

Teacher unions, teacher pay and student performance in India: A pupil fixed effects approach

Geeta Kingdon^a
Institute of Education, London University

Francis Teal^b
Department of Economics, University of Oxford

April 2009

Abstract

This paper examines the relationship between teacher unionization, student achievement and teachers' pay using a cross-section of data from private schools in India. We use differences in student mark across subjects to identify within-pupil variation in achievement and find that union membership of the teacher appears to strongly reduce pupil achievement. We find no evidence this could be due to the unobservables not controlled for by this procedure. A school fixed effects equation of teacher pay shows that union membership raises pay and in this case too we find that remaining unobservables are unlikely to explain this outcome. We thus have in this data evidence that unions raise cost and reduce student achievement. We discuss the policy implications of the findings and show that teacher credentials do not necessarily improve teacher performance.

JEL classification: I21

Keywords: teacher unions, teacher salaries, student achievement, productivity, India

^aCorresponding author: Geeta Kingdon, Institute of Education, London University, 20 Bedford Way, London WC1H 0AL. Telephone: 44-20-3073-8316; Fax: 44-20-3073-8316.
Email: g.kingdon@ioe.ac.uk

^bCoauthor: Francis Teal, Department of Economics, University of Oxford, Manor Road, Oxford, OX1 3UQ. Telephone: 44-1865-271077; Fax: 44-1865-281447.
Email: francis.teal@economics.ox.ac.uk

Acknowledgments: This paper has benefited from the comments of Joshua Angrist, Martin Browning, Paul Collier, Esther Duflo, Paul Glewwe, Eric Hanushek, Helen Ladd, Stephen Machin, Karthik Muralidharan, Lant Pritchett, Måns Söderbom and Alessandro Tarozzi, as well as the comments of participants at the NEUDC, Royal Economic Society and CESifo/PEPG conferences. We are greatly indebted to Todd Elder for sharing STATA routines with us. Any errors are ours. We would like to thank the Council for Indian Secondary Certificate Examinations for enabling data collection from affiliated schools. Funding from ESRC Global Poverty Research Group and the DFID RECOUP consortium on 'Outcomes of Education' is gratefully acknowledged.

1 Introduction

Development policy is centrally concerned with the cost of supplying education and its outcomes. In this paper we have data that enables us to address an important element of such costs in the role of teacher unions. We can link unionization to measures of student achievement which, recent work suggests, is much more closely correlated with economic growth than measures of educational outcomes based on years of schooling (Hanushek and Woessmann, 2008). If learning outcomes are critical for economic success then understanding how those outcomes can be improved at lower cost is an important part of the development policy agenda.

Under different theoretical models, teacher unionization can lead to higher or lower student achievement. Conceptually there are two major reasons why teachers may become union members (Hoxby, 1996). The first is that teachers maximise the same objective function as parents, namely student achievement, but have superior information about the correct input mixes. Union membership provides teachers with a collective voice to implement these input mixes. This may include for instance asking for lower class sizes or higher salaries which help to attract and retain superior teachers. The second potential reason for teachers joining a union is that they have a different objective function than parents or school management, possibly one in which school policies that directly affect them, such as teacher salaries, receive greater weight than policies that only indirectly affect them, i.e. membership of a rent-seeking teachers' union¹.

¹ Loveless (2000), Ballou (2001) and Ballou and Podgursky (2000) argue that unions retard school reform. On the other hand, Henderson, Urban and Wolman (2004) argue that in recent times the public “has been exposed to a barrage of pejorative views” about teacher unions which “by and large condemn unions as obstacles to efficiency and quality” but do not recognise what unions have contributed towards building a quality public education system. Stone (2000) and

A rent-seeking union may block reform of incentives to improve instruction, e.g. by tying salaries to seniority rather than to performance and by protecting ineffective teachers from dismissal. Under rent seeking, unions may also lower achievement if their pursuit of higher salaries diverts resources away from other school inputs that raise achievement and if teacher union strikes disrupt teaching. Finally, since teachers interact with other inputs in order to produce education, rent seeking unions could lower the efficiency of the other inputs, such that more money for schools may not matter (Burtless, 2000; Lemke, 2004; Hoxby, 1996).

The sign of the relationship between teacher union membership and student achievement is thus an empirical question. This paper addresses this empirical question asking how teacher unions affect student achievement and teacher's pay. We pose these questions with a unique dataset on private secondary schools from India which permits pupil fixed effects estimation of the achievement function, enabling us to address what are arguably the most important sources of endogeneity bias.

As has been widely recognised in the achievement production function literature, imputing causality from any cross-section correlation between teacher attributes and student outcomes is problematic as students may match to schools and teachers endogenously. While randomized experiments provide a good solution to the problem of endogeneity in general (see Duflo and Hanna, 2005, for one educational application in India), they cannot be used here since union membership cannot be randomly allocated. Although quasi-experimental approaches such as propensity score matching methods have been used to evaluate education program impacts

Eberts (2007) summarize the US literature on the effect of teacher unions on student achievement.

(e.g. Machin and McNally, 2004), they require the assumption that matching based on pre-treatment observables adequately captures all relevant characteristics of treated units. Valid instrumental variables are relatively difficult to find though some studies have convincingly tested impact effects in education using the IV approach (e.g. Angrist and Lavy, 1999). The IV approach is often infeasible in developing countries because even where official (exogenous) rules, e.g. about maximum class-size or school start age etc., exist in law, they are rarely adhered to in practice. The panel data approach requires longitudinal data and this is scarce.

We examine the effect of teacher union membership on pupil learning using the standard cross-section achievement production function but allowing for pupil fixed effects. This is possible because the data used here provide each student's marks in five different subjects (English, second language, history/geography, math, science). We link the average characteristics of the teachers who teach that subject within the school to the achievement of the student. This approach allows us to control for all student and family unobservables and examine whether the union membership status of different subject teachers in a school is related to a student's marks across those subjects. In other words, we estimate a within-pupil across-subject equation of the achievement production function rather than a within-pupil across-time one. After investigating the union effect on pupil learning, we also examine the teacher pay schedule to ask whether unionized teachers are paid more once we control for their observable skills and the unobservable characteristics of the school.²

² While much has been written on the effect of teacher unions on student achievement in the US, both by economists and non-economists (Eberts and Stone, 1987; Betts, 1996; Hoxby, 1996; Ballou and Podgursky, 2000; Lemke, 2004; Eberts, 2007; Loveless, 2000; Carini, 2002;

While the above approach can control for many possible sources of endogeneity, it cannot control for all. In particular, it cannot control for the possible correlation between teachers' unobserved characteristics and their union membership status. To see if the remaining unobservables could be explaining our results, we adopt the procedure developed by Altonji, Elder and Taber (2005) to assess how important these unobservables would need to be to explain our result. How important unobservables will be depends on what we can observe so in the next section we set out the data and model explicitly the potential role of unobservables in explaining our results. Section 3 presents estimates of how union membership affects achievement and section 4 the results for the effect of union membership on teacher pay. A final section concludes.

2 Observables and Unobservables for Student Achievement

The data for this study come from a sample of 186 schools affiliated to the Council for Indian Secondary Certificate Examinations (CISCE) which is an English Medium exam board. The schools were chosen by a stratified random sampling procedure within 16 major Indian states (the strata). The sampling procedure is explained in Kingdon (2006). Postal questionnaires sent by the Exam Board were filled by all students of grade 10 in the sample schools, and by the teachers that teach them, as well as by the school Principal. Grade 10 students are aged approximately 16 years old and the grade 10 board examination in CISCE schools is equivalent to the High School board examination in other Indian exam boards, such as the state examination boards or the Central Board of Secondary Education. The overwhelming proportion of CISCE affiliated schools are private unaided schools, i.e. run without state aid (95%), 3.2% are aided schools (mainly in West

Henderson, Urban and Wolman, 2004), we are not aware of any research on this topic in developing countries in general and in India in particular.

Bengal), and only 1.6% are government or local body schools. Thus the sample represents mainly English Medium private secondary schools in India which are generally the more privileged fee-charging secondary schools.

The student questionnaire captured information on a wide range of the child's personal characteristics including age, gender, health, disability, and time-use, as well as detailed family characteristics such as household demographics, asset/wealth ownership, parental education and occupation. Pupils were also given a test of ability/IQ based on 36-items in Sets A, C and D of the Ravens Progressive Matrices test. The teacher questionnaire collected information on a range of teacher characteristics and the school questionnaire elicited data on student and teacher numbers, school facilities and resources, length of the instructional program, school fees, and management and teacher motivation aspects.

Examination results data were subsequently provided by the exam board and matched to students using a unique pupil identifier code. Students of grade 10 in the CISCE exam board have 5 compulsory and one optional subject. The optional subject is chosen from among a large number of subject choices and thus varies a great deal between students. We wish to use exam marks of students in the 5 compulsory subjects: English, Second language, History/Geography, Math and Science³. Table 1 shows mean mark by subject. All subjects are marked out of 100 so

³ As the ICSE exam board is an English Medium exam board, all subjects in affiliated schools are taught in English. The local state language is considered the 'Second language' and it is typically a child's mother tongue. While History and Geography are tested separately, their marks are pooled by the examination board and provided together, i.e. they are treated as one subject. Similarly, while physics, chemistry and biology are tested separately, their marks are lumped together. When matching student marks in a subject to the characteristics of the teacher that

the given marks may be interpreted as percentages. The second language mark has the lowest dispersion and the highest mean across the subjects while the maths mark has the largest distribution, with a SD of 18.4 around a mean of 69 marks. In order to render the marks in the different subjects comparable, we standardize the mark in each subject by the national mean mark in the subject, i.e. we use the z-score of achievement as our dependent variable. The z-score is a student's mark in a subject less the national mean mark in that subject, divided by the standard deviation of mark in the subject. Thus, by construction, mean z score in any given subject is 0 and its standard deviation is 1. Appendix Table 1 defines variables and shows descriptive statistics. The OLS regressions of standardized achievement by subject are presented in Appendix Table 2, as an extension of the descriptive statistics, but we do not discuss these.

As a result of this data collection we have marks across subjects for the same student within a school as well as an extensive range of controls for both the student and the teacher. We have taken the average of characteristics of the subject teachers within a school so we have variation in these characteristics for a student across subjects. It is this cross-subject variation we intend to exploit in establishing if there is an effect from unions onto student achievement.

$$(1) \quad A_{ijk} = \alpha + \beta X_{ij} + \gamma TC_{jk} + \theta U_{jk} + \delta S_j + (\mu_i + \varepsilon_j + \eta_k) + v_{ijk}$$

A_{ijk} is achievement of the i^{th} student in the j^{th} school in the k^{th} subject, X is a vector of characteristics of the i^{th} student in the j^{th} school, TC_{jk} is all teacher characteristics that are

taught them that subject, we have taken the *average* characteristics of the teachers that taught them that subject. Thus, if history and geography were taught to grade 10 by different teachers, we have taken the average characteristics of the history and geography teachers as the relevant characteristics to match with a student's history-geography subject-row.

hypothesized to be independent of the unionization term, U_{jk} in the j^{th} school for subject k . S_j a vector of characteristics of the j^{th} school. μ_i , ε_j and η_k represent the unobserved characteristics of the student, the school and the subject respectively. The remaining unobservables are in the error term, v_{ijk} , and it is their potential correlation with the observables, in particular the union variable, that is the key to being able to identify a causal effect from unionization onto achievement.

The teacher characteristics that we are going to use in our achievement production function are:

$$(2) \quad TC_{jk} = \{High_{jk}, Train_{jk}, Divison_{jk}, Christian_{jk}, Female_{jk}, Age_{jk}\}$$

where $High_{jk}$ is whether the teacher has a higher qualification meaning an MA or PhD qualification, $Train_{jk}$ is whether the teacher has pre-service teacher training, $Division_{jk}$ is whether the teacher had obtained first division in his/her higher secondary exam (a proxy for teacher's own cognitive skills), $Christian_{jk}$ is whether the teacher is a Christian, $Female_{jk}$ indicates their gender and Age_{jk} their age. All these characteristics can be viewed as pre-determined for the union variable. We also have information on other teacher characteristics which arguably may not be. We denote these as TC'_{jk} :

$$(3) \quad TC'_{jk} = \{Leave_{jk}, Experience_{jk}, Tenure_{jk}, Permanent_{jk}\}$$

where $Leave_{jk}$ is the days of sick leave last year, $Experience_{jk}$ is the total years of teaching experience, $Tenure_{jk}$ is years of tenure in current school and $Permanent_{jk}$ is whether the teacher has a permanent position. All the arguments of TC'_{jk} may be at least in part determined by the

union status of the teacher so including them in the student achievement function would bias down the effect of unionization. To test whether our results are sensitive to their inclusion we report regressions which include the arguments of both TC_{jk} and TC'_{jk} . We apply a similar argument to our equation for Teacher's pay

$$(4) \quad P_{jk} = \phi\{TC_{jk}\} + \varphi U_{jk} + u_{jk}$$

where P_{jk} is the pay of the teacher. If we wish to identify the effect of unionization on pay it seems necessary to exclude variables from the regression which are in part determined by unionization. Again we test if our results are sensitive to the exclusion of the variables in the TC'_{jk} vector.

Equations (1) and (4) are our equations of interest and for both we need to allow for the possibility that there remain unobservables in the error terms correlated with the union variable. Before considering this problem we outline how across-subject differencing has two important methodological advantages over across-time differencing.

Firstly, the across-subject approach does not suffer from the problem of non-random attrition of teachers and students over time that occurs in panel data. For instance, in their panel study relating student achievement to teacher characteristics using North Carolina data, Clotfelter, Ladd and Vigdor (2006) highlight the difficulty of determining whether a higher coefficient on teacher experience reflects a teacher's improved effectiveness with experience or the differentially higher attrition of the less effective teachers⁴, and Rivkin et al (2005) also

⁴ They say that due to the technical difficulty of including both pupil and teacher fixed effects in one equation they attempt to address the problem by using the sub-sample of teachers who remain teachers for three or more years.

address non-random attrition. Across-subject estimation obviates this problem since estimation is within pupil at one point in time. While the potential for endogenous selection into the ‘surviving’ teachers’ group is the same in both approaches, the across-time technique relies on change in teacher over time (over which non-random attrition can take place) as part of the estimation strategy, while the across-subject technique does not.

The second methodological advantage is that the across-subject approach provides a means to circumvent the potential problem of non-random matching of students to particular teachers within the school on the basis of their unobserved characteristics – whether it be brighter students matching to abler teachers or school policy deliberately matching slower students to abler teachers. Across-subject estimation bypasses the problem either by averaging the characteristics of all teachers by grade and subject within the school, or by restricting the sample to schools where any given subject is taught to the student’s grade by only *one* teacher within the school. Either way, the student is by construction matched to a single set of teacher characteristics in each subject within the school.

It however remains the case that subject-varying school unobservables remain in the error term and may be correlated with the Union variable. Moreover, it is required that teachers’ unobserved characteristics be unrelated to the union variable. Since such omitted characteristics may be correlated both with Unionization status and with student achievement we cannot say that pupil fixed effects estimation of achievement – even with no subject-specific student and school unobserved heterogeneity – permits us to interpret the effects of Unionization as causal.

We propose to address this issue by borrowing from the procedures set out in Altonji, Elder and Taber (2005) to assess the potential size of any bias due to the unobservables in the equation (an appendix summarises their algebra which we use below) . In their paper they show

that the notion that “selection on observables is the same as selection on unobservables” is equivalent to a condition that in our model (where we drop subscripts for clarity) is:

$$(5) \quad \frac{Cov(v, U)}{Var(v)} = \frac{Cov(\gamma TC, U)}{Var(\gamma TC)}$$

The asymptotic bias from OLS is $\frac{Cov(v, \tilde{U})}{Var(\tilde{U})}$, which is equivalent to $\frac{Cov(v, U)}{Var(\tilde{U})}$ if v and TC are

orthogonal (tildes over a variable denote the residuals from a regression of that variable on TC).

We can estimate this bias by noting that:

$$(6) \quad \frac{Cov(v, U)}{Var(\tilde{U})} = \frac{Cov(v, U)}{Cov(\gamma TC, U)} \frac{Var(\gamma TC)}{Var(v)} \frac{Cov(\gamma TC, U)}{Var(\gamma TC)} \frac{Var(v)}{Var(\tilde{U})}$$

$$(7) \quad \frac{Cov(v, U)}{Var(\tilde{U})} = \frac{Cov(\gamma TC, U)}{Var(\gamma TC)} \frac{Var(v)}{Var(\tilde{U})}$$

where (7) follows from (6) using (5).

In reporting our regression results we will also show the results from applying this method. We do this by reporting the estimate of bias from equation (7). As we will show, for the school and student fixed effects regressions, the sign of the union effect is negative while the estimate of the potential bias induced by a correlation between the unobservables and the Union variable is positive. In this case the evidence points to the school and student fixed effects estimates of the union membership effect as being underestimates of the true negative impact. Where the sign of the effect and the bias are the same we report their ratio as this is open to the interpretation from Altonji, Elder and Taber (2005) that it measures the size of the shift in the distribution of the unobservables necessary to explain away the implied effect from the Union treatment. We turn now to assessing the role of both observables and unobservables in determining student achievement.

3 The Achievement Production Function

The main results of the paper for student achievement are presented in Tables 2 and 3, which pool all five subjects. In Table 2 the only teacher characteristic included is teacher union membership; in the first three columns of Table 3 we include those teacher covariates in the vector TC_{jk} set out above, and in the fourth column we include the teacher covariates in the vector TC'_{jk} .

The first column of Table 2 is an OLS achievement production function with state dummy variables. The second and third columns show school and pupil fixed effects results. The school fixed effects equation includes the full set of pupil characteristics, including a measure of pupil ability (as measured by the ravens progressive matrices test). The achievement production function changes dramatically when moving from the OLS (Column 1) to the within school analysis of Column (2). Exploiting within pupil variation in Column (3) does not change the union effect significantly. While higher scoring schools are more likely to have unionized teachers, within a school the students of unionized teachers have sharply lower achievement levels.

Table 3 includes the teacher covariates from TC and shows a larger negative effect of unionization on achievement. These teacher characteristics do have a highly significant effect on student achievement and are positively correlated with the Union variable. Thus controlling for these observable aspects of the teacher increases the adverse effect of unionization on achievement. These results suggest that the achievement of a student in a subject that is taught by a unionized teacher is 0.23 standard deviations lower than his/her achievement in a subject taught by a non-unionized teacher.

The results from the procedures of Altonji, Elder and Taber (2005) are shown in the bottom part of the Table. There is evidence of a potentially substantial bias in the OLS results. However once we move to the school and student fixed effects results, we find evidence of a positive correlation between the unobservables in the achievement equation and the Union variable, implying that the negative effect of Unionization on achievement is underestimated⁵.

The last column of Table 3 tests whether our results are sensitive to the inclusion of experience, tenure, permanent status and days of leave taken (arguments in equation (3)), which may be at least in part determined by the union status of the teacher. The inclusion of these variables does not affect the coefficient on union membership.

Table 4 checks the robustness of the union membership effect to differencing across various three- and four-subject combinations, instead of differencing across all five compulsory subjects, as in our main results of Table 3. Though the magnitude of the coefficient on union membership varies by sub-sample, it is negative and statistically significant in virtually all sub-samples and the average union effect is -0.252 in the three-subject combinations and -0.232 in the four-subject combinations, i.e. very close to the union effect with five-subjects in Table 3⁶. The bottom part of Table 4 tests whether the size of the union effect varies by political leaning of

⁵ See the Appendix for the equation of union membership on the *TC* vector of variables.

⁶ An equation of teacher union membership shows that science teachers are more likely to join a teacher union than other subjects' teachers. The results excluding science achievement (in 4 subject and 3-subject combinations) show that our main results in the paper – based on differencing across 5 subjects – are not driven by science teachers' unobserved characteristics. The union effect was negative and statistically highly significant in pupil fixed effects achievement equations fitted separately for students of different gender, religion and ability (not reported). Unions reduce achievement across the whole range of student ability (lowest quartile, middle 50% and the top quartile of achievers), though the coefficient was much smaller for low achievers than for middle and high achievers. This contrasts somewhat with syntheses of the US literature which find that unions modestly improve the achievement level of students who are middle-range performers but reduce that of students who are low and high performers (Eberts, 2007; Stone, 2000; Carini, 2002; Betts, 1996).

the state and by degree of school competition. It might be expected, a priori, that the negative effect of teacher union-membership on student achievement will be lower in the ‘Left’ leaning states (Kerala and West Bengal, which have had communist parties in power in much of the post-Independence period) than in other Indian states because, in their stronger pro-labour environment, teacher demands are satisfied more quickly without unionized teachers losing too many teaching days through strike action. This is indeed what we find in Table 4, and the difference in the union effect between the ‘left’ and other states is statistically significant. Secondly, it might be expected that the union effect will be largest in schools that face high competition (i.e. that have plenty of other nearby schools) since unionized teachers are likely to be less afraid of shirking or rent-seeking in high school-concentration areas, where the teacher labor market is tighter i.e. the prospect of getting other teaching jobs is greater. School competition is measured by the number of schools situated within a 5 km radius of the school. Fewer than 4 schools within this radius corresponds to ‘low’ and greater than 12 corresponds to ‘high’ competition. The negative union effect is greatest in high competition schools and it is statistically significantly higher there than in middle and low competition schools.

As a final robustness check we present in Table 5 the pupil fixed effects achievement equation using the sub-sample of schools that had any unionized teachers, which increases the proportion of the sample of children taught by a unionized teacher to 37 per cent. The coefficient on the union membership variable is very similar to that in Table 3. Table 5 also shows that teachers’ possession of MA qualifications and teachers’ cognitive skills (proxied by first division in board exam) raise student achievement only in the non-unionized schools, i.e. the productivity of teacher characteristics is greater in non-union schools, a result similar to that in Hoxby (1996) where the beneficial effects of school inputs (such as lower class-size), and of teacher salary, on

pupil achievement are significantly lower in unionized than in non-unionized US school districts⁷.

These findings are important for policy. In this sample of English-Medium private secondary schools, the conventional measures of teacher quality – whether teacher has MA qualifications, pre-service teacher training and a first division in her/his own Higher Secondary exam (a proxy for teacher’s cognitive skills) – do have payoffs in terms of higher student achievement. For instance, teacher credentials such as having MA, training and a first division in Higher Secondary exam are each associated with approximately 0.09 – 0.10 SD higher pupil achievement. However, as is seen in Table 5, these pay-offs are generally not available in unionized schools. Thus better credentials do not necessarily improve teacher performance.

4 Teacher pay schedule

How do unions impact on pay? We attempt to answer this question by examining the teacher pay schedule. All employers with more than 10 workers in India have to *de jure* abide by minimum wage laws and ‘recognized’ private schools are required to pay teachers salaries on a par with the government teacher salary scales. *De facto*, many private schools pay teachers significantly less than the government prescribed minima. For instance, at the middle school level, Kingdon (1996) found that private teachers’ mean salary was only 60% of the public teachers’ mean salary. Kingdon and Teal (2007) confirm that salary structures in public and private schools in India are very different.

The survey for this study collected information on all teachers that taught grade 10 students in sample schools. This yielded a sample of 2103 teachers of whom 1731 are teachers of

⁷ Interacting teacher qualifications with union membership in the full sample of all schools (not reported) showed that the positive effect of MA qualifications exists only for non-union teachers. Thus the positive effect of MA qualifications, on achievement, is absent for unionized teachers.

the five compulsory subjects taught at grade 10 level. The first column of Table 6 presents an OLS equation of log of teacher pay but our data also have within-school variation in teacher pay, which is used to estimate a school fixed effects equation in column 2. The coefficients on some of the variables – and especially on the union membership variable – change when moving from across to within school estimation. Since teachers may sort into schools on the basis of their unobserved characteristics, we rely more on the school fixed effects regression in the second column.

The variable of interest is union membership. It is seen that the coefficient on union membership is positive and statistically significant at the 5% level. Within a school, a union member earns a wage premium of 14.9% over non-union members. This regression does not control for the teacher characteristics identified above as potentially determined by unionization. If we include these, the union effect falls to 9.5% as is shown in Table 7. This is similar to the teacher union wage effect of 7.6% estimated by Lemke (2004) for Pennsylvania using data from school districts after accounting for aspects of teacher quality. It is also virtually identical to the union wage effect of 8-10% associated with other US public sector unions.

Applying a similar test for the importance of unobservables to those already reported for the achievement equations, we do now find a positive potential bias on the positive union coefficient, but it is small compared with the size of the union pay effect. The ratio of effect to bias is 1.65 which implies that the role of unobservables that determine pay would have to be more than 1.65 times the role of observables for the entire union pay effect to be explained away by the unobservables, which seems unlikely given the rich set of observables we include, including the school fixed effects. This suggests that part of the union pay effect is real.

The above findings suggest two questions: Why would schools hire union teachers given they cost more and produce less? Also, if there is a 9% union wage premium, then why do not all

teachers join a union? On the first, whether or not a teacher is a member of a union is unlikely be observed at the time when they are hired. The process of hiring a teacher in a private school is highly specific to the individual school and there are no mandated rules about whether, at the time of hiring, a teacher has to declare any union memberships. Even if there were such rules, compliance with these would be patchy as with so many other rules in the education sector in India⁸. Applicants to private schools are unlikely to admit to being union members at the time of recruitment because of the fear of being identified as potential ‘trouble makers’.

We have investigated what kinds of teachers join unions by estimating an equation of teacher union membership (not reported), with teachers’ age, gender, religion, qualifications, training and cognitive skills (proxied by teacher’s ‘division’ or grade in her/his own Higher Secondary exams). This showed that persons with pre-service training (i.e. those who prepared for teaching as a career) are 10 percentage points more likely to join a union than those who do not acquire training. It is also the case that older teachers are significantly more likely to join a union than younger teachers. Both these findings are consistent with teachers who join a union being more career oriented than those who do not and thus more interested in union benefits both in terms of pay and as a protection from dismissal.

The second question raised by the results is why, given the union premium, all teachers do not join a union. Clearly union premiums exist in many industries and incentives to join unions differ greatly across workers in those labour markets. In our sample of English Medium private schools, well-heeled English-educated individuals enter teaching and 71% of them are women; as such, they may not wish to be associated with labour movements for a relatively modest union wage

⁸ For example, as is well known, the mandated 40:1 pupil teacher ratio is routinely violated when deploying or transferring teachers even in the government school sector. Also, a high proportion of government ‘recognised’ private schools do not fulfil the conditions required for gaining ‘government recognition’.

premium. If the management of schools is hostile to unionised teachers, and our results suggest this is possible, then there may be costs of membership which we cannot observe resulting from joining a union.

5 Conclusions

Using a methodology for identification that allows for controls at the level of the pupil, we find that the achievement level of a student in a subject that is taught by a unionized teacher is about a quarter of a standard deviation lower than his/her achievement in a subject that is taught by a non-unionized teacher. A school fixed effects equation of teacher pay shows that union membership is rewarded with substantially higher pay. In other words, union membership of teachers reduces student achievement and raises salary costs in private schools in India.

While we controlled for what are commonly regarded as the most important sources of omitted variable bias in the student achievement equation – namely the non-random matching of students to schools and teachers on the basis of their unobserved characteristics, *teacher* unobservables are still in the error term and could be related with observed teacher characteristics. We sought to deal with this potential source of bias by using the method of Altonji, Elder and Taber (2005) and find, once we allow for school or student fixed effects, that the negative union effect on student achievement is, if anything, underestimated. However it remains a possibility that certain teacher types join unions. If this is the mechanism by which the adverse effect of unions operates then it is not an omitted variable in the sense Altonji, Elder and Taber (2005) are investigating, but more that poor teachers reveal their type by joining a union.

If unions do reduce student performance then policy needs to focus on how such effects can be mitigated. One policy, which has been argued to raise student achievement in India, is linking teacher pay with student achievement (Muralidharan and Sundararaman, 2008).

References

- Altonji, J., T. Elder and C. Taber (2005) "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113 (1): 151-184, February.
- Angrist, J. and V. Lavy (1999) "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics*, 114(2): 533-75.
- Ballou, D. (2001). Pay for performance in public and private schools, *Economics of Education Review*, 20, pp. 51-61.
- Ballou, D. and M. Podgursky (2000) "Teacher Unions and Education Reform: Gaining Control of Professional Licensing and Advancement." in Tom Loveless (ed.) *Conflicting Missions: Teacher Unions and Education Reform*. Washington DC: The Brookings Institution.
- Betts, Julian (1996) "Is there a link between school inputs and earnings? Fresh scrutiny of an old literature" in Burtless, G. (ed.) "*Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*", Brookings Institution Press.
- Burtless, G. (2000) "*Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*", Brookings Institution Press.
- Carini, Robert M. 2002. "Teacher Unions and Student Achievement." Pp. 197-215 in *School Reform Proposals: The Research Evidence*, edited by Alex Molnar. Greenwich, CT: Information Age Publishing.
- Clotfelter, C., H. Ladd and J. Vigdor (2006) "Teacher-Student Matching and the Assessment of Teacher Effectiveness", NBER Working Paper No. 11936, January.
- Duflo, Esther and Rema Hanna (2005) "Monitoring Works: Getting Teachers to Come to School", mimeo, Poverty Action Lab, Massachusetts Institute of Technology. September.
- Eberts, R. W. and J. A. Stone (1987) "The effect of teacher unions on student achievement", *Industrial and Labor Relations Review*, 40 (3): 354-363. April.
- Eberts, Randall (2007) "Teacher unions and student performance: Help or hindrance?", *The Future of Children*, 17 (1): 175-200. Spring.
- Hanushek, Eric A., and Ludger Woessmann (2008) "The role of cognitive skills in economic development." *Journal of Economic Literature* 46, no.3 (September).
- Henderson, R., W. Urban and P. Wolman (2004) *Teacher Unions and Education Policy: Retrenchment or Reform?* Oxford. Elsevier.
- Hoxby, C. (1996) "How teachers' unions affect education production", *Quarterly Journal of Economics*, 111 (4): 671-718.

Kingdon, G. G. (1996) “The Quality and Efficiency of Public and Private Schools: A Case Study of Urban India”, *Oxford Bulletin of Economics and Statistics*, 58, No.1: 55-80, Feb. 1996.

Kingdon, G. G. and F. Teal (2007) “Does Performance Related Pay for Teachers Improve Student Achievement? Some Evidence from India”, *Economics of Education Review*, 26, No. 4: 473-86. August 2007.

Lemke, R. (2004) “Estimating the union wage effect for public school teachers when all teachers are unionized”, *Eastern Economic Journal*, 30(2): 273-91. Spring.

Loveless, T. (ed.) (2000) *Conflicting missions? Teachers unions and educational reform*. Washington, DC: Brookings Institution.

Machin, S. and S. McNally (2004) “The Literacy Hour”, Discussion Paper CEE 0043, Centre for Economics of Education, London School of Economics.

Muralidharan, K. and V. Sundararaman (2008) “Teacher performance pay: Experimental evidence from India.” Program on Education Policy and Governance, JF Kennedy School of Government, PEPP 08-16, Harvard University (May).

Rivkin, S., E. Hanushek and J. Kain (2005) “Teachers, Schools and Academic Achievement”, *Econometrica*, 73(2): 417-58.

Table 1
Mean and standard deviation of raw mark, by subject

Subject	Mean mark	SD	Minimum	Maximum
English	67.15	15.3	20	97
Second language	79.73	8.9	28	99
History-geography	70.09	14.3	20	99
Math	68.92	18.4	15	99
Science	65.72	17.0	20	99

Note: Descriptive statistics of standardized mark by subject are not reported as, by construction, the mean of standardized mark is 0 and standard deviation is 1 for each subject.

Table 2
Achievement production function
‘member of a union’ as the only included variable

	OLS	School FE	Pupil FE
Member of union	0.1816** (4.22)	-0.1755** (-5.28)	-0.1671** (-5.05)
Subject dummies	yes	yes	yes
Pupil variables	yes	yes	no
School variables	yes	no	no
N	49089	49089	49089
R-squared	0.26	0.35	---
Number of clusters	16	172	10016
F (p-value)	90.36 (0.000)	58.06 (0.000)	12.29 (0.000)

Table 3
Achievement production function
Including teacher characteristics

	Including only pre-determined teacher characteristics (TC)			Including characteristics in TC'
	OLS	School FE	Pupil FE	Pupil FE
Member of union	0.1788** (4.17)	-0.2342** (-6.91)	-0.2266** (-6.72)	-0.2347** (-6.91)
MA or PhD qualification	0.1206** (9.07)	0.0929** (9.63)	0.0914** (9.58)	0.0923** (9.51)
Pre-service teacher training	0.0602** (4.66)	0.0871** (10.0)	0.0885** (10.2)	0.0911** (10.5)
First division in hisec exam	0.1020** (5.35)	0.1092** (7.74)	0.1053** (7.52)	0.1113** (7.94)
Christian	0.0214 (1.43)	0.0254* (2.49)	0.0252* (2.49)	-0.0060** (-4.21)
Female	-0.0384** (-2.74)	0.0039 (0.40)	0.0020 (0.21)	0.0247* (2.45)
Age	0.0459** (10.2)	0.0272** (8.69)	0.0260** (8.38)	0.0049 (0.52)
Age square	-0.0005** (-9.51)	-0.0003** (-8.21)	-0.0003** (-7.84)	0.0240** (7.00)
Days of sick leave last year				-0.0002** (-6.38)
Total experience				-0.0040** (-5.37)
Tenure				0.0030** (4.07)
Permanent status				0.0293* (2.16)
Subject dummies	yes	yes	yes	yes
Pupil variables	yes	yes	--	no
School variables	yes	--	--	no
Observations	49089	49089	49089	49089
R-squared	0.27	0.35	---	---
Number of clusters	16	172	10016	10016
Estimated Bias (Equation (7))	0.57	0.27	0.91	0.88
Ratio (a)	0.31	--	--	--

(a) This is the ratio of the coefficient on union membership and the estimated bias.

Note: Constant included but not shown. Standard errors were corrected for clustering of errors between subjects within a pupil, i.e. we have used pupil id as the clustering variable. In all three columns, the fixed effects (state, school and pupil respectively) were highly statistically significant in F-tests (not shown). ** represents statistical significance at the 1% level and * at the 5% level.

Table 4
Robustness tests – the union coefficient in different samples

	Coefficient on union variable	Robust t-value
Political leaning of the state^(a)		
Non-Left state	-0.2187**	-3.21
Left state	-0.1497**	-3.64
Degree of school competition^(b)		
Low	-0.0943	-1.26
Medium	-0.1435**	-3.10
High	-0.4637**	-6.94
Three subject combinations		
English, second language, history-geography	-0.2942**	-4.70
English, second language, math	-0.4728**	-7.92
English, second language, science	-0.4480**	-9.25
English, history-geography, math	-0.0962*	-1.77
English, history-geography, science	-0.1399**	-3.03
English, math, science	-0.0353	-1.02
Second language, history-geography, math	-0.1481**	-2.93
Second language, history-geography, science	-0.1883**	-4.99
History-geography, math, science	-0.4429**	-11.01
Four subject combinations		
English, second language, history-geography, math	-0.2569**	-5.59
English, second language, history-geography, science	-0.2566**	-6.94
English, history-geography, math, science	-0.1049**	-2.74
English, second language, math, science	-0.2006**	-6.04
Second language, history-geography, math, science	-0.3431**	-8.47

Notes: All equations estimated using the pupil fixed effects estimator and using the exact specification of column 3 in Table 3, i.e. including controls for teacher variables and including subject dummy variables. ** represents statistical significance at the 1% level and * at the 5% level.

(a) 'Left leaning' states are West Bengal and Kerala, which have had communist parties in power in much of the post-independence period.

(b) School competition is measured by the number of schools situated within a 5 km radius of the school. Fewer than 4 nearby schools corresponds to 'low' and greater than 12 corresponds to 'high' competition.

Table 5
Pupil fixed effects achievement equation, on sub-sample of unionized schools

	All schools§	Unionized schools
Member of union	-0.2266** (-6.72)	-0.2428** (-5.94)
MA or PhD qualification	0.0914** (9.58)	-0.0631* (-2.10)
Pre-service teacher training	0.0885** (10.2)	0.0793** (3.52)
First division in hisec exam	0.1053** (7.52)	-0.1247* (-2.56)
Christian	0.0252* (2.49)	-0.0604 (-1.79)
Female	0.0020 (0.21)	-0.2096** (-7.88)
Age	0.0260** (8.38)	0.0261 (1.61)
Age square	-0.0003** (-7.84)	-0.0003 (-1.34)
Subject dummies	yes	Yes
Observations	49089	4790
Number of clusters (i.e. pupils)	10016	961

Notes: § these are results for the whole sample, taken from the third column of Table 3. ** represents statistical significance at the 1% level and * at the 5% level.

Note: In the full sample of (10016) students, i.e. in column 1, for 9.6% of all students at least one (out of their 5) teachers is unionized, and for only 0.4% of all students, all 5 of their teachers are unionized. Taking just the sub-sample of 961 students who study in schools that are unionized (column 2), all 961 students have at least 1 teacher who is unionized but only 3.95% of them have all 5 of their teachers unionized.

Table 6
Regression of log of teacher's monthly pay
(Teachers of grade 10 who teach the 5 main subjects only)

	Across school <u>(OLS)</u>	Within school <u>(school fixed effects)</u>
Teacher characteristics		
Member of union	0.3580* (2.33)	0.1498* (2.26)
MA or PhD	0.0555** (2.65)	0.0401** (2.84)
Teacher training	0.0697** (3.39)	0.0596** (4.40)
First division in hsec exam	-0.0315 (-1.00)	0.0134 (0.68)
Christian	0.0607* (2.25)	0.0618** (3.76)
Female	-0.0230 (-0.93)	-0.0013 (-0.092)
Age	0.0546** (8.08)	0.0340** (5.33)
Age square	-0.0005** (-6.05)	-0.0003** (-3.57)
Subject dummies	yes	yes
School variables	yes	---
State dummies	yes	---
N	1731	1731
R-squared	0.59	0.22
Number of scid_n	---	183
F (p-value)	---	35.43 (0.000)
Estimated Bias (Equation (7))	0.45	0.09
Ratio (a)	0.89	1.65

(a) This is the ratio of the coefficient on union membership and the estimated bias.

Table 7
Regression of log of teacher's monthly pay
(including the teacher characteristics in TC ')
(Teachers of grade 10 who teach the 5 main subjects only)

	Across school (OLS)	Within school (school fixed effects)
Teacher characteristics		
Member of union	0.2932* (2.13)	0.0947* (1.96)
MA or PhD	0.0681** (3.62)	0.0530** (4.30)
Teacher training	0.0427* (2.19)	0.0316** (2.72)
First division in hsec exam	-0.0537 (-1.77)	-0.0036 (-0.20)
Christian	0.0392 (1.52)	0.0397** (2.85)
Female	-0.0218 (-0.95)	-0.0170 (-1.29)
Age	0.0256** (3.90)	0.0107 (1.79)
Age square	-0.0003** (-3.68)	-0.0001 (-1.68)
Days of sick leave last year	-0.0073* (-2.14)	-0.0007 (-0.50)
Total experience	0.0061** (2.85)	0.0075** (4.30)
Tenure	0.0133** (6.68)	0.0095** (5.43)
Permanent contract	0.1756** (5.46)	0.1639** (5.54)
Subject dummies	yes	yes
School variables	yes	---
State dummies	yes	---
N	1731	1731
R-squared	0.65	0.39
Number of schools	---	183
F (p-value)	---	35.43 (0.000)
Estimated Bias (Equation (7))	0.44	0.08
Ratio (a)	0.74	1.20

(a) This is the ratio of the coefficient on union membership and the estimated bias.

Note: t-values are in parentheses and are robust to clustering at the school level. For School fixed-effects estimation, the number of groups is 183, i.e. estimation is within 183 schools. Mean number of teachers within a school is 9.5 (minimum=4, maximum=30). The F-test shows that the school dummies are jointly significant at the 0.000 level. Constant included but not shown. ** represents statistical significance at the 1% level and * at the 5% level.

Appendix

Selection on Unobservables

In the Altonji, Elder and Taber (2005) paper the issue is to assess the potential importance of unobservables in a bivariate probit model of a school outcome Y as a function of attending a catholic school (CH). To gauge the role of selection bias in a simple way they ignore the fact that Y is estimated by a probit and treat θ (the parameter of interest to establish the casual affect of catholic schooling on outcomes) as if it were estimated by a regression of the latent variable Y^* on X and CH where X is a vector of observable determinants of both the Y and CH variables. In our application the outcome, student achievement as measured by the standardised mark, is a continuous variable but that does not affect the logic of the approach.

We need to use an equation of the form:

$$(A1) \quad A = \theta \tilde{U} + [\gamma + \theta\beta]TC + \varepsilon$$

where we have dropped subscripts for clarity and

$$(A2) \quad U = \beta TC + \tilde{U}$$

which is equivalent to estimating our equation of interest which is:

$$(A3) \quad A = \theta U + \gamma TC + \varepsilon$$

By construction \tilde{U} and TC are orthogonal in (A1) and the standard formula for the potential bias in the θ parameter is

$$(A4) \quad \text{Plim } \hat{\theta} \approx \theta + \frac{\text{Cov}(\tilde{U}, \varepsilon)}{\text{Var}(\tilde{U})} = \frac{\text{Cov}(U, \varepsilon)}{\text{Var}(\tilde{U})}$$

The second equality follows from the fact that ε and TC are orthogonal.

The condition in the Altonji, Elder and Taber (2005) paper which defines the notion that “selection on the unobservables” is the same as selection on observables is equivalent to the condition that:

$$(A5) \quad \frac{\text{Cov}(U, \varepsilon)}{\text{Var}(\varepsilon)} = \frac{\text{Cov}(\gamma TC, U)}{\text{Var}(\gamma TC)}$$

The bias in (A4) can be written as:

$$(A6) \quad \frac{Cov(U, \varepsilon)}{Var(\tilde{U})} = \frac{Cov(U, \varepsilon)}{Cov(\gamma TC, U)} \frac{Var(\gamma TC)}{Var(\varepsilon)} \frac{Cov(\gamma TC, U)}{Var(\gamma TC)} \frac{Var(\varepsilon)}{Var(\tilde{U})}$$

Using (A5) this can be written as:

$$(A7) \quad \frac{Cov(U, \varepsilon)}{Var(\tilde{U})} = \frac{Cov(\gamma TC, U)}{Var(\gamma TC)} \frac{Var(\varepsilon)}{Var(\tilde{U})}$$

In their paper Altonji et. al. are concerned that a positive bias could be induced by the positive covariance between X and CH . In our application of their procedure we are concerned with the correlation between γTC and U . In fact these are strongly positively correlated. This explains our argument in the text that selection on unobservables suggests that our estimation of the negative effect is a lower bound to the true negative effect of unionisation on student achievement. We report in Appendix Table A1 the results for both the student achievement mark and for union membership as a function of the TC variables.

Appendix Table A1

**Pupil fixed effects Linear probability model of union membership
(union member=1; non-member=0) and of achievement score**

	Achievement score	Union membership
MA or PhD qualification	0.0903** (9.48)	0.0050** (4.19)
Pre-service teacher training	0.0805** (9.39)	0.0351** (19.67)
First division in hisec exam	0.1072** (7.65)	-0.0088** (-3.60)
Christian	0.0297** (2.94)	-0.0198** (-11.36)
Female	0.0001 (0.01)	0.0085** (5.22)
Age	0.0262** (8.44)	-0.0009** (-4.89)
Age square	-0.0003** (-7.93)	0.0000** (6.54)
Subject dummies	yes	yes
Observations	49089	49089
Number of clusters	10016	10016

Appendix Table A2
Variable definitions and descriptive statistics

Variable	Definition	Mean	SD
<u>Pupil characteristics</u>			
stdmark	Standardized mark	0.000	1.00
Male	Child is male*	0.491	0.50
Age in months	Child's age in months	195.055	8.90
Age in months squared	Age squared	38.126	3.56
No. younger brothers	Number of younger brothers	0.418	0.60
No. of older brothers	Number of older brothers	0.327	0.64
No. of younger sisters	Number of younger sisters	0.333	0.57
No. of older sisters	Number of older sisters	0.365	0.70
Father's education	Father's education in years	14.886	2.32
Mother's education	Mother's education in years	13.645	2.73
Wears spectacles	Child wears spectacles*	0.282	0.45
Disabled	Child is disabled*	0.003	0.06
Sibling in similar school	Any sibling studies in ICSE school	0.301	0.46
Raven test score	Score on Raven's ability test	23.824	6.40
Hours per day study at home	Hours per day spent in study at home	4.033	1.61
Hours per day domestic work	Hours per day spent in domestic work	0.692	0.57
Hours per day play	Hours per day spent in playing	1.631	0.86
Hours per day travel to school	Hours per day spent in travel to school	0.614	0.48
Hours per day spent in school	Hours per day spent in school	6.223	0.73
<u>Household characteristics</u>			
Household size	Household size	5.361	2.62
Religion Sikh	Religion is Sikh*	0.058	0.23
Religion Christian	Religion is Christian*	0.078	0.27
Religion Muslim	Religion is Muslim*	0.069	0.25
Wealth index	Index of asset ownership	21.629	6.31
Wealth index squared	Wealth squared	507.608	256.85
Other backward caste	'other backward caste'*	0.062	0.24
Schedule caste	'scheduled caste'*	0.012	0.11
Schedule tribe	'scheduled tribe'*	0.017	0.13
<u>Teacher characteristics</u>			
Gross monthly pay	Teachers' gross monthly pay (Rupees)	7864.421	2997.72
Log monthly pay	Log of teacher's gross monthly pay	8.884	0.40
MA or PhD	Teacher has MA/MSc/PhD*	0.741	0.35
Teacher training	Years of teacher training	0.920	0.36
First division in hisec exam	Teacher got first division in Higher Secondary exam (Proxy for teacher's cognitive skills)	0.092	0.23
Days of sick leave	Days of sick leave taken by teacher last year	2.182	2.70
Member of union	Teacher is member of a teacher union*	0.036	0.16
Christian	Teacher is Christian*	0.251	0.36
Female	Teacher is female*	0.723	0.39
Age	Teacher's age	41.122	7.92
Age square	Teachers age squared	1778.512	667.40
Total experience	Years of work experience in teaching	13.388	7.17

Tenure	Years of work experience in this school	10.037	6.67
Permanent contract	Teacher's contract is permanent?* Yes=1; no=0	0.878	0.27
<u>School characteristics</u>			
Total strength	Total number of pupils in the school	1648.789	852.18
Pupils in grade 10 (class size)	Number of pupils in grade 10	111.175	59.55
Principal influence index	Index of principal's influence in school decision-making: lowest=1, highest=5	4.161	0.93
Log monthly fee in grade 10	Log of monthly fee (rupees)	6.349	0.48
Girls school	Is a girls-only school*	0.210	0.41
Higher secondary school	Is a higher secondary school*	0.567	0.50
Index of school resources	Index of school resources	20.598	14.56

Note: Variables marked with a * are 0/1 indicator variables with yes=1 and no=0. The means of all variables (pupil, teacher and school) in this table have been computed using the pupil-subject level dataset.

Appendix Table A3
OLS Achievement production function, by subject

	<u>English</u>	<u>Second language</u>	<u>History- geography</u>	<u>Math</u>	<u>Science</u>
Pupil characteristics					
Male	-0.1686** (-8.71)	-0.3601** (-16.1)	-0.0148 (-0.69)	0.0367 (1.74)	0.0354 (1.66)
Age in months	0.1287** (6.63)	0.0816** (3.21)	0.1469** (6.34)	0.1807** (7.42)	0.1752** (7.26)
Age in months squared	-0.3505** (-7.24)	-0.2403** (-3.77)	-0.4093** (-7.05)	-0.4973** (-8.14)	-0.4862** (-8.04)
No. younger brothers	-0.0478** (-3.35)	-0.0275 (-1.59)	-0.0205 (-1.27)	-0.0278 (-1.66)	-0.0308 (-1.89)
No. of older brothers	-0.0392** (-2.98)	-0.0372* (-2.13)	-0.0284 (-1.89)	-0.0270 (-1.70)	-0.0342* (-2.20)
No. of younger sisters	-0.0248 (-1.66)	-0.0355* (-1.99)	-0.0180 (-1.07)	-0.0340* (-2.00)	-0.0459** (-2.70)
No. of older sisters	-0.0225 (-1.85)	0.0013 (0.094)	-0.0209 (-1.51)	-0.0155 (-1.08)	-0.0153 (-1.08)
Father's education	0.0319** (7.44)	0.0201** (3.90)	0.0319** (6.89)	0.0305** (6.37)	0.0356** (7.64)
Mother's education	0.0491** (12.8)	0.0193** (4.44)	0.0372** (9.07)	0.0376** (8.77)	0.0429** (10.3)
Wears spectacles	0.1377** (7.97)	0.0778** (4.06)	0.1120** (5.96)	0.1191** (6.27)	0.1173** (6.26)
Disabled	-0.0438 (-0.43)	-0.0703 (-0.43)	0.0018 (0.012)	-0.1633 (-1.02)	0.0622 (0.40)
Sibling in similar school	0.0231 (1.23)	-0.0690** (-3.10)	0.0395 (1.88)	0.0396 (1.85)	0.0258 (1.20)
Raven test score	0.0325** (26.6)	0.0253** (17.1)	0.0344** (24.3)	0.0438** (30.1)	0.0393** (28.0)
Hours per day in study at home	-0.0045 (-0.82)	0.0033 (0.54)	-0.0069 (-1.14)	0.0174** (2.89)	0.0125* (2.08)
Hours per day in domestic work	-0.0557** (-3.84)	-0.0230 (-1.35)	-0.0728** (-4.60)	-0.0831** (-5.11)	-0.0818** (-5.06)
Hours per day in play	-0.0186 (-1.95)	-0.0128 (-1.16)	-0.0174 (-1.63)	-0.0245* (-2.23)	-0.0277* (-2.54)
Hours per day in travel to school	0.0182 (1.13)	0.0046 (0.24)	-0.0070 (-0.39)	0.0212 (1.16)	0.0249 (1.39)
Hours per day spent in school	-0.0016 (-0.097)	0.0478** (2.40)	0.0605** (3.63)	0.0484** (2.63)	0.0043 (0.25)
Household characteristics					
Household size	-0.0060 (-1.86)	-0.0001 (-0.036)	-0.0012 (-0.33)	-0.0025 (-0.62)	-0.0057 (-1.50)
Religion Sikh	-0.1114** (-2.80)	-0.0976* (-2.19)	-0.0785 (-1.69)	-0.1285** (-2.77)	-0.0610 (-1.30)
Religion Christian	0.0575 (1.72)	-0.3537** (-8.58)	-0.0730* (-2.06)	-0.2140** (-5.62)	-0.1927** (-5.36)
Religion Muslim	-0.0240 (-0.77)	-0.1757** (-4.69)	-0.0824* (-2.42)	-0.1598** (-4.36)	-0.1039** (-2.95)
Wealth index	0.0132* (2.27)	-0.0002 (-0.027)	0.0178** (2.74)	0.0216** (3.23)	0.0115 (1.72)

Wealth index squared	-0.0004** (-2.78)	-0.0003 (-1.74)	-0.0006** (-3.77)	-0.0007** (-4.26)	-0.0005** (-3.14)
Other Backward Caste	-0.0370 (-1.09)	-0.0069 (-0.16)	-0.1041** (-2.79)	-0.0980** (-2.60)	-0.0697 (-1.90)
Schedule caste	0.0938 (1.22)	-0.0040 (-0.046)	-0.0011 (-0.013)	-0.0307 (-0.37)	0.0272 (0.34)
Schedule tribe	-0.0917 (-1.37)	-0.0087 (-0.11)	-0.2323** (-3.21)	-0.2467** (-3.29)	-0.1752* (-2.48)
Teacher characteristics					
Member of teacher union	0.1969** (2.95)	0.2667** (4.08)	0.0759 (1.00)	0.0614 (1.00)	-0.0124 (-0.23)
Log of monthly pay	0.4612** (11.2)	0.3730** (8.30)	0.4346** (9.67)	0.2501** (5.95)	0.3740** (8.45)
MA or PhD qualification	-0.0034 (-0.11)	0.1832** (5.42)	0.0407 (1.23)	0.0371 (1.30)	0.1737** (4.16)
Pre-service teacher training	0.0181 (0.67)	0.0130 (0.37)	0.0572 (1.74)	0.0285 (1.13)	-0.0702* (-2.11)
First division in hisec exam	-0.0150 (-0.32)	0.2420** (6.05)	0.0332 (0.66)	-0.1088* (-2.43)	-0.2201* (-2.39)
Days of sick leave last year	-0.0062 (-1.63)	-0.0134** (-4.34)	-0.0072 (-1.81)	-0.0157** (-4.39)	-0.0122** (-2.65)
Christian	0.0729** (3.16)	0.1151** (2.79)	-0.1503** (-4.69)	-0.1309** (-4.17)	-0.0268 (-0.67)
Female	0.0641* (2.07)	0.0410 (1.09)	-0.1059** (-3.66)	-0.0588* (-2.27)	0.0033 (0.080)
Age	0.0035 (0.42)	-0.0108 (-0.75)	0.0221 (1.58)	0.0255* (2.08)	-0.0057 (-0.34)
Age square	-0.0001 (-1.05)	0.0001 (0.89)	-0.0002 (-1.11)	-0.0004* (-2.44)	0.0000 (0.080)
Total experience	0.0051* (2.03)	-0.0003 (-0.15)	-0.0210** (-7.22)	0.0106** (4.00)	0.0180** (4.34)
Tenure	-0.0043 (-1.76)	-0.0036 (-1.74)	-0.0079** (-3.05)	-0.0037 (-1.73)	-0.0046 (-1.24)
Permanent status	-0.0047 (-0.15)	0.5208** (7.78)	0.2888** (6.33)	0.1260** (3.24)	0.0506 (1.21)
School characteristics					
Total strength	0.0001** (6.69)	0.0001** (5.90)	0.0002** (8.94)	0.0003** (13.4)	0.0002** (10.3)
Pupils (class-size) in grade 10	-0.0008* (-2.40)	-0.0013** (-3.78)	-0.0013** (-3.93)	-0.0030** (-8.73)	-0.0020** (-6.32)
Principal influence index	0.0537** (4.94)	0.0672** (5.12)	0.0733** (5.68)	0.0453** (3.84)	0.0353** (3.02)
Log monthly fee in grade 10	0.0875** (3.24)	-0.0004 (-0.013)	-0.1131** (-3.88)	-0.0595* (-2.12)	-0.0400 (-1.44)
Girls school	0.3391** (13.4)	0.0198 (0.69)	0.1826** (6.45)	0.1326** (4.35)	0.1479** (4.95)
Higher secondary school	-0.0987** (-4.99)	-0.0533* (-2.27)	-0.0588** (-2.74)	0.0592** (2.77)	-0.0034 (-0.15)
Index of school resources	0.0136** (14.1)	-0.0001 (-0.043)	0.0075** (6.88)	0.0044** (4.18)	0.0049** (4.76)
State dummies					
Andhra Pradesh	-0.3024** (-5.05)	-0.8926** (-11.6)	-0.4498** (-7.20)	-0.3454** (-5.34)	-0.2704** (-4.35)

Bihar	-0.5308** (-9.15)	-0.2934** (-3.89)	-0.3865** (-6.06)	-0.1288 (-1.86)	-0.2684** (-3.66)
Gujarat	-0.7744** (-8.71)	-0.9335** (-8.74)	-0.6923** (-6.71)	-0.3956** (-4.00)	-0.3686** (-4.05)
Haryana	-0.8423** (-14.1)	-1.0750** (-14.9)	-0.4381** (-6.30)	-0.2394** (-3.34)	-0.4024** (-5.48)
Himachal	-0.3235** (-5.56)	-0.6683** (-9.31)	-0.1949** (-3.15)	-0.3302** (-5.18)	-0.3363** (-5.28)
Karnataka	-0.3213** (-6.00)	-1.1284** (-16.4)	-0.2983** (-5.13)	-0.4948** (-8.18)	-0.4173** (-6.91)
Maharashtra	-0.6101** (-11.2)	-0.9220** (-12.6)	-0.4095** (-6.25)	-0.4357** (-6.75)	-0.4213** (-6.57)
Madhya Pradesh	-1.0217** (-15.3)	-0.9566** (-12.5)	-0.8276** (-11.5)	-0.5693** (-7.51)	-0.6783** (-9.46)
Orissa	-0.3292** (-5.98)	-0.5465** (-8.27)	-0.2795** (-4.85)	-0.1355* (-2.19)	-0.3109** (-5.00)
Punjab	-0.8400** (-13.8)	-0.6425** (-8.75)	-0.4538** (-6.54)	-0.3788** (-5.62)	-0.5958** (-8.42)
Rajasthan	-0.8771** (-10.5)	-0.8967** (-8.95)	-0.0721 (-0.84)	-0.3362** (-3.37)	-0.4704** (-4.69)
Tamil Nadu	-0.3212** (-4.78)	-0.3762** (-4.31)	-0.2887** (-3.66)	-0.1803* (-2.48)	-0.1857** (-2.60)
Uttar Pradesh	-0.8643** (-16.8)	-0.8614** (-13.9)	-0.7276** (-12.9)	-0.5292** (-8.86)	-0.7383** (-12.1)
West Bengal	-0.4901** (-9.36)	-1.0550** (-16.1)	-0.5533** (-9.27)	-0.4222** (-7.01)	-0.5451** (-8.98)
Chattisgarh	-0.8375** (-11.0)	-1.0733** (-9.78)	-0.7433** (-7.16)	-0.7562** (-7.40)	-0.6671** (-6.13)
_constant	-18.3856** (-9.26)	-11.1105** (-4.35)	-18.9495** (-8.17)	-21.2663** (-8.70)	-20.7108** (-8.54)
Observations	9772	9864	9833	9903	9717
R-squared	0.44	0.24	0.33	0.30	0.33

Note: the standard errors and thus t-values are corrected for the correlation of the errors between pupils within a school, i.e. we have used the pupil id as the clustering variable. The reference category for State is Kerala.