685	1,249,922	1,163,200	1,163,200	1,163,200	1,163,200	1,239,521
Observations	10,403,870	10,403,870	10,403,870	10,403,870	11,726,051	10,229,591	10,229,591	10,229,591	10,229,591	11,504,662

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the constituency-year level in parentheses. The dependent variable is a dummy variable that takes the value of one if the school reports any teacher absenteeism in that year. Each specification includes a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 also include a lagged dependent variable. The election year mean is 0.145. Column 9 corresponds to Panel A of Figure 2.

A7 Full Results Tables

This section provides the full results tables for the figures presented in Figures 2 to 7. The results for Panel A of Figure 2 is presented in Column 9 of Table A10. The results for Panel B of Figure 2 is presented in Column 9 of Table A11. Irrespective of the specification, we still see evidence of an electoral cycle in government schools, with point estimates ranging from 2.3 to 3.8 percentage point reduction in absenteeism in the years before elections. The difference between the point estimate for one year before the election and one year after the election is also significant across all specifications.

The results for Panel A of Figure 4 are presented in Column 9 of Table A12. The results for Panel B of Figure 4 are presented in column 4 of Table A13. Columns 1 to 4 and 6 to 9 provide robustness

Table A11: Log Average Absence in Government Schools over the Electoral Cycle Using DISE Data

	Log Average Absence									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2 or More Years Before Election	-0.156 (0.104)	-0.153 (0.115)	-0.099 (0.085)	-0.079 (0.104)	-0.112 (0.103)	-0.154 (0.101)	-0.156 (0.112)	-0.099 (0.086)	-0.081 (0.104)	-0.125 (0.105)
1 Year Before Election	-0.162* (0.082)	-0.155* (0.081)	-0.128 (0.073)	-0.116* (0.058)	-0.126** (0.057)	-0.161* (0.082)	-0.157* (0.078)	-0.130* (0.072)	-0.123** (0.056)	-0.139** (0.056)
1 Year After Election	0.051 (0.133)	0.101 (0.115)	0.058 (0.116)	0.128 (0.096)	0.061 (0.108)	0.051 (0.131)	0.093 (0.111)	0.058 (0.117)	0.123 (0.095)	0.047 (0.113)
2 or More Years After Election	-0.103 (0.070)	-0.056 (0.070)	-0.098 (0.078)	-0.012 (0.078)	-0.016 (0.077)	-0.071 (0.077)	-0.015 (0.070)	-0.081 (0.089)	0.023 (0.089)	0.006 (0.089)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Teacher Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of Schools	1,167,685	1,167,685	1,167,685	1,167,685	1,249,922	1,163,200	1,163,200	1,163,200	1,163,200	1,239,521
Observations	10,403,870	10,403,870	10,403,870	10,403,870	11,726,051	10,229,591	10,229,591	10,229,591	10,229,591	11,504,662

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the constituency-year level in parentheses. The dependent variable is the log number of average absences per school. Each specification includes a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 also include a lagged dependent variable. The election year mean is 0.005. Column 9 corresponds to Panel B of Figure 2.

checks to modeling choices with and without year and school fixed effects, and Columns 5 and 10 runs the analysis without the lagged dependent variable. All specifications include a dummy for rural schools and columns 6 to 10 include the number of teachers in a school as a measure of the size of the school. Columns 1 to 5 do not control for the number of teachers in the school in case politicians also manipulate the number of teachers in a school around an electoral cycle. The results are substantively similar in all specifications. Like the results in the main body of the paper, we do not see evidence of an electoral cycle in private schools.

Table A14 presents the analysis from Panel A of Figure 3 in Column 1, and Panel B of Figure 3 in Column 2.

Table A12: Any Absence over the Electoral Cycle in Private Schools Using DISE Data

	Absent									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2 or More Years Before Election	-0.006 (0.010)	-0.006 (0.011)	-0.003 (0.007)	-0.003 (0.008)	-0.004 (0.008)	-0.006 (0.010)	-0.006 (0.011)	-0.004 (0.007)	-0.003 (0.008)	-0.006 (0.009)
1 Year Before Election	-0.011 (0.011)	-0.011 (0.011)	-0.006 (0.008)	-0.007 (0.008)	-0.004 (0.009)	-0.011 (0.011)	-0.011 (0.011)	-0.007 (0.008)	-0.007 (0.008)	-0.005 (0.009)
1 Year After Election	-0.005 (0.012)	-0.002 (0.010)	-0.005 (0.009)	-0.002 (0.007)	-0.004 (0.007)	-0.005 (0.012)	-0.002 (0.010)	-0.005 (0.009)	-0.002 (0.008)	-0.005 (0.008)
2 or More Years After Election	-0.009 (0.009)	-0.008 (0.010)	-0.008 (0.005)	-0.005 (0.007)	-0.007 (0.007)	-0.008 (0.010)	-0.006 (0.011)	-0.008 (0.006)	-0.004 (0.008)	-0.005 (0.008)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Teacher Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of Schools	522,520	522,520	522,520	522,520	606,929	514,240	514,240	514,240	514,240	591,590
Observations	3,003,404	3,003,404	3,003,404	3,003,404	3,591,621	2,950,197	2,950,197	2,950,197	2,950,197	3,500,943

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors clustered at the constituency-year level in parentheses. The dependent variable is a dummy variable that takes the value of one if the school reports any teacher absenteeism in that year. Each specification includes a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 also include a lagged dependent variable. The election year mean is 0.047. Column 9 corresponds to Panel A of Figure 4.

Table A15 presents the analysis from Panel A of Figure 5 in Column 1, and Panel B of Figure 5 in Column 2.

Table A16 presents the full results in Table form of Figure 7.

Table A13: Log Average Absence in Private Schools over the Electoral Cycle Using DISE Data

	Log Average Absence									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2 or More Years Before Election	-0.035 (0.056)	-0.037 (0.060)	-0.020 (0.041)	-0.015 (0.047)	-0.026 (0.046)	-0.036 (0.055)	-0.038 (0.059)	-0.020 (0.042)	-0.016 (0.048)	-0.032 (0.050)
1 Year Before Election	-0.056 (0.060)	-0.059 (0.062)	-0.033 (0.042)	-0.035 (0.044)	-0.023 (0.048)	-0.056 (0.060)	-0.060 (0.060)	-0.034 (0.042)	-0.037 (0.045)	-0.028 (0.049)
1 Year After Election	-0.031 (0.063)	-0.017 (0.052)	-0.032 (0.049)	-0.016 (0.040)	-0.026 (0.042)	-0.031 (0.062)	-0.018 (0.052)	-0.031 (0.050)	-0.014 (0.042)	-0.029 (0.046)
2 or More Years After Election	-0.054 (0.052)	-0.048 (0.056)	-0.045 (0.028)	-0.029 (0.037)	-0.040 (0.040)	-0.048 (0.054)	-0.037 (0.059)	-0.043 (0.030)	-0.021 (0.040)	-0.028 (0.043)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Teacher Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of Schools	522,520	522,520	522,520	522,520	606,929	514,240	514,240	514,240	514,240	591,590
Observations	3,003,404	3,003,404	3,003,404	3,003,404	3,591,621	2,950,197	2,950,197	2,950,197	2,950,197	3,500,943

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the constituency-year level in parentheses. The dependent variable is the log number off average absences per school. Each specification includes a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 also include a lagged dependent variable. The election year mean is 0.005. Column 9 corresponds to Panel B of Figure 4.

Table A14: Absence Over Electoral Cycle in Government Schools Using IHDS Data

	Absent	On Official Duty
	(1)	(2)
2 or More Years Before Election	0.113*** (0.030)	-0.127*** (0.042)
1 Year Before Election	0.060** (0.027)	-0.063 (0.044)
1 Year After Election	-0.007 (0.022)	-0.022 (0.057)
2 or More Years After Election	0.074* (0.038)	-0.029 (0.070)
Male	0.006 (0.009)	0.056*** (0.021)
Age	-0.001** (0.000)	0.002** (0.001)
Muslim	-0.029 (0.021)	0.050 (0.043)
Other Religion	0.036 (0.023)	0.121 (0.080)
Upper Caste	-0.006 (0.012)	0.025 (0.027)
SC/ST	0.001 (0.010)	0.006 (0.025)
Other Caste	-0.055** (0.027)	-0.118 (0.077)
Distance from School	0.002*** (0.001)	-0.001 (0.001)
N	14,498	2,383
Election Year Mean	0.15	0.03

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Linear models of the probability of a teacher being absent from a government school the day IHDS surveyors surveyed the school, as well as the probability a teacher was absent on official duty as outlined in Equation 1. The dependent variable is either a dummy variable that takes the value of one if the teacher was absent from the school on the day of the survey or, conditional on being absent, whether they were absent on official duty on the day of the survey.

Table A15: Absence Over Electoral Cycle in Private Schools Using IHDS Data

	Absent	On Official Duty
	(1)	(2)
2 or More Years Before Election	0.113*** (0.030)	-0.028 (0.037)
1 Year Before Election	0.082*** (0.025)	-0.003 (0.030)
1 Year After Election	-0.019 (0.018)	0.047 (0.052)
2 or More Years After Election	0.048 (0.032)	0.088 (0.089)
Male	0.013 (0.010)	0.012 (0.018)
Age	-0.001 (0.000)	0.005*** (0.001)
Muslim	0.009 (0.018)	0.017 (0.036)
Other Religion	0.040** (0.018)	0.070 (0.055)
Upper Caste	-0.032*** (0.011)	-0.035 (0.022)
SC/ST	0.012 (0.013)	-0.046** (0.022)
Other Caste	0.049 (0.045)	-0.105** (0.053)
Distance from School	0.001 (0.001)	-0.001 (0.001)
N	15,935	1,975
Election Year Mean	0.08	0.01

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Linear models of the probability of a teacher being absent from a private school the day IHDS surveyors surveyed the school, as well as the probability a teacher was absent on official duty as outlined in Equation 1. The dependent variable is either a dummy variable that takes the value of one if the teacher was absent from the school on the day of the survey or, conditional on being absent, whether they were absent on official duty on the day of the survey.

Table A16: Test Scores Improve in Election Years Relative to Non-Election Years

	Government Schools				Private Schools			
	Overall Score	Reading	Math	Writing	Overall Score	Reading	Math	Writing
2 or More Years Before Election	0.068*** (0.010)	0.050*** (0.012)	0.072*** (0.011)	0.082*** (0.014)	0.046*** (0.011)	0.002 (0.013)	0.053*** (0.013)	0.084*** (0.016)
1 Year Before Election	-0.003 (0.009)	-0.008 (0.011)	-0.008 (0.010)	0.010 (0.013)	0.031*** (0.010)	0.029** (0.011)	0.034*** (0.011)	0.028** (0.014)
1 Year After Election	0.113*** (0.011)	0.076*** (0.013)	0.134*** (0.011)	0.129*** (0.015)	0.050*** (0.013)	0.024* (0.014)	0.062*** (0.014)	0.063*** (0.017)
2 or More Years After Election	-0.006 (0.012)	-0.048*** (0.014)	0.033*** (0.013)	0.000 (0.017)	0.008 (0.014)	-0.034** (0.016)	0.011 (0.016)	0.047** (0.019)
Male	0.013* (0.007)	0.010 (0.008)	0.026*** (0.007)	0.004 (0.009)	-0.004 (0.008)	-0.010 (0.009)	0.020** (0.009)	-0.020* (0.011)
Age	0.011*** (0.004)	0.010** (0.004)	0.010** (0.004)	0.014*** (0.005)	-0.005 (0.004)	-0.003 (0.005)	0.000 (0.005)	-0.010* (0.006)
Class	0.077*** (0.003)	0.087*** (0.003)	0.072*** (0.003)	0.071*** (0.004)	0.072*** (0.003)	0.074*** (0.003)	0.077*** (0.003)	0.067*** (0.004)
Local Teacher	0.017** (0.007)	0.011 (0.008)	0.022*** (0.007)	0.015 (0.010)	0.019** (0.008)	0.014 (0.009)	0.020** (0.009)	0.020* (0.011)
Election Year	0.075** (0.033)	0.136*** (0.038)	0.028 (0.034)	0.062 (0.045)	0.445*** (0.036)	0.510*** (0.041)	0.299*** (0.041)	0.516*** (0.051)
Observations	6,365	6,426	6,404	6,379	4,300	4,351	4,335	4,315

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors are in parentheses. I run results for test scores overall, and four reading, math, and writing comprehension separately. The overall scores are a sum of the other three scores and each score is rescaled from 0 to 1. For reading, a child is scored as unable to read, able to read letters, words, paragraphs, or able to read an entire story. In math, a child is scored as unable to recognize a number, whether they can recognize a number, whether they can subtract one-digit numbers, or whether they can divide a two-digit number by a one-digit number. For writing, a child is scored by whether they cannot write, can write a paragraph with two mistakes or fewer, or can write with no mistakes. This presents the results from Figure efigure:testscorecycle.

Table A17: Any Administrative Visits Over the Electoral Cycle in Government Schools Using DISE Data

	Administrative Visits									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2 or More Years Before Election	-0.296 (0.746)	-0.229 (0.760)	-0.315 (0.365)	-0.299 (0.354)	0.020 (0.444)	-0.320 (0.757)	-0.230 (0.770)	-0.352 (0.365)	-0.320 (0.355)	-0.015 (0.446)
1 Year Before Election	-0.292 (0.393)	-0.209 (0.427)	-0.217 (0.343)	-0.189 (0.367)	-0.050 (0.374)	-0.324 (0.404)	-0.221 (0.438)	-0.252 (0.357)	-0.201 (0.382)	-0.078 (0.395)
1 Year After Election	0.579 (0.570)	0.473 (0.563)	0.432 (0.469)	0.270 (0.447)	0.633 (0.527)	0.524 (0.587)	0.461 (0.590)	0.426 (0.486)	0.282 (0.465)	0.620 (0.536)
2 or More Years After Election	-0.145 (0.597)	-0.203 (0.614)	0.097 (0.514)	-0.033 (0.511)	0.003 (0.488)	-0.341 (0.645)	-0.368 (0.664)	0.012 (0.510)	-0.086 (0.513)	-0.087 (0.485)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Teacher Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of Schools	1,168,053	1,168,053	1,168,053	1,168,053	1,250,185	1,163,182	1,163,182	1,163,182	1,163,182	1,239,469
Observations	10,514,932	10,514,932	10,514,932	10,514,932	11,839,336	10,229,464	10,229,464	10,229,464	10,229,464	11,504,370

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the constituency-year level in parentheses. The dependent variable is a dummy variable that takes the value of one if the school reports any visits from a block or district level officer in that year. Each specification includes a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 also include a lagged dependent variable. The election year mean is 7.876. Column 9 corresponds to Panel A of Figure 6.

Full Results Tables for Administrative Visits

In Tables A17 and A18, I present the full set of results for the effects of the electoral cycle on administrative visits (Table A17) and SMC meetings (Table A18). Column 9 in both tables represents the results presented in Panels A and B of Figure 6 respectively.

Table A18: Any SMC Meetings Over the Electoral Cycle in Government Schools Using DISE Data

	SMC Meetings									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2 or More Years Before Election	-0.332 (0.695)	-0.244 (0.549)	-0.258 (0.428)	-0.231 (0.217)	-0.356 (0.308)	-0.347 (0.684)	-0.208 (0.511)	-0.258 (0.425)	-0.213 (0.203)	-0.341 (0.292)
1 Year Before Election	-0.381 (0.764)	-0.228 (0.740)	-0.531 (0.833)	-0.416 (0.772)	-0.440 (0.777)	-0.396 (0.768)	-0.200 (0.751)	-0.542 (0.843)	-0.380 (0.775)	-0.418 (0.777)
1 Year After Election	-0.511 (0.567)	-0.501 (0.510)	-0.330 (0.500)	-0.323 (0.451)	-0.370 (0.498)	-0.514 (0.565)	-0.424 (0.498)	-0.335 (0.507)	-0.297 (0.430)	-0.356 (0.477)
2 or More Years After Election	-0.383 (0.691)	-0.270 (0.738)	-0.518 (0.586)	-0.467 (0.592)	-0.318 (0.660)	-0.580 (0.689)	-0.586 (0.603)	-0.482 (0.605)	-0.631 (0.561)	-0.460 (0.625)
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Lagged DV	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Teacher Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of Schools	1,085,297	1,085,297	1,085,297	1,085,297	1,129,172	1,080,380	1,080,380	1,080,380	1,080,380	1,121,356
Observations	7,197,071	7,197,071	7,197,071	7,197,071	7,497,993	7,036,164	7,036,164	7,036,164	7,036,164	7,321,588

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the constituency-year level in parentheses. The dependent variable is a dummy variable that takes the value of one if the school reports any SMC meetings in that year. Each specification includes a dummy for whether the school is in a rural area. Columns 1-4 and 6-9 also include a lagged dependent variable. The election year mean is 6.062. Column 9 corresponds to Panel B of Figure 6.

Supplementary Materials References

- Adukia, Anjali, Sam Asher and Paul Novosad. 2019. "Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction." *American Economic Journal: Applied Economics* 12(1):348–376.
- Banerjee, Abhijit V. and Esther Duflo. 2006. "Addressing Absence." *Journal of Economic Perspectives* 20(1):117–132.
- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan and F. Halsey Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *Journal of Economic Perspectives* 20(1):91–116.
- Desai, Sonalde, Reena Vanneman and National Council of Applied Economic Research, New Delhi. 2015. "India Human Development Survey-II (IHDS-II), 2011-12. ICPSR36151-v2."
- Gelman, Andrew. 2008. "Scaling Regression Inputs by Dividing by Two Standard Deviations." *Statistics in Medicine* 27:2865–2873.
- Gulzar, Saad and Benjamin J. Pasquale. 2017. "Politicians, Bureaucrats, and Development: Evidence from India." *American Political Science Review* 111(01):162–183.
- Herrera, Yoshiko M. and Devesh Kapur. 2007. "Improving Data Quality: Actors, Incentives, and Capabilities." *Political Analysis* 15(4):365–386.
- Horowitz, Joel L. and Charles F. Manski. 2000. "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data." *Journal of the American Statistical Association* 95(449):77–84.
- Jensenius, Francesca R. and Gilles Verniers. 2017. "Studying Indian Politics with Large-Scale Data: Indian Election Data 1961–Today." *Studies in Indian Politics* 5(2):269–275.
- Jerven, Morten. 2013. *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It*. 1 edition ed. Ithaca: Cornell University Press.

- Kingdon, Geeta Gandhi and Vandana Sipahimalani-Rao. 2010. "Para-Teachers in India: Status and Impact." *Economic and Political weekly* pp. 59–67.
- Kremer, Michael, Karthik Muralidharan, Nazmul Chaudhury, Jeffrey Hammer and F Halsey Rogers. 2005. "Teacher Absence in India: A Snapshot." *Journal of the European Economic Association* 3(2-3):658–667.
- Lindgren, Karl-Oskar, Sven Oskarsson and Christopher T. Dawes. 2016. "Can Political Inequalities Be Educated Away? Evidence from a Large-Scale Reform." *American Journal of Political Science* .
- Martínez, Luis R. 2022. "How Much Should We Trust the Dictator's GDP Growth Estimates?" *Journal of Political Economy* 130(10):2731–2769.
- Muralidharan, Karthik, Jishnu Das, Alaka Holla and Aakash Mohpal. 2017. "The Fiscal Cost of Weak Governance: Evidence from Teacher Absence in India." *Journal of Public Economics* 145:116–135.