ONLINE APPENDIX to "Teachers, Electoral Cycles and Learning in India" by Sonja Fagernäs and Panu Pelkonen

1. Cleaning of the DISE school data

The sample of interest covers lower primary schools governed by the Department of Education, Tribal/Social Welfare Department or another local body. The raw database for 2005-2011 includes roughly 6 million school observations. The size of the sample used in the analysis is smaller for three reasons. Firstly, we have excluded schools for which there is some doubt about the robustness of the school code across time. This procedure excludes 8.7% of observations. On average, the excluded schools are slightly smaller than others (2.61 versus 2.77 teachers per school). In practice, we have excluded schools that go through a 'substantial' name change as defined by a simple algorithm, while keeping the same school code. This can lead to the exclusion of schools, which have genuinely changed name, but since the analysis uses school fixed effects throughout, we strive to ensure that school panels are genuine. Secondly, from the remaining sample, 8.9% of observations are deemed to be outlier observations with respect to some key variable of interest. Outlier status is first assigned to observations with undoubtedly unrealistic values. With uncertain cases, the top (and/or bottom) 0.5% of the values are regarded as outliers. Outliers, on average, relate to larger schools than others (3.02 teachers). Finally, 2.2% of the remaining observations include missing values for some variables of interest. The initial and final samples in terms of a few characteristics are showed in Table X1 below.

Table X1

Sample selection in the DISE school-level data

	Raw data		Change Sch. Code		Outliers		Final		Regression sample	
Observations	6059856		524356		494636		5040864		4929221	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Year	2008.1	2.0	2008.0	2.0	2007.9	2.3	2008.1	2.0	2008.1	2.0
# of Teachers	2.77	1.95	2.61	1.80	3.02	3.15	2.76	1.81	2.76	1.80
Urban school	0.07	0.25	0.05	0.22	0.10	0.30	0.07	0.25	0.07	0.25

2. Test for caste favoritism

In this appendix we test whether caste-based transfers are more common in the post-election phase, in particular if the caste of the local incoming Member of the Legislative Assembly (MLAs) matches the caste of the teacher. These data come from the Election Commission of India. A limitation of this analysis is that a large share of variation in politicians' caste is due to the delimitation exercise of 2008 which affected the proportions of seats reserved to particular castes.

Table X2 shows the results of the model in equation (1), where the post-election dummy variable is interacted with the proportion of the district's MLAs, who report being either scheduled caste (SC) or scheduled tribe (ST), the reference group being the general caste. The models are estimated for three sub-samples of teachers: general caste, scheduled caste (SC) and scheduled tribe (ST).[i] There are two observations. Firstly, a larger share of SC politicians is associated with a higher frequency of transfers for SC teachers, although this relationship is not statistically significant. The post election peak in transfers for Scheduled Tribe (ST) teachers is to some extent attributable to changes in the share of Scheduled Caste MLAs because the post election dummy on its own becomes statistically insignificant, whereas its interaction with the MLA share of tribal politicians is significant (Table X2, column 3). However, ST teachers represent a relatively small fraction of all teachers (12.4%). Overall, we can conclude that caste does appear to play some role in teacher transfers, but rule out the possibility that caste-based politics would be a key driver of post-election transfers.

Table X2

Caste of politicians and transfers, OLS estimates

	[1]	[2]	[3]
Dependent: Transfer	Sample:		
	General	SC	ST
Electoral cycle main effects:			
[4]	0.0632	0.0476	0.0484
	[0.0567]	[0.0587]	[0.0471]
[5] 'Election year'	-0.011	-0.0287	-0.0333
	[0.027]	[0.0244]	[0.0201]
[1] 'Post-Election'	0.0744*	0.0647	0.0385
	[0.0325]	[0.0387]	[0.0276]
[2]	0.0173	0.0273	0.0276
	[0.013]	[0.019]	[0.0156]
Politician's caste main effects:			
% MLAs SC	0.0073	0.112	0.0293
	[0.0493]	[0.0898]	[0.0549]
% MLAs ST	-0.0005	-0.0421	-0.116*
	[0.0572]	[0.0693]	[0.0518]
Interactions:			
[1] 'Post-Election' and % MLAs SC	-0.0377	-0.0414	0.038
	[0.0648]	[0.0612]	[0.0573]
[1] 'Post-Election' and % MLAs ST	0.0086	-0.0009	0.0789*
	[0.0572]	[0.0458]	[0.03]
Observations	4889095	992871	715477
R-squared	0.0120	0.0280	0.0254

Notes: All estimates are school level fixed effects estimates. MLA stands for Member of Legislative Assembly, elected from each constituency. The percentage shares of MLAs of each caste are district level averages, and the mean of '% MLAs SC' across all teachers is 0.1477, and the mean of '% MLAs ST' is 0.1141. The reference category is General caste teachers, which in this case include Other backward classes (OBC) given that the caste of politicians is reported in this way in the election data. Standard errors are clustered at the state level. (+, *, **) refer to statistical significance at 10%, 5% and 1% levels, respectively.

3. Effects of the election cycle on learning: An alternative specification

This specification compares the skills of 4^{th} grade pupils, depending on which election cycle (out of the five possible) they have missed during their schooling. Essentially, models (2) and (X1) below will convey the same message, but are framed differently.

$$\text{(X1)} \quad zscore_{itd} = A_i + Female_i + \Lambda_t + \Omega_d + \beta Miss_y + u_{it}, \qquad t \in [2005,2011], \qquad y \in [1,5].$$

The sample is restricted to grade 4 pupils. The variable of interest is the $Miss_y$ dummy. Assuming that no grade repetition took place, this dummy variable indicates whether a pupil was not attending school in the school year that begins over a certain phase of the election cycle (y). For example, $Miss_1$ refers to missing the school year that begins in the post-election year. Again, the actual election phases are instrumented with the intended election phase.

Table X3 shows the results for model specification X1. We estimate separate models for pupils in government (public) and private schools and separate models for Reading and Mathematics. The rows refer to a missed phase of the election cycle. It is important to note that each row-column cell represents a coefficient from a separate estimation.

The results in column 1 show that not being in school in the school year beginning during the post-election year is beneficial for Reading outcomes. Avoiding this year increases Reading scores on average by 0.084 standard deviations for those currently in grade 4. With respect to the treatments in Table 6, the result implies that fourth graders would be best off if they experience 'Treatment 2'; they begin their schooling in phase 2, and enter fourth grade in phase 5 (the election year).

Table X3

Learning outcomes of grade 4 pupils, IV models

	[1]	[2]	[3]	[4]
	Government		Private	
	Reading	Maths	Reading	Maths
Treatment / Election phase missed by grade 4				
T2 / Miss school year beginning in the post-election year	0.0843*	0.115**	0.0133	0.0481+
	[0.0362]	[0.0409]	[0.0221]	[0.0273]
T3 / ..phase 2	-0.013	-0.0131	-0.0017	-0.0114
	[0.0263]	[0.0278]	[0.0139]	[0.0162]
T4 / ..phase 3	-0.0719**	-0.0703**	-0.0188	-0.032
	[0.026]	[0.0267]	[0.024]	[0.0287]
T5 / ..phase 4	0.0056	-0.0191	-0.0108	-0.0047
	[0.025]	[0.022]	[0.017]	[0.0171]
T1 / Miss school year beginning in the election year	0.0064	0.0004	0.0164	-0.002
	[0.0254]	[0.0302]	[0.0191]	[0.0199]
Observations (pupils)	317762	316104	83699	83261
Number of districts	562	562	562	562

Notes: Each row-column cell represents the coefficient from a separate regression model (equation X1). Treatments T1-T5 are explained in Table 6. Each model includes district fixed effects, survey year controls, age and gender controls. Standard errors are clustered at the state level. (+, *, **) refer to statistical significance at 10%, 5% and 1% levels.

Those who miss the election phase 3; three years after elections, appear to do worse than others. Referring back to Table 6, we can see that fourth graders who miss phase 3 are in 'Treatment 4'; they experience elections in the same year that they begin grade 2, and reach grade 4 two years after the elections. Out of all the treatments linked with experiencing the post-election year in Table 6 (T1, T3, T4 and T5), those in Treatment 4 are in the 'worst' position. They are tested in phase 2, having just experienced the school year that begins in the post-election year (phase 1). This is the year associated with an increase in teacher transfers. Those in treatments T1 and T5, who experienced the election phase with an increase in transfers earlier during their schooling, have better skills. This result suggests that there is a degree of decay in the effect of the post-election shock.

Column 2 shows the results for similar models for Mathematics. The results are broadly similar to those for Reading, but the estimated coefficients are somewhat larger. A fourth grade pupil scores 0.115 standard deviations higher, if she has missed the school year starting in the post-election year during her primary schooling.

In columns 3 and 4 of Table X3, the same estimates are repeated for children who attend private schools. If the effects on learning are attributed to the government school system, we would not expect to see effects on learning for private school pupils. For Reading, there is no evidence that the phases of the election cycle would make a difference for learning; all coefficients in the Table are statistically insignificant. The same is true for Mathematics in column 4, with phase 1 being only marginally statistically significant. Overall, the results for private schools suggest that the findings in columns 1 and 2 are not mere statistical artifacts, but represent variations in the quality of how government schools are run across the electoral cycle.

---

i General caste teachers include 'other backward classes', as this is how the caste of politicians is reported.