ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Goel, Deepti; Barooah, Bidisha

Working Paper

Drivers of Student Performance: Evidence from Higher Secondary Public Schools in Delhi

GLO Discussion Paper, No. 231

Provided in Cooperation with:
Global Labor Organization (GLO)

Suggested Citation: Goel, Deepti; Barooah, Bidisha (2018) : Drivers of Student Performance: Evidence from Higher Secondary Public Schools in Delhi, GLO Discussion Paper, No. 231, Global Labor Organization (GLO), Maastricht

This Version is available at:
https://hdl.handle.net/10419/180665

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Drivers of Student Performance: Evidence from Higher Secondary Public Schools in Delhi

Deepti Goel and Bidisha Barooah[1]

July 2018

Abstract: We examine the role of teachers and students in the formation of grade 12 test scores in public schools in Delhi. There is substantial variation in teacher and student quality within schools. Over the period spanning grades 11 and 12, being taught by a standard deviation better than average teacher raises test scores by 0.373 standard deviation; being a standard deviation better than average student raises it by 0.799 standard deviation. Being tenured positively predicts teacher effectiveness, while educational qualifications, training, experience and personality traits have no predictive power. Relative to families where only fathers earn, those where both parents earn negatively predict student effectiveness, while religion, caste and parents' education have no predictive power. There is some evidence that principals' rating of teachers and value added measures of teacher quality are positively correlated. We recommend teacher compensation be based in-part on ex-post evaluation of value-added measures and principal ratings.

JEL classification codes: I20, I23

Keywords: Value added, Test scores, Teacher quality, Student quality, India

[1] Deepti Goel (deepti@econdse.org), corresponding author, is an Assistant Professor at the Delhi School of Economics, Department of Economics, University of Delhi, Delhi 110007, India. Bidisha Barooah (bbarooah@3ieimpact.org) is an Evaluation specialist at International Initiative for Impact Evaluation (3ie), 202/203 Rectangle 1 D4, District Saket Centre, Saket Delhi 110017, India. The authors are grateful to The International Growth Centre (IGC), for financial support for this study.

1 Introduction

Cognitive skills, typically captured using test scores, are an important component of human capital that lead to higher individual earnings (Hanushek, Schwerdt, Wiederhold and Woessmann 2015), and boost a country's economic growth (Hanushek and Woessmann 2012). Understanding how they are formed is particularly crucial for developing countries that seek higher economic growth. In this paper, we examine this issue for India. This is of relevance to India because Indian students perform poorly in international tests and also exhibit greater test score variance.[2] The Annual Status of Education Reports (ASER) provide a more recent picture of poor learning outcomes among children in rural India: in 2016, only 43 percent of eight graders could correctly carry out a 3-digit by 1-digit division problem (ASER 2017). Given that gaps in test scores explain corresponding gaps in college attainment rates and earnings (Chay, Guryan and Mazumder 2014; Johnson and Neal 1998), it is conceivable that differences in skill acquisition are in part responsible for India's economic disparities which have been rising since the early nineties (Basole and Basu 2015; Subramanian and Jayaraj 2015). Against this background, we examine the role of teachers and students in the formation of test scores at the higher secondary level (grade 12) in public (government) schools in Delhi, India. We measure teacher and student effectiveness using the value added approach (Todd and Wolpin 2003). Accordingly, a higher quality teacher/student is characterized by greater ability to raise test scores after accounting for past and present inputs, including students' prior teachers and prior performance.[3]

A rich body of research has confirmed that variation in teacher assignment explains a large part of the variation in students' test scores (see Hanushek and Rivkin 2012 and Ladd 2008 for reviews of this

[2] Based on 2005 data for ninth graders from two Indian states, Das and Zajonc (2010) find that the median child failed to meet a low international benchmark in a standardized mathematics test, and the variance of test scores was amongst the highest in the world.

[3] We define teacher quality solely in terms of the teacher's ability to raise test scores and do not account for other important contributions that a teacher may make such as instilling curiosity or imparting good civic sense. In the absence of an alternative comprehensive measure, the ability to raise test scores provides a reasonable proxy for teacher quality and enables us to undertake quantitative analyses.

literature). Most of this work is from high income countries, predominantly the United States.[4] Of late, new research is emerging from middle income countries: Araujo, Carneiro, Cruz-Aguayo and Schady 2016 from Ecuador, Azam and Kingdon 2015 from India, Bau and Das 2017 and Talanché 2017 from Pakistan, and Metzler and Woessmann 2012 from Peru. We add to this growing literature by being the first to examine the role of teachers in *public* schools in India. Our paper complements that by Azam and Kingdon (2015) who also examined teacher quality at the higher secondary level, but did so for a *private* consortium of schools in the Indian state of Uttar Pradesh. There are many reasons why private and public schools might function differently and should therefore be studied separately.[5] First, on average, public schools are less productive compared to private schools: their test scores are either the same (Chudgar or Quin 2012) or worse (Azam, Kingdon and Wu 2016; Muralidharan and Sundararaman 2015; Singh 2015), while their per-student expenditure is larger, mainly on account of higher teacher salaries (Pritchett and Aiyar 2015).[6] Second, teachers in public schools have different characteristics compared to their private counterparts: they are less likely to have a college degree, more likely to have a formal teacher training certificate, more likely to be absent, and less likely to be teaching when present (Desai, Dubey, Vanneman and Banerji 2008; Muralidharan and Kremer 2009). Finally, being subsidized by the state, public schools largely cater to students from lower income backgrounds which may lead to different peer effects and teacher-student dynamics within these schools. The only other paper to have looked at teacher quality in public schools at the higher secondary level is Slater et al. 2012 who study schools in England. In the concluding section we compare our findings with both Azam and Kingdon 2015 and Slater et al. 2012.

[4] Aaronson, Barrow and Sander 2007; Chetty, Friedman and Rockoff 2014; Clotfelter, Ladd and Vigdor 2010; Hanushek, Kain, O'Brien and Rivkin 2005; Rivkin, Hanushek and Kain 2005; and Rockoff 2004 are select studies from the United States; Slater, Davies and Burgess 2012 from the United Kingdom; and Leigh 2010 from Australia.
[5] Even though the share of private schools in enrolment is increasing in India, public schools still account for the larger share. At the all-India level, in 2014, 56 percent of students at the secondary and higher secondary levels (grades 10 through 12) studied in public schools. The corresponding figure for Delhi is higher at 71 percent (National Sample Survey 2016).
[6] The studies cited here are based on analyses at the primary (grades 1 to 5) or secondary (grade 10) levels, and not at the higher secondary level. We could not find any study comparing public with private schools at the higher secondary level.

The broad consensus in the education literature on the correlates of teacher effectiveness is that most observable teacher characteristics such as gender, experience, educational qualifications and training are not strongly associated with value added measures of teacher quality (Hanushek and Rivkin 2010).[7] In addition to examining whether these standard observable characteristics matter in our context, we also examine specific personality traits of teachers. We measure personality traits using *The Big Five* test, the most widely accepted taxonomy of traits in psychology (John, Naumann and Soto 2008). Its personality dimensions are: extraversion, agreeableness, conscientiousness, neuroticism and openness, each of which summarizes a large number of more specific personality traits.[8] Recently, economists have also begun to conceptualize (Almlund, Duckworth, Heckman and Kautz 2011; Borghans, Duckworth, Heckman and ter Weel 2008) and test (Lindqvist and Vestman 2011; Mueller and Plug 2006) the importance of personality traits for labor market outcomes. Lindqvist and Vestman show that, compared to cognitive ability, personality traits are stronger predictors of labor force participation and earnings at the low end of the earnings distribution. Motivated by this evidence, and an intuitive expectation that personality traits of teachers should influence the relationship between them and their students and thereby affect test scores, we examine whether the Big Five can predict teacher effectiveness.

There is a substantial body of work that has looked at the importance of family background on human capital (Björklund and Salvanes 2011 provide a review).[9] The link between the two is indicative of equality of opportunity wherein a stronger association signifies greater inequality as it captures the salience of factors that are outside the control of an individual. This literature has focused on several specific characteristics including household income (Chevalier, Harmon, Sullivan and Walker 2013;

[7] A notable exception to this conclusion is Clotfelter et al. 2010. Additionally, some studies (Hanushek et al. 2005; Leigh 2010) have found that early experience adds to teacher effectiveness.
[8] Select trait adjectives associated with each dimension are: talkative and assertive with extraversion; sympathetic and kind with agreeableness; organized and thorough with conscientiousness; tense and anxious with neuroticism; wide interests and imaginative with openness. See John et al. 2008 for an extensive list of trait adjectives.
[9] They focus on completed years of schooling as the outcome of interest and conclude that more than 50 percent of the variation in it can be attributed to factors shared by siblings.

Thompson 2014), parental education (Lundborg, Nilsson and Rooth 2014), parental time investments (Boca, Monfardini and Nicoletti 2017; Cunha and Heckman 2008) and family size and birth order (Haan, Plug and Rosero 2014, Sen and Clemente 2010). The value added specification that we estimate includes both teacher and student fixed effects. The latter captures all student factors (that are not subject specific), including individual ambition level and family background. We examine whether some of the aforementioned background characteristics studied in the literature can predict student effectiveness in the context we study.

Finally, we also document the correlation between subjective principal ratings of individual teachers and objective value added measures of teacher quality embodied in our estimated teacher effects. The literature on such comparisons is limited, and as far as we know spans only the United States (Harris and Sass 2014; Jacob and Lefgren 2008; Rockoff and Speroni 2011). We add to this nascent literature by providing the first such comparison from a developing country context. Principals have the opportunity to observe teachers closely and to evaluate them on a broader spectrum of outcomes beyond test scores, some of which may be hard to measure but are valued by parents and the society at large. Thus, there is merit in including principal evaluations as one of the factors that determines teacher compensation. It would then be useful to know how they compare with value added measures, which may be used as another determinant of teacher pay.

Before we present the analysis we briefly review our main findings. Using data collected from 22 schools for the academic year 2015-16, we find substantial variation in teacher quality within schools: over the two year period spanning grades 11 and 12, being taught by a one standard deviation better than average teacher raises a student's test scores by 0.373 standard deviation (averaged across all subjects). This would move a student at the middle of the test score distribution to the 65^{th} percentile. In terms of identifying an effective teacher, we find that consistent with most studies, educational qualifications, teacher training, and seniority do not characterize a better teacher; being permanent (tenured or confirmed

in the government job) positively predicts teacher effectiveness; and contrary to our expectations none of the Big Five personality traits correlate with teacher effectiveness. Turning to student factors, we find that the within school variation in student quality is at least twice that in teacher quality. In terms of correlates of student effectiveness, religion, caste and parents' education have no predictive power; interestingly, relative to families where only the fathers earn, those where both parents earn negatively predict student performance. Finally, we find weak evidence (significant at 10 percent) that principals' assessment of teachers regarding how well they can increase students' understanding of the subject is positively correlated with valued added measures of teacher quality, with this correlation increasing when the principals have known the teachers for a longer duration.

2 Methodology

Consider the following production function for test scores:

$$A_{izt} = \Phi_1(T_{izt} + T_{izt-1} + \cdots + T_{iz0}) + \Phi_2(\boldsymbol{X}_{izt} + \boldsymbol{X}_{izt-1} + \cdots + \boldsymbol{X}_{izo}) + \Phi_3(\theta_{it} + \theta_{it-1} + \cdots + \theta_{i0}) + \varepsilon_{izt} \quad (1)$$

where A_{izt} is the test score of student i, in subject z, in year t. Educational inputs in year t are: T_{izt}, the quality of the student's subject specific teacher; $\boldsymbol{X}_{izt}$, a vector of subject specific student inputs such as time spent studying at home; and θ_{it}, the student's *subject invariant* ability such as general motivational level. Lagged values refer to corresponding inputs in previous years. ε_{izt} is the residual error.

2.1 Estimating Teacher and Student Effects: Two-level Fixed Effects

Under standard assumptions in this literature (Todd and Wolpin 2003),[10] equation (1) can be written as:

$$A_{izt} = \beta T_{izt} + \alpha\beta T_{izt-1} + \cdots + \alpha^t \beta T_{iz0} + \boldsymbol{\gamma X}_{izt} + \alpha\boldsymbol{\gamma X}_{izt-1} + \cdots + \alpha^t \boldsymbol{\gamma X}_{izo}$$

[10] We make the following assumptions: there is a linear relationship between each input ($T, \boldsymbol{X}$ and θ) and the test score; the marginal impact of each input ($\beta, \boldsymbol{\gamma}$, and δ) is age-invariant; and the impacts of all past inputs decay at a constant annual rate, $1 - \alpha$.

$$+\delta\theta_{it} + \alpha\delta\theta_{it-1} + \cdots + \alpha^{t}\delta\theta_{i0} + \varepsilon_{izt} \quad (2)$$

Combining terms we get,

$$A_{izt} = \alpha A_{izt-1} + \beta T_{izt} + \boldsymbol{\gamma X}_{izt} + \delta\theta_{it} + u_{izt} \quad (3)$$

where $u_{izt} = \varepsilon_{izt} - \alpha\varepsilon_{izt-1}$.[11] Value addition due to current (subject specific) teacher quality, βT_{izt}, and current student ability, $\delta\theta_{it}$, are estimated using subject specific teacher fixed effects, and student fixed effects, respectively.

Introducing subscripts for teacher j and school s, superscripts for grades 10 and 12, and suppressing the time subscript (we study a single cohort of students),[12] the two-level fixed effects specification that we estimate is as follows:

$$A^{12}_{izjs} = \alpha A^{10}_{iz} + \{\tau_{zjs}\} + \boldsymbol{\gamma X}_{iz} + \{\omega_i\} + u_{izjs} \quad (4)$$

where $\{\tau_{zjs}\}$ and $\{\omega_i\}$ are teacher[13] and student fixed effects, respectively. We measure A^{12}_{izjs} in terms of subject specific z-scores which accounts for some of the subject specific idiosyncrasies in the achievement distribution. However, previous research (Hanushek and Rivkin 2012), has shown that teacher effects differ by subject in spite of using z-scores.[14] We therefore regress the teacher effects on

[11] For Ordinary Least Squares (OLS), to give consistent estimates of model parameters we either require ε to be serially correlated with the degree of serial correlation equal to α (so that u is independent identically distributed), or we need instruments for lagged test score, A_{izt-1}, such as past inputs or past test scores, (Todd and Wolpin 2003). In the absence of valid instruments we assume the former. This assumption has been implicitly followed by others including Azam and Kingdon (2015) and Slater et al. (2012).

[12] Lack of multi-year data would be a limitation if we wanted to arrive at a ranking of individual teachers that was persistent over time (McCaffrey, Sass, Lockwood and Mihaly 2009). However, assuming stationarity of the distribution of teacher effects, data for a single year is sufficient to characterize the overall distribution in any given year.

[13] Whenever an individual teacher teaches multiple subjects, a distinct fixed effect is assigned for each subject. Thus, although we refer to our estimated effects as teacher effects, they are more precisely subject-teacher effects.

[14] Hanushek and Rivkin (2012) collate evidence from multiple studies for the United States and document an average standard deviation of teacher effects of 0.13 for reading, and 0.17 for math.

subject dummies and calculate the standard deviation of the resulting residuals as our measure of the variation in teacher quality. This estimator captures the importance of teacher assignment on test score performance. Similarly, the standard deviation of student effects measures the variation in student quality.[15]

Since teachers and students do not shift schools within the school year, it is not possible to separate their effects from school effects. We therefore report within school variation in estimated teacher and student effects.[16] As pointed out by Slater et al. (2012), the within school estimator is a lower bound for the actual degree of variation in quality across all public schools, if, as is likely, teachers and students cluster in schools on the basis of their quality.

The inclusion of *contemporaneous* student fixed effects implies that teacher effects are derived from within student across subject variation in test scores at a point in time. To a large extent this addresses non-random matching of students and teachers. To fully address it, we require that, having conditioned for lagged subject specific scores, students should not be matched with teachers according to other *subject specific* abilities.[17]

---

[15] The precision of estimated teacher and student effects depends on the number of students taught and the number of subjects taken, respectively. To account for this we calculate weighted standard deviations, using students taught and subjects taken as weights. Further, to account for sampling variation in estimated teacher and student effects, we use bootstrap (150 replications) to estimate the standard errors of the standard deviation of (residual) teacher and student effects.

[16] We implement equation (4) in STATA using `felsdvregdm'. Unlike previous commands that deal with over parameterization by dropping an arbitrary hold out unit, felsdvregdm imposes a sum to zero constraint within each reference collection, school in our case. School means are not separately identified in the presence of student fixed effects (Mihaly, McCaffrey and Lockwood 2010)

[17] The inclusion of contemporaneous student fixed effects allows for dynamic tracking of students (Rothstein 2010), wherein students are matched with teachers according to their most recent ability measure. For consistent estimates of teacher effects we require this ability measure to be *subject-invariant* or to be based on grade 10 subject scores. It would be problematic if students were matched to teachers according to unobserved expectation of subject specific performance, or according to revealed subject specific performance in grade 11. In our schools the section of a student determines teacher assignments. Most students do not change sections between 11 and 12. For the ones that do, it happens because sections are merged into larger ones due to attrition in grade 11. In our school visits we did not find any indication of selective matching of individual students with subject specific teachers.

In our empirical specification, the vector $\boldsymbol{X}$ in equation (4) includes the following subject specific inputs applied by the student: whether or not tuitions were taken, study time outside school, and interest (captured using self-reported ranking of subjects according to the student's liking). It may be argued that some of these inputs are responses to perceived or actual teacher quality. If this is the case and our interest is in the overall (policy) effect of a change in teacher quality, then these student inputs should not be included as controls. Alternatively, if interest lies in production function parameters, or if one believes that these inputs are driven by unobserved student subject specific heterogeneity such as parental motivation to excel in particular subjects, then they should be included as controls. We therefore present two separate specifications, with and without $\boldsymbol{X}$, which bound our estimator of teacher influence. It is not possible to predict which specification would lead to a larger standard deviation of teacher effects. If the inputs in $\boldsymbol{X}$ are substitutes for (complements to) teacher quality, then controlling for them would give larger (smaller) estimates.

2.2 Identifying Correlates of Teacher and Student Quality

To identify the correlates that distinguish an effective teacher/student, estimates of teacher/student fixed effects are regressed on individual characteristics.[18]

2.3 Correlating Principal Ratings with Value Added Measures of Teacher Quality

Principals were required to rate teachers along multiple performance dimensions on an ordinal scale of 1 (lacking a lot) to 5 (exceptional). Given that principals may have different subjective scales of rating, we normalize the ratings along each performance dimension by subtracting from each rating the principal specific mean for that dimension and then dividing by the principal specific standard deviation for that dimension. Recall that the value added measures captured by teacher effects have also been normalized within school, making the comparison between ratings and teacher effects a sensible one.

[18] As in the calculation of the standard deviation of teacher effects, the estimated teacher effects are purged of subject effects before they are regressed on individual characteristics.

Our value added measures are based on a single cohort of students and consequently include year-specific idiosyncrasies in teacher performance. On the other hand, it is likely that principals' ratings are based on a more stable metric of performance. As noted by Jacob and Lefgren (2008), the estimation error in value added measures results in underestimating the correlation between principal ratings and the true teacher effects. We adopt the correction suggested by them (explained briefly in Appendix 1) and present both the un-adjusted and corrected correlations between principal ratings and estimated teacher effects.

3 Data

Data come from 22 higher secondary schools of the Directorate of Education (DOE), Delhi Government, and were obtained using a combination of primary surveys and schools' administrative records. The schools were selected using convenience sampling. In Appendix 2 we examine whether they constitute a representative sample of the population of DOE schools and find no evidence to suggest otherwise.

We study a single cohort of students (and their teachers) who appeared for the higher secondary examinations held in March 2016.[19] A total of 2,207 students from our study schools took these exams.[20] Of these, we are able to study 1,733 students (78.5 percent) who form our analysis sample. The remaining students had to be dropped for the following reasons: (a) could not obtain grade 10 records (14.7 percent);[21] (b) two or more students from the same class had the same name (8.1 percent);[22] (c) section

[19] The Central Board of Secondary Education (CBSE), conducts these examinations. They are ideal to study teacher and student effectiveness because they are based on material that the teachers are supposed to teach, students take them seriously as they determine admission into colleges and other professional institutions, and there is little scope for manipulation of scores because they are set and graded outside the schools.

[20] These students belong to either the Arts, Commerce or Science streams.

[21] Each school is well represented in the analysis sample even after excluding these students as we have grade 10 records for at least 70 percent of the students in each school.

[22] These students had to be dropped as they could not be matched with their corresponding grade 12 scores.

information for grade 12 and/or grade 11 was missing (1.0 percent).[23] In the next paragraph we talk about the implications of dropping these students.

Under the value added approach, each subject in grade 12 has to be matched with a counterpart in grade 10. We used some discretion to find the closest match, but given the wide range of subjects taught in grade 12 it was not possible to find a match for each one.[24] Consequently, we restrict the analysis to 16 subjects for which a match could be found, and which were chosen by at least 100 students. Table 1 lists the subjects we study along with their matches in grade 10.[25]

Table 1: Subject-wise Performance of Students in the Analysis Sample (1,733 Students)

	Grade 12 (Z-Scores)			Grade 10 (Percent Scores)		
	Mean	**Std. Dev.**	**No. of Students**	**Matching Subject**	**Mean**	**Std. Dev.**
English	0.059	0.986	1,572	English	56.0	14.2
Hindi	0.097	0.986	1,233	Hindi	53.3	14.0
Political Science	0.077	0.986	860	Social Science	51.9	15.2
Economics	0.026	1.008	823	Social Science	63.4	14.7
History	0.087	1.001	776	Social Science	51.9	15.1
Accountancy	0.022	1.000	536	Mathematics	51.3	15.7
Business Studies	0.015	1.005	535	Social Science	65.2	14.7
Mathematics	0.033	0.983	452	Mathematics	65.7	16.0
Home Science	0.069	0.989	416	Social Science	57.3	16.5
Geography	0.069	0.987	336	Social Science	50.8	13.8
Chemistry	0.033	0.989	203	Science	72.9	14.8
Physics	0.036	1.007	203	Science	72.9	14.8
Sanskrit	0.007	1.015	164	Sanskrit	44.5	10.3
Sociology	0.087	1.000	112	Social Science	44.6	15.1
Biology	0.077	1.015	110	Science	72.9	14.7
Agriculture	0.153	1.012	100	Science	41.8	11.0

The dependent variable for estimating teacher and student effects is a student's subject specific z-score in grade 12. It is constructed using the subject mean and subject standard deviation calculated over those

[23] These students had to be dropped as they could not be matched with their subject teachers.
[24] We do not study students from the Vocational stream because for most vocational subjects it was not possible to find a matching subject in grade 10.
[25] 76.3 percent of students in the analysis sample had taken 5 of these 16 subjects, 18.0 percent had taken 4, 5.3 percent had taken 6, and 0.4 percent had taken 3 subjects.

students (among all 2,207 students in our study schools) who opted for that subject. Table 1 presents the subject wise mean z-scores in grade 12 for our analysis sample.[26] The positive (instead of 0) mean z-scores for all subjects suggest that our analysis sample consists of better performing students. Using the Kolmogorov–Smirnov test, we find that for English, Hindi, political science (the three most popular subjects), history, home science, geography, agriculture and biology, we reject that the distributions are identical between students in the analysis sample and those who were dropped, in favor of our analysis sample having higher test scores. For the remaining subjects we cannot reject that the distributions are identical across the two groups. Since identification of teacher fixed effects relies on within student across subject comparison, positive or no selection for all subjects should allow us to consistently estimate teacher effects for the type of students (i.e. better performing ones) we study. Student effects are understandably estimated for these better performers.

Recall that for estimating teacher and student fixed effects, the unit of observation is a student-subject-(grade)-teacher match. We used class-wise teaching schedules (time-tables) to link each student's subject specific z-score in grade 12 to the individual teachers who taught that subject in grades 11 and 12. This resulted in 18,552 student-subject-(grade)-teacher observations,[27] arising from 273 subject-teachers (these stem from 254 individual persons).[28]

We conducted voluntary surveys of teachers, students and principals to collect information on their individual characteristics and to capture principals' ratings of teachers. Students were surveyed within

[26] It also shows the mean percent score for each matching subject in grade 10. For grade 10 we only had letter grades which were converted to percent scores using the mid-point of the range for each letter grade.
[27] For a given student-subject test score in grade 12, let N1 and N2 be the numbers of teachers that taught the subject in grades 11 and 12, respectively. This would result in a total of (N1+N2) observations related to that student-subject test score. During estimation, we assign a weight of 1/(N1+N2) to each observation. In this way, each student-subject test score in grade 12 receives equal weight irrespective of the number of contributing teachers, and each contributing teacher is assumed to have made an equal contribution.
[28] For 327 of 18,552 student-subject-grade combinations (1.8 percent), we could not find information on the teacher. These are treated as missing in the analysis.

schools on the days that we visited, while teachers and principals were allowed to fill up their questionnaires at leisure which were later collected by our enumerators. Take up rates for the teacher, student and principal surveys are 68.1 (186/273), 68.7 (1,190/1,733) and 72.7 (16/22) percent, respectively. We discuss selection issues arising from partial take up in the results section.

The teacher survey included the 44-item Big Five test (see John and Srivastava 1999 for the Big Five instrument and its scoring sheet). Against each item, the teacher was required to choose between one and five on a Likert scale indicating the extent to which he/she agreed or disagreed with the statement made. On the basis of these responses a score was calculated for each personality dimension wherein a higher score indicates greater strength of that dimension. The Big Five scores can only be calculated when *all* 44 items are answered. While 186 subject teachers took the teacher survey, the Big Five component was fully answered by 145 of them.[29] For a more meaningful interpretation of regression results, we normalize the Big Five scores for each dimension using the mean and standard deviation for that dimension across all 145 teachers who fully answered the test.

Table 2 presents summary statistics for surveyed teachers and students. 87 percent of teachers are post graduates, 71 percent have received special teacher training, 84 percent are permanent (tenured), and the average teaching experience is about 20 years. Turning to students, only 38 percent of their fathers and 21 percent of their mothers passed grade 12. Thus, many of them would become first generation graduates. In terms of parental earnings, 92 percent of their fathers and 19 percent of their mothers earn, while in 5 percent of the cases neither parent earns. 82 percent of the students have a commute time to school (one-way) that is less than an hour.

[29] Cronbach's alpha is a measure of internal consistency of a test. It ranges between 0 and 1, the larger the value the greater is the inter-correlation among test items. As a rule of thumb, a value greater than 0.7 is considered as indicating an acceptable level of internal consistency (Nunnally and Bernstein 1994). For our 44-item Big Five, the Cronbach's alpha for agreeableness, conscientiousness, extraversion, neuroticism and openness are, 0.817, 0.792, 0.659, 0.618 and 0.832, respectively.

Table 2: Summary Statistics for Surveyed Teachers and Students		
Teacher Characteristics	**Mean (Std. Dev.)**	**No. of Teachers**
Female (in %)	51.1	186
Married (in %)	89.2	186
Hindu (in %)	93.0	186
General Caste (in %)	68.3	186
Post Graduate Teacher (in %)	86.6	183
Trained (in %)	71.0	186
Permanent (in %)	83.9	186
Teaching Experience (in years)	20.1 (10.2)	184
Big Five Raw Scores		
Agreeableness (max. possible score 45)	28.8 (3.2)	145
Conscientiousness (max. possible score 45)	28.8 (3.7)	145
Extraversion (max. possible score 40)	25.3 (3.7)	145
Neuroticism (max. possible score 40)	17.1 (4.7)	145
Openness (max. possible score 50)	36.8 (5.1)	145
Student Characteristics	**Mean (Std. Dev.)**	**No. of Students**
Age (in years)	17.8 (1.0)	1190
Female (in %)	45.6	1190
Hindu (in %)	83.1	1190
General Caste (in %)	61.9	1190
Father Passed Twelfth (in %)	37.7	1190
Mother Passes Twelfth (in %)	21.4	1190
Parents' Earning Status (in %)		
Only Father Earns	76.1	1190
Only Mother Earns	3.3	1190
Both Parents Earn	15.6	1190
Neither Parent Earns	5.0	1190
Commute less than 1 hour (one way, in %)	81.8	1165
Birth Order (in %)		
One	32.3	1119
Two	34.2	1119
Three	20.6	1119
Four or Higher	13.0	1119
Number of Siblings	2.4 (1.4)	1004

4 Results

We first present our findings related to teachers and students, followed by a comparison between principals' ratings and value added measures of teacher quality.

4.1 Importance of Teacher Assignment

Table 3 presents the standard deviation of estimated teacher effects from multiple specifications that differ in terms of subject specific student inputs.

Table 3: Importance of Teacher Assignment

Specification	Weighted Std. Dev. of TFE[1]	Bootstrap Std. Err.[2]	Number of Subject Teachers
Without student inputs	0.373	0.005	273
Only tuitions	0.385	0.007	263
Only study time	0.385	0.007	263
Only interest	0.381	0.006	263
All three student inputs	0.383	0.007	263

Notes: TFE stands for teacher fixed effects.

1 Weighted by the number of observations per subject teacher.

2 Bootstrap standard errors calculated using 150 replications

When none of the student inputs are included, the estimate is 0.373, and when all three inputs are included it is 0.383. These are to be interpreted as follows: over the two year period spanning grades 11 and 12, being taught by a one standard deviation better than average teacher adds 0.373 (or 0.383) standard deviation to a student's test score (averaged across subjects). This would move a student at the middle of the test score distribution to the 65^{th} percentile (in both cases). Thus, there exists substantial variation in teacher quality within schools. Another important observation is that although the increase in standard deviation from 0.373 to 0.383 suggests that student inputs are substitutes for teacher quality, the magnitude of increase is small in real terms, making no difference to the percentile of student performance.

4.2 Identifying an Effective Teacher

We use information on surveyed teachers to identify the correlates of teacher effectiveness. Given 68.1 percent completion rate for the teacher survey, it is important to examine whether surveyed teachers differ from non-surveyed ones. In Appendix 3 we examine teacher effectiveness by surveyed status and find no

evidence that surveyed teachers are more or less effective compared to non-surveyed ones. While this by itself does not preclude individual characteristics of the two groups from being different, it makes it less likely that it is so, and makes the discussion below more meaningful. In Appendix 3 we also look at teacher effectiveness according to whether teachers fully completed the Big Five component of the survey. Once again, we find no evidence that teachers who fully completed the Big Five are different compared to those who did not.

Table 4 presents the correlates of teacher effectiveness. We use teacher effects estimated using the specification without controlling for student inputs as it gives us estimates for all 273 subject teachers. Three different specifications are shown: (1) includes only standard observable characteristics typically studied in this literature, (2) includes only the normalized Big Five scores, and (3) includes both these. The results show that being permanent (i.e. tenured or confirmed in the job) positively predicts teacher effectiveness. The effect size is large and remains so irrespective of whether we control for personality dimensions. Contrary, to our expectations, the personality dimensions have very little predictive power and none of them are significant in predicting teacher effectiveness. Consistent with a large body of existing literature, gender, educational qualifications, teacher training, and seniority cannot distinguish a good from a bad teacher.

Table 4: Predicting Teacher Effectiveness						
Dependent variable is Teacher Fixed Effects (TFE) from the specification without student inputs						
	(1)		**(2)**		**(3)**	
	Coefficient	**Std. Error**	**Coefficient**	**Std. Error**	**Coefficient**	**Std. Error**
Female	0.001	0.065			0.040	0.073
Married	0.146	0.100			0.167	0.138
Hindu	-0.190	0.110			-0.265	0.215
General Caste	0.020	0.071			-0.019	0.083
Post Graduate Teacher	-0.160	0.098			-0.116	0.128
Trained	0.063	0.059			0.147*	0.084
Permanent	0.297***	0.084			0.231**	0.099
Teaching Experience (omitted >20 years)						
Less than equal to 2	0.288***	0.062			0.209	0.149
Greater than 2 but less than equal to 5	0.319*	0.175			0.294	0.181
Greater than 5 but less than equal to 10	-0.015	0.080			0.030	0.099
Greater than 10 but less than equal to 20	-0.008	0.102			0.003	0.113
Personality Dimensions						
Agreeableness			0.069	0.035	0.071	0.043
Conscientiousness			-0.080	0.072	-0.076	0.063
Extraversion			0.043	0.037	0.063	0.039
Neuroticism			0.044	0.047	0.042	0.047
Openness			-0.018	0.065	-0.038	0.069
Number of subject teachers (observations)	184		145		144	
R squared	0.079		0.044		0.149	
Mean TFE among 186 surveyed teachers is -0.046, with a standard deviation of 0.400.						
Regressions are weighted using number of observations per subject teacher.						
Standard errors clustered at the school level (22 clusters).						
* indicates significant at 10 percent, ** significant at 5 percent and *** significant at 1 percent						

4.3 Importance of Student Factors

We calculate the standard deviation of student effects only for estimated effects from the specification that does not include any student inputs.[30] The estimated standard deviation is 0.799 with a (bootstrap) standard error of 0.009.[31] This means that being a one standard deviation better than average student translates into a 0.799 standard deviation improvement in test scores over a period of two years. This

[30] For this specification, effects for all 1,733 students are estimated; whereas for specifications which include student inputs, effects for only surveyed students are estimated.
[31] The standard deviation is a weighted by the number of observations per student.

would move a student at the middle of the test score distribution to the 79th percentile. The standard deviation of student effects is about twice the standard deviation of teacher effects. Moreover, as described in the data section, our analysis is based on better performing students. Consequently, had we considered all students our estimate of the variance in student quality would have been even higher. Thus, within schools, the variation in student quality is much greater than the variation in teacher quality.

4.4 Identifying an Effective Student

The take up rate for the student survey is 68.7 percent. It is therefore important to check whether surveyed students form a select sample. In Appendix 4 we examine whether surveyed students differ from non-surveyed ones in terms estimated student effects and do not find support for this.

Table 5 presents the correlates of student effectiveness. We use student effects estimated using the specification without controlling for student inputs as it gives us effects for all 1,733 students in the analysis sample. Three different specifications are shown: (1) includes only socio-demographic and parent characteristics, (2) also includes commute time to school, (3) also includes birth order and total number of sibling. We find that being older negatively predicts student effectiveness. This is not surprising as age is most likely picking up the effect of having failed the examination previously and repeating it again. This suggests that repeaters continue to lag behind in subsequent attempts. There is weak evidence that female students perform better than their male counterparts. The coefficient is fairly stable across specifications, though it is significant only at the 10 percent level in some of them. Belonging to the dominant Hindu religion, or belonging to the General caste category (a proxy for the historically privileged upper castes), are not correlated with student quality. Surprisingly, parents' education, in terms of being higher secondary graduates themselves, also does not predict student effectiveness. A consistent finding across all three specifications is that relative to families where only the father earns, those where both parents earn correlate negatively with student performance. We can only conjecture that, conditioning on one parent working, mothers being present at home positively influences a child's performance. We do not

comment on the remaining variables as the results are not robust across specifications or the sample size drops considerably due to missing observations.

Table 5: Predicting Student Effectiveness						
Dependent variable is Student Fixed Effects (SFE) from the specification without student inputs						
	(1)		**(2)**		**(4)**	
	Coefficient	**Std. Error**	**Coefficient**	**Std. Error**	**Coefficient**	**Std. Error**
Age	-0.132***	0.026	-0.127***	0.028	-0.119***	0.027
Female	0.296**	0.134	0.258*	0.137	0.244*	0.135
Hindu	0.086	0.084	0.089	0.081	0.111	0.069
General Caste	0.046	0.049	0.041	0.045	0.074	0.048
Father Passed Twelfth	0.053	0.078	0.070	0.078	0.115	0.081
Mother Passes Twelfth	0.071	0.074	0.067	0.071	0.047	0.071
Parents' Earning Status (omitted only father earns)						
Only Mother Earns	0.068	0.145	0.039	0.146	0.080	0.176
Both Parents Earn	-0.168***	0.057	-0.199***	0.057	-0.275***	0.068
Neither Parent Earns	-0.110	0.090	-0.104	0.102	-0.018	0.152
Commute time less than 1 hour			0.150	0.074	0.247***	0.071
Birth Order (omitted first born)						
Two					0.056	0.034
Three					0.027	0.065
Four or Higher					0.060	0.078
Number of siblings					-0.016	0.014
Stream dummies	Yes		Yes		Yes	
Number of students (observations)	1190		1165		986	
R squared	0.177		0.176		0.173	
Mean SFE among 1190 surveyed students is -0.421 with a standard deviation of 0.800.						
Regressions are weighted using number of observations per student. Standard errors clustered at the school level (22 clusters).						
* indicates significant at 10 percent, ** significant at 5 percent and *** significant at 1 percent						

4.5 Principal Ratings and Value Added Measures of Teacher Quality

Principals were required to rate all teachers in their school who taught grades 11 and 12. Principals from 16 schools completed the survey.[32] Of a total of 196 subject teachers in these 16 schools, principals rated 159 of them, a completion rate of 81 percent.[33] Appendix 5 examines whether incomplete ratings result in

[32] In two of our schools the principal's post was vacant and the vice-principal was the acting head of the school. The survey was completed by the vice-principals in these schools.

[33] While principals may have consciously not rated some teachers, some ratings may be missing for the following reason. Some teachers may have taught the higher secondary level only in the academic year

a select sample by testing whether teacher effects differ by rating status. We find no evidence that this is the case.

Table 6 presents summary statistics for surveyed principals. All of them are post graduates, with a teaching experience of about 27 years on average.

Table 6: Summary Statistics for Surveyed Principals		
	Mean (Std. Dev.)	**No. of Principals**
Age (in years)	54.2 (5.1)	15
Female (in %)	50.0	16
Hindu (in %)	93.8	16
General Caste (in %)	75.0	16
Post Graduate (in %)	100.0	16
Duration as Principal of School (in years)[1]	4.7 (5.0)	16
Teaching Experience (in years)	26.9 (5.6)	16
1 One principal reported a tenure of 22 years, while the rest reported a tenure of less than 9.		
Omitting this outlier, the average tenure value drops to 3.6 years with a standard deviation of 2.3		

Table 7 provides summary statistics for principals' ratings of teachers. The ratings encompass six performance dimensions, each marked on a scale of 1 (lacking a lot) to 5 (exceptional). As seen, the ratings are high in general, with a mean of 3.92, and a 10-90 percentile range from 3 to 5. In Appendix 6 we present the pairwise correlations between the six performance dimensions. All correlations are positive and significant at the 1 percent level. The lowest pairwise correlation is between increasing students' understanding of the subject and positive relationship with colleagues (0.44), and the highest is between increasing students' understanding of the subject and liked by students (0.76).

2014-15, and not in 2015-16. These teachers may have been omitted as the principal survey was administered in February/March 2016 and principals were instructed to rate teachers who taught at the higher secondary level in the *current* academic year.

Table 7: Summary Statistics for Principals' Ratings of Teachers (raw ratings on a scale of 1 to 5)			
Performance Dimension	**Mean (Std. Dev.)**	**10th Percentile**	**90th Percentile**
Dedication and Work Ethic	3.96 (0.78)	3	5
Increasing Students' Understanding of the Subject	3.91 (0.66)	3	5
Liked by Students	3.87 (0.64)	3	5
Maintaining Discipline Among Students	4.03 (0.76)	3	5
Positive Relationship with Colleagues	3.97 (0.65)	3	5
Organizational/Administration Skills	3.77 (0.79)	3	5
These statistics are based on 159 subject teachers in the analysis sample.			

Table 8 presents the un-adjusted and corrected correlations between estimated teacher effects and principal's ratings. We focus on two performance dimensions, namely, (1) a teacher's dedication and work ethic, and (2) a teacher's ability to increase students' understanding of the subject as (1) has been commonly used in this literature, and (2) comes closest to measuring the same underlying construct as our value added measures of teacher effectiveness.[34] The first column presents the unadjusted correlations. As stated in the methodology section, these correlations are biased towards zero. For performance measures (1) and (2) the estimate values are 0.08 and 0.12, respectively, with only the latter being significant at the 10 percent level. Once we apply the correction for estimation error in the value added measures, the correlations increase to 0.4 and 0.6, respectively. Once again, only the latter value is significant at the 10 percent level. Thus, we find weak evidence of principal ratings being positively correlated with our estimated value added measures of teacher effectiveness.

[34] When subjective and objective measures do not capture the same underlying construct, the correlations are biased downwards (Jacob and Lefgren 2008).

Table 8: Correlation between Principal Ratings and Value Added Measure of Teacher Effectiveness				
	Un-adjusted		Corrected	
	Correlation	Std. Error	Correlation	Std. Error
Dedication and Work Ethic	0.079	0.087	0.400	0.445
Increasing Students' Understanding of the Subject	0.118	0.071	0.602	0.358
Liked by Students	0.117	0.083	0.596	0.415
Maintaining Discipline Among Students	0.097	0.084	0.496	0.434
Positive Relationship with Colleagues	0.107	0.093	0.546	0.477
Organizational/Administration Skills	0.167	0.072	0.852	0.368
Correlations based on ratings of 159 subject teachers from 16 schools in the analysis sample.				
Standard errors calculated using 150 bootstrap replications				

In Table 9, we present the results by replicating the exercise on a restricted sample of teachers that the principals have known for at least 3 years. Ex-ante we expect the correlations to be higher for this select sample. As seen, the correlations increase for four of the six performance measures.

Table 9: Correlation between Principal Ratings and Value Added Measure of Teacher Effectiveness				
For teachers known to the Principal for at least 3 years				
	Un-adjusted		Corrected	
	Correlation	Std. Error	Correlation	Std. Error
Dedication and Work Ethic	0.054	0.111	0.265	0.545
Increasing Students' Understanding of the Subject	0.171	0.093	0.849	0.451
Liked by Students	0.193	0.093	0.956	0.44
Maintaining Discipline Among Students	0.172	0.107	0.853	0.535
Positive Relationship with Colleagues	0.072	0.130	0.355	0.652
Organizational/Administration Skills	0.193	0.086	0.955	0.423
Correlations based on ratings of 106 subject teachers from 13 schools in the analysis sample.				
Standard errors calculated using 150 bootstrap replications				

5. Conclusions and Policy Implications

We conclude with a discussion of findings from our study of higher secondary public schools in Delhi. We find substantial variation in teacher quality within schools: over the two year period spanning grades 11 and 12, being taught by a one standard deviation better than average teacher raises a student's test scores by 0.373 standard deviation (estimate from the specification without student inputs). This would

move a student at the middle of the test score distribution to the 65th percentile. Our estimate underestimates the overall variation in teacher quality as it does not account for between school variations. To the best of our knowledge these are the first quantitative estimates from India of the importance of teacher assignment in public schools. Azam and Kingdon (2015) undertook a similar exercise for a consortium of private schools in the adjoining Indian state of Uttar Pradesh, while Slater et al (2012) did so for public schools in England. Like us, both looked at the higher secondary level and found the standard deviation of teacher effects to be 0.366 and 0.233, respectively.[35] A student at the middle of the achievement distribution would move to the 64th percentile (Azam and Kingdon), and the 59th percentile (Slater et al.), in their contexts. Thus, at the higher secondary level, teachers in Delhi's public schools play a comparable role to that documented in a study of private schools in Uttar Pradesh. Given our rich data on students, we could account for subject specific student inputs (specifically, tuitions, study time and interest) in the estimation of test scores. When these inputs are accounted for, the standard deviation of teacher effects increases from 0.373 to 0.383, implying that student inputs are substitutes for teacher quality. However, the magnitude of increase is small in real terms, not translating to a percentile position change in student performance. This should allay fears that students may drastically cut back on their own private inputs as a reaction to institutional attempts to raise teacher quality.

In terms of identifying an effective teacher, we find that being permanent positively predicts teacher effectiveness. Permanency means being confirmed in the government job, which at least in Delhi, is characterized by regular pay (much higher than what is paid in the private sector), with little possibility of being fired until retirement age. Our finding that it positively predicts a better teacher is consistent with Maslow's Hierarchy of Needs (Maslow 1943). According to this theory, a person is able to make greater contributions to society and realize his/her full potential only after their lower level needs, which includes

[35] Neither study controlled for student inputs, so we contrast their values with our estimate of 0.373.

economic security, are met. At first glance, this may seem at variance with the recent emphasis laid on financial incentives for teachers in order to improve learning outcomes (Duflo, Hanna and Ryan 2012; Muralidharan and Sundararaman 2011). We do not see the contradiction as incentives need not become redundant in the presence of job security. Further, contrary to our expectations, we find that none of the Big Five personality traits are correlated with teacher effectiveness. Finally, post-graduation degree, specialized teacher training, and higher teaching experience do not characterize a better teacher, a finding consistent with Azam and Kingdon, and Slater et al. Unfortunately, in the present institutional set up, it is these observable characteristics that count in determining teacher promotions and compensation. To move closer to a system that incentivizes good performance, policy makers should seriously consider fixing part of a teacher's salary on ex-post evaluation of value-added performance measures. This is already being done in several states in the United States and lessons can be drawn from their experience.[36]

In addition to studying the variation in teacher quality, we also analyze the importance of student factors in the formation of test scores. Being a one standard deviation better than average student translates into a 0.799 standard deviation improvement in test scores over a period of two years. This would move a student at the middle of the test score distribution to the 79^{th} percentile. Thus, the within school variance in student quality is twice that in teacher quality, and this is an underestimate as our analysis is based on better performing students. This is consistent with Slater et al. who find the ratio to be 2.7 for higher secondary public schools in England. Turning to identifying an effective student, we find that being older negatively predicts student effectiveness suggesting that repeaters continue to lag behind in subsequent attempts to clear the exam. Contrary to our expectations, parents' education does not predict student effectiveness. Interestingly, relative to families where only the fathers earn, those where both parents earn negatively predict student performance. We can only conjecture that, conditioning on one parent working,

[36] It would be unwise to move to a reward structure that is solely based on value added measures as teachers may then only teach to the test and neglect other aspects of being a good educator.

stay at home mothers positively influence their child's performance. It would be interesting to explore the robustness and mechanisms for this result in future work.

Finally, we find weak evidence (significant at the 10 percent level) that principals' assessment of teachers on how well they can increase students' understanding of the subject is positively correlated with valued added measures of teacher quality. The correlation increases when the principals have known the teachers for at least three years. We recommend that principal evaluations be incorporated in the reward structure for teachers. Compared to ex-post evaluation of value added measures which requires investment in creating and maintaining large administrative datasets, this is a low cost alternative to incentivizing better teacher performance. Moreover, principals may be able to evaluate teachers on a broader set of outcomes than just their effectiveness in improving test scores.

References

Aaronson, Daniel, Lisa Barrow and William Sander (2007): "Teacher and Student Achievement in Chicago Public High Schools," *Journal of Labor Economics,* Vol 25(1), pp 95-135.

Almlund, Mathilde, Angela Lee Duckworth, James J. Heckman and Tim D. Kautz (2011): "Personality Psychology and Economics." In Eric A. Hanushek, Stephen Machin and Ludger Woessmann (Eds) Handbook of the Economics of Education Vol 4, pp 1-181.

Araujo, M. Caridad, Pedro Carneiro, Yyannu Cruz-Aguayo and Norbert Schady (2016): "Teacher Quality and Learning Outcomes in Kindergarten," *The Quarterly Journal of Economics,* Vol 131, No 3, pp 1415-1453.

ASER (2017): "Annual Status of Education (Rural) Report 2016 (Provisional)," New Delhi.

Azam, Mehtabul, and Geeta G. Kingdon (2015): "Assessing Teacher Quality in India," *Journal of Development Economics,* Vol 117, pp 74-83.

Azam, Mehtabul, Geeta G. Kingdon and Kin Bing Wu (2016): "Impact of Private Secondary Schooling on Cognitive Skills: Evidence from India," *Education Economics,* Vol 24, No 5, pp 465-480.

Boca, Daniela Del, Chiara Monfardini and Cheti Nicoletti (2017): "Parental and Child Time Investments and the Cognitive Development of Adolescents," *Journal of Labor Economics,* Vol 35, No 2, pp 565-608.

Basole, Amit, and Deepankar Basu (2015): "Non-food Expenditures and Consumption Inequality in India," *Economic and Political Weekly,* Vol L, No 36, pp 43- 53.

Bau, Natalie, and Jishnu Das (2017): "The Misallocation of Pay and Productivity in the Public Sector: Evidence from the Labor Market for Teachers," Policy Research Working Paper No. 8050, World Bank, Washington, DC.

Björklund, Anders, and Kjell G. Salvanes (2011): "Education and Family Background: Mechanisms and Policies," In Eric A. Hanushek, Stephen Machin and Ludger Woessmann (Eds) Handbook of the Economics of Education Vol 3, pp 201-247.

Borghans, Lex, Angela Lee Duckworth, James J. Heckman and Bas ter Weel (2008a): "The Economics and Psychology of Personality Traits," *The Journal of Human Resources,* Vol 43, No 4, pp 815-858.

Chay, Kenneth Y., Jonathan Guryan and Bhashkar Mazumder (2014): "Early Life Environment and Racial Inequality in Education and Earnings in the United States," Working Paper 20539, NBER.

Chetty, Raj, John N. Friedman and Jonah E. Rockoff (2014): "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value –Added Estimates," *American Economic Review,* Vol 104, No 9, pp 2593-2632.

Chevalier, Arnaud, Colm Harmon, Vincent O' Sullivan and Ian Walker (2013): "The impact of parental income and education on the schooling of their children," *IZA Journal of Labor Economics,* Vol 2, No 8, pp 2-22.

Chudgar, Amita, and Elizabeth Quin (2012): "Relationship between Private Schooling and Achievement: Results from Rural and Urban India," *Economics of Education Review,* Vol 26, No. 6, pp 376-390.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor (2010): "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects," *Journal of Human Resources,* Vol 45, No 3, pp 655-681.

Cunha, Flavio, and James J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources,* Vol 43, No 4, pp 738-782.

Das, Jishnu, and Tristan Zajonc (2010): "India Shinning and Bharat Drowning: Comparing two Indian States to the Worldwide Distribution in Mathematics Achievement," *Journal of Development Economics,* Vol 92, pp 175-187.

Desai, Sonalde, Amaresh Dubey, Reeve Vanneman and Rukmini Banerji (2008): "Private Schooling in India: A New Educational Landscape," *India Policy Forum,* Vol 5, No 1, pp 1-58.

Duflo, Esther, Rema Hanna and Stephen P. Ryan (2012): "Incentives Work: Getting Teachers to Come to School," *American Economic Review,* Vol 102, No 4, pp 1241-1278.

Haan, Monique De, Erik Plug and Jose Rosero (2014): "Birth Order and Human Capital Development: Evidence from Ecuador," *The Journal of Human Resources,* Vol 49, No 2, pp 359-392.

Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold and Ludger Woessmann (2015): "Returns to Skills around the World: Evidence from PIAAC," *European Economic Review,* Vol 73, pp 103-130.

Hanushek, Eric A., John F. Kain, Daniel M. O'Brien and Steven G. Rivkin (2005): "The Market for Teacher Quality," Working Paper 11154, NBER.

Hanushek, Eric A., and Ludger Woessmann (2012): "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation," *Journal of Economic Growth,* Vol 17, pp 267-321.

Hanushek, Eric A., and Steven G. Rivkin (2010): "Generalizations about Using Value-Added Measures of Teacher Quality, *American Economic Review: Papers and Proceedings* 100, pp 267-271.

_____ (2012): "The Distribution of Teacher Quality and Implications for Policy," *Annual Review of Economics,* Vol 4, No 1, pp 131-157.

Harris, Douglas N., and Tim R. Sass (2014): "Skills, productivity and the evaluation of teacher performance," *Economics of Education Review,* Vol 40, pp 183-204.

Jacob, Brian A., and Lars Lefgren (2008): "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education," *Journal of Labor Economics,* Vol 26, No 1, pp 101-136.

John, Oliver P., Laura P. Naumann and Christopher J. Soto (2008): "Paradigm Shift to the Integrative Big Five Trait Taxonomy: History, Measurement, and Conceptual Issues." In Oliver P. John, Richard W. Robins and Lawrence A. Pervin (Eds) Handbook of Personality: Theory and Research 3rd Edition, pp 114-158, New York, NY, US: Guilford Press.

John, Oliver P., and Sanjay Srivastava (1999): "The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives." In Lawrence A. Pervin and Oliver P. John (Eds) Handbook of Personality: Theory and Research 2nd Edition, New York, NY, US: Guilford Press.

Johnson, William R., and Derek Neal (1998): "Basic Skills and the Black-White Earnings Gap," In: Jencks C, Phillips M (Eds) The Black-White Test Score Gap. Brookings Institution, Washington DC.

Ladd, Helen F. (2008): "Teacher effects: What do we know?" *Journal of Educational Behavioral Statistics*, Vol 29, No 1, pp 37-66.

Leigh, Andrew (2010): "Estimating Teacher Effectiveness from Two Year Changes in Students' Test Scores," *Economics of Education Review*, Vol 29, pp 480-488.

Lindqvist, Erik, and Roine Vestman (2011): "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment," *American Economic Journal: Applied Economics*, Vol 3, pp 101-128.

Lundborg, Petter, Anton Nilsson and Dan-Olof Rooth (2014): "Parental Education and Offspring Outcomes: Evidence from the Swedish Compulsory School Reform," *American Economic Journal: Applied Economics*, Vol 6, No 1, pp 253-278.

Maslow, Abraham H. (1943): "A Theory of Human Motivation," *Psychological Review*. Vol 50, No 4, pp 370-96.

McCaffrey, Daniel F., Tim R. Sass, J.R. Lockwood and Kata Mihaly (2009): "The Inter-Temporal Variability of Teacher Effect Estimates," *Education Finance and Policy*, Vol 4, No 4, pp 572-606.

Metzler, Johannes, and Ludger Woessmann (2012): "The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation," *Journal of Development Economics*, Vol 99, pp 486-496.

Mihaly, Kata, Daniel F. McCaffrey and J. R. Lockwood (2010): "Centering and reference groups for estimates of fixed effects: Modifications to felsdvreg," *The STATA Journal*, Vol 10, No 1, pp 82-103.

Mueller, Gerrit, and Erik Plug (2006): "Estimating the Effect of Personality on Male and Female Earnings," *Industrial and Labor Relations Review*, Vol 60, pp 3-22.

Muralidharan, Karthik, and Michael Kremer (2009): "Public-Private Schools in Rural India." In Rajashri Chakrakbarti and Paul Peterson (Eds), School Choice International: Exploring Public-Private Partnerships. Cambridge, MA: MIT Press.

Muralidharan, Karthik, and Venkatesh Sundararaman (2011): "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, Vol 119, No 1, pp 39-77.

_____ (2015): "The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India," *The Quarterly Journal of Economics*, Vol 130, No 3, pp 1011-1066.

National Sample Survey (2016): "Education in India," Ministry of Statistics and Programme Implementation, National Sample Survey Office, Government of India, New Delhi.

Nunnally, Jum C., and Ira H. Bernstein (1994): "Psychometric Theory." 3rd edition. New York: McGraw–Hill.

Pritchett, Lant, and Yamini Aiyar (2015): "Value Subtraction in Public Sector Production: Accounting versus Economic Cost of Primary Schooling in India, Working Paper No. 297, Center for International Development, Harvard University.

Rivkin, Steven G., Eric A. Hanushek and John F. Kain (2005): "Teachers, Schools, and Academic Achievement," *Econometrica*, Vol 73, No 2, pp 417-458.

Rockoff, Jonah E. (2004): "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *The American Economic Review*, Vol 94, No 2, pp 247-252.

Rockoff, Jonah E., and Cecilia Speroni (2011): "Subjective and objective evaluations of teacher effectiveness: Evidence from New York City," *Labour Economics*, Vol 18, pp 687-696.

Rothstein, Jesse (2010): "Teacher Quality in Educational Production: Tracking, Decay and Student Achievement," *The Quarterly Journal of Economics*, Vol 125, No 1, pp 175-214.

Sen, Anindya, and Anthony Clemente: "Intergenerational correlations in educational attainment: Birth order and family size effects using Canadian data," *Economics of Education Review*, Vol 29, No 1, pp 147-155.

Singh, Abhijeet (2015): "Private School Effects in Urban and Rural India: Panel Estimates at Primary and Secondary School Ages," *Journal of Development Economics*, Vol 113, pp 16-32.

Slater, Helen, Niel M. Davies and Simon Burgess (2012): "Do Teachers Matter? Measuring the Variation in Teacher Effectiveness in England," *Oxford Bulletin of Economics and Statistics*, Vol 74, No 5, pp 629-645.

Subramanian, S, and D Jayaraj (2015): "Growth and Inequality in the Distribution of India's Consumption Expenditure: 1983 to 2009-10," *Economic and Political Weekly*, Vol L, No 32, pp 39- 47.

Talanché, Marine de (2017): "Better Teachers, Better Results? Evidence from Rural Pakistan," *The Journal of Development Studies*, Vol 53, No 10, pp 1697-1713.

Thompson, Owen (2014): "Economic Background and Educational Attainment: The Role of Gene-Environment Interactions," *The Journal of Human Resources*, Vol 49, No 2, pp 263-294.

Todd, Petra E., and Kenneth I. Wolpin (2003): "On the Specification and Estimation of the Production Function for Cognitive Achievement," *The Economic Journal*, Vol 113, pp F3-F33.

Appendix 1: Correcting for Estimation Error in the Correlation between Principal Ratings and Value Added Measures of Teacher Effectiveness

This discussion is based on the exposition in Jacob and Lefgren 2008. Let $\hat{p}$ denote principal rating and $\hat{\delta}$ denote estimated value added measure of teacher effectiveness i.e. the estimated teacher effect. $\hat{\delta} = \delta + e$, where δ is the true teacher effect and e is the estimation error.

The un-adjusted correlation between principal rating and estimated value added measure is given by:

$$Corr(\hat{p}, \hat{\delta}) = \frac{Cov(\hat{p}, \hat{\delta})}{\sqrt{Var(\hat{p})Var(\hat{\delta})}}$$

The corrected correlation is the correlation between principal rating and true teacher effect and is given by:

$$Corr(\hat{p}, \delta) = Corr(\hat{p}, \hat{\delta}) \frac{\sqrt{Var(\hat{\delta})}}{\sqrt{Var(\delta)}}$$

where,

$$Var(\delta) = Var(\hat{\delta}) - Mean\ of\ [std\ error(\hat{\delta})^2]$$

We estimate $std\ error(\hat{\delta})$ using 150 bootstrap replications of $\hat{\delta}$.

Appendix 2: School Selection

We followed convenience sampling based on school location. Figure A2.1 presents a map of Delhi showing the 22 sampled schools. As seen, there is geographical clustering of schools. Here we check whether our sampled schools represent the population of DOE schools.

Figure A2.1: Map of Delhi showing sampled schools

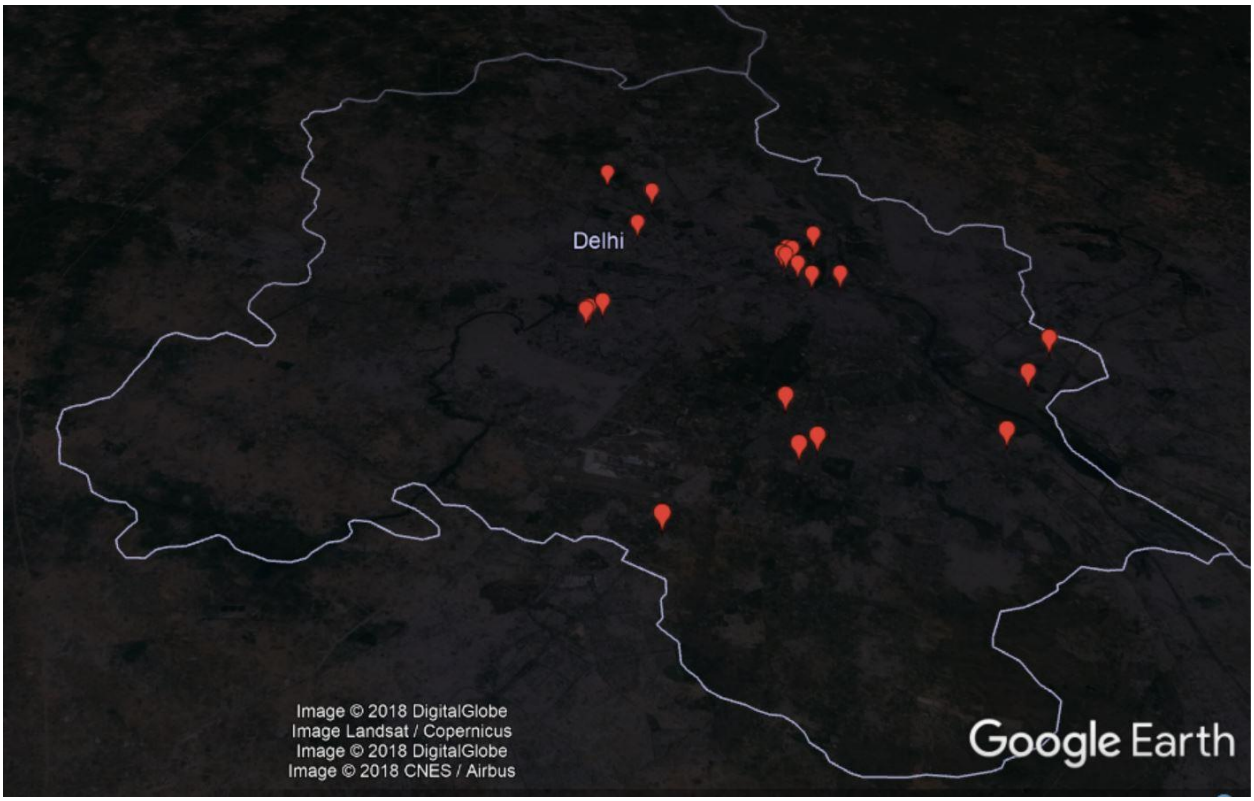


Table A2.1 shows select descriptive statistics by sampling status. At the 10 percent level of significance, it would seem that we oversampled smaller schools. However, the Bonferroni corrected critical value for testing 15 characteristics at this level is 0.0067. Given that all the p-values are above this cut off, we do not reject the null that our sample is representative.

Table A2.1: Descriptive Statistics for Higher Secondary Public Schools by Sampling Status, 2015-16					
	All Schools	**Sampled**	**Not Sampled**	**Difference**	**p-value**
Number	871	22	849		
Gender, Percent					
Girls	43.7	36.4	43.9	-7.5	0.4767
Boys	41.2	45.5	41.1	4.4	0.6927
Co-ed	15.0	15.0	18.2	-3.2	0.7048
Shift, Percent					
Morning	69.9	81.8	69.6	12.2	0.1544
Evening	30.1	18.2	30.4	-12.2	
Mean Number of Students	150.8 (111.6)	109.6 (62.9)	151.9 (112.4)	-42.3	0.080
Streams Taught, Percent					
Has Arts	99.1	95.5	99.2	-3.7	0.4143
Has Commerce	69.4	72.7	69.3	3.4	0.7247
Has Science	30.8	40.9	30.5	10.4	0.3377
Pupil Teacher Ratio	26.5 (14.7)	24.7 (9.6)	26.6 (14.8)	-1.9	0.565
Pupil Classroom Ratio	39.4 (17.6)	35.9 (11.8)	39.5 (17.7)	-3.2	0.347
Mean Pass Percentage					
Arts	89.6 (12.8)	90.5 (13.2)	89.5 (12.8)	1.0	0.742
Commerce	91.6 (12.1)	93.6 (8.2)	91.6 (12.2)	2.0	0.509
Science	95.2 (9.9)	98.1 (4.2)	95.1 (10.1)	3.0	0.371
Mean Marks per Student					
Arts	285.6 (32.0)	295.0 (31.3)	285.3 (32.0)	9.7	0.171
Commerce	296.2(37.6)	302.6 (30.0)	296.0 (37.8)	6.6	0.488
Science	336.3 (37.4)	349.5 (20.4)	335.9 (37.8)	13.6	0.284
Source: Directorate of Education, Government of National Capital Territory of Delhi					
Standard deviation in parentheses. p-values are for the null hypothesis that the 'Difference' is zero.					

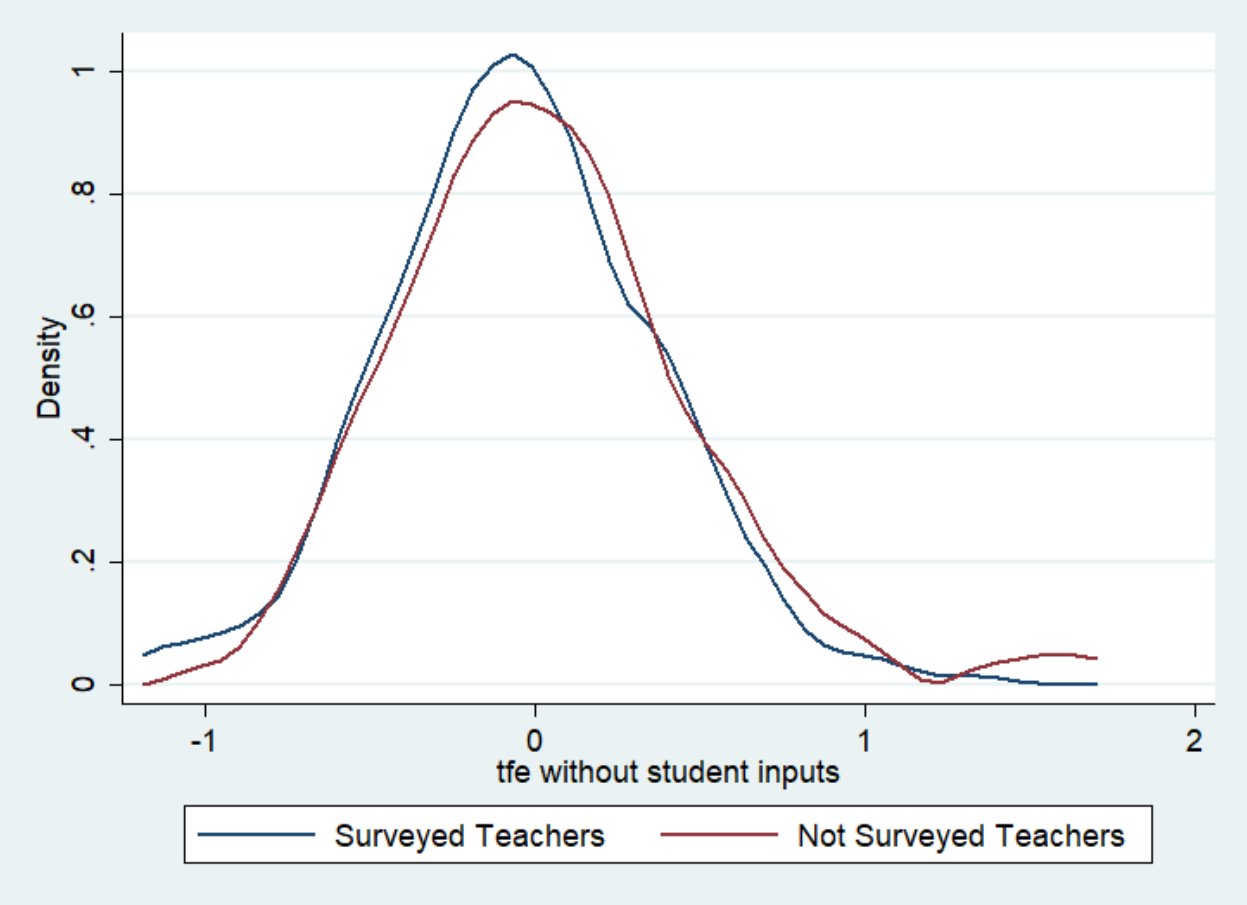
Another way to look at whether our sample is representative is to calculate the probability of drawing a random sample with our sample characteristics. We do this exercise for the mean marks per student. Assuming that in each stream, the mean marks per student follows a normal distribution with the population mean and standard deviation, the probability of drawing a random sample of size 22 with a mean within a 5 percent interval of the observed sample mean in Arts, Commerce and Science is 0.783, 0.858 and 0.704, respectively. These are reasonably high to claim that our sample is representative of the population of all DOE schools.

Appendix 3: Teacher Effects by Surveyed Status

Out of a total of 273 subject teachers, 186 filled up the teacher survey. Since we have estimates of teacher effects for all 273 teachers, we examine whether surveyed teachers are more or less effective compared to those who chose not to participate in our survey.

Figure A3.1 shows the estimated kernel densities of teacher fixed effects (from the specification without controlling for student inputs) for surveyed and non-surveyed teachers. Using the Kolmogorov-Smirnov test for equality of distribution functions we cannot reject that the two distributions are identical (p-value 0.579).

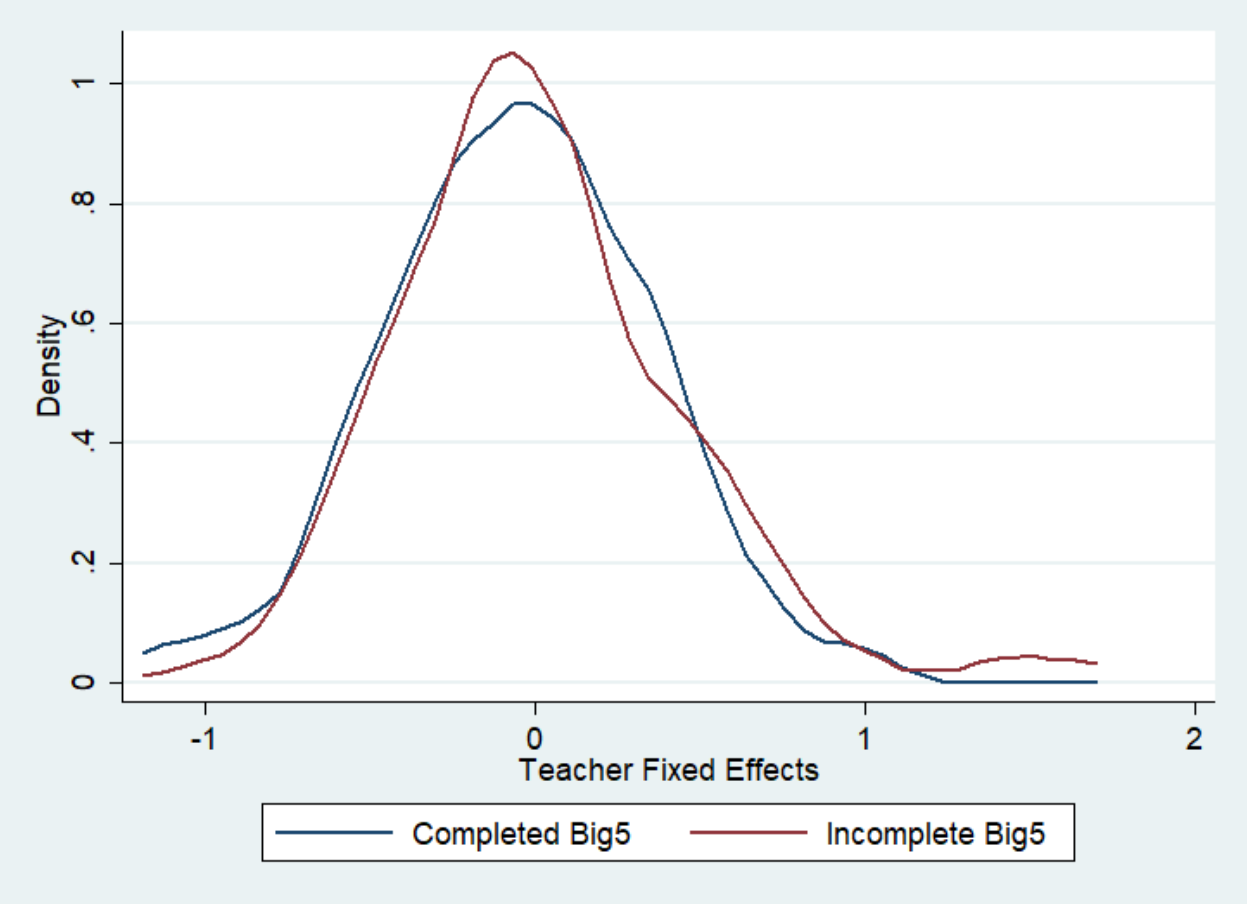
Figure A3.1 Teacher Fixed Effects Density Functions by Survey Status



A rather long question in the teacher survey was the 44-item Big Five personality test. The responses to this test are useful only when all 44 items have been answered, without skipping a single item. Of the 186

teachers who participated in the survey, 145 fully completed the Big Five component (53.1 percent completion rate when considering all 273 teachers). Figure A3.2 shows the estimated kernel densities of teacher fixed effects by Big Five completion status. When we carry out the Kolmogorov-Smirnov test we cannot reject that the two distributions are identical (p-value 0.883).

Figure A3.2 Teacher Fixed Effects Density Functions by Big Five Completion

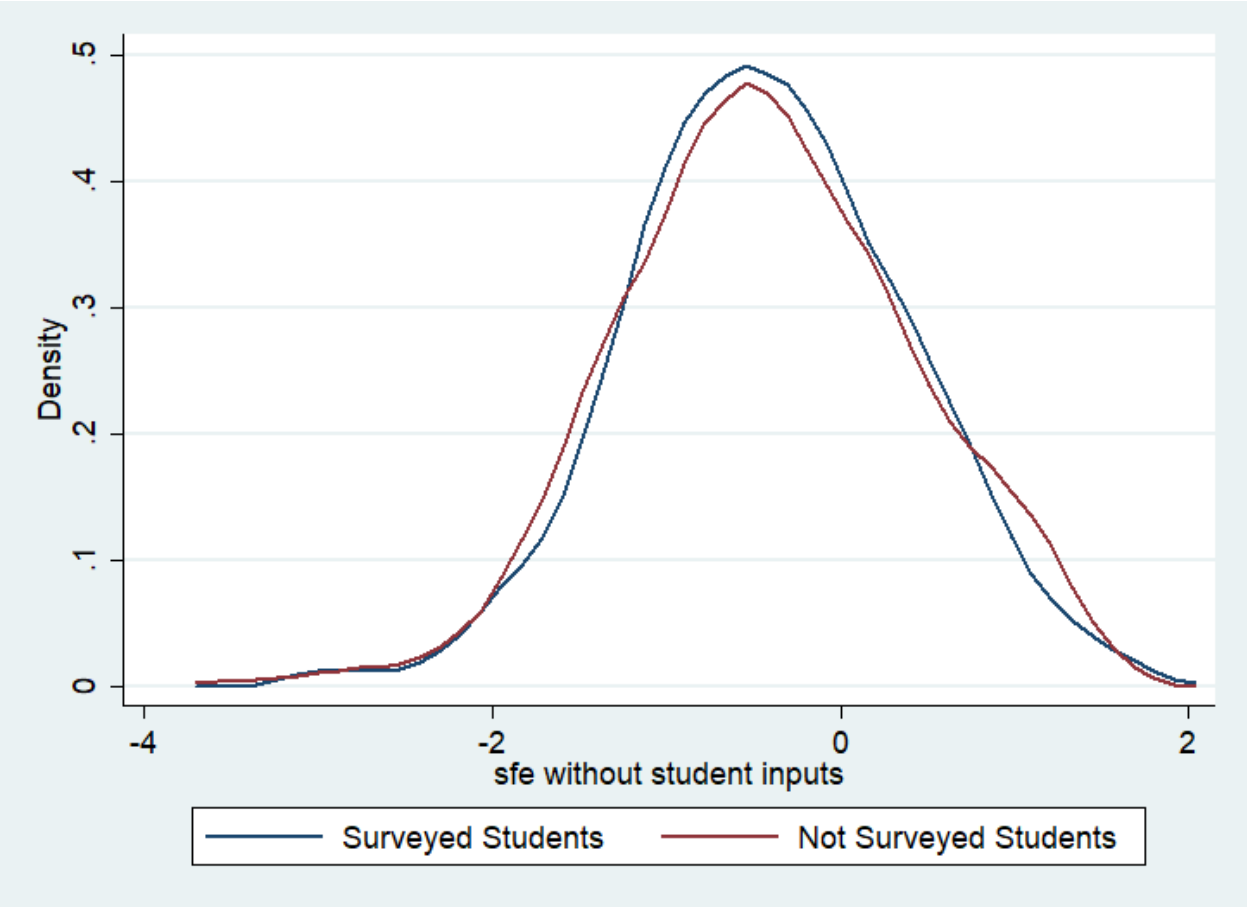


Appendix 4: Student Effects by Surveyed Status

Out of a total of 1,733 students in our analysis sample, 1,190 were surveyed. We surveyed them in school towards the end of the academic year. The survey was administered in class on the days we visited the school. If a student was absent on the day that his/her class was surveyed, or if he/she chose not to fill up the survey, we do not have information on their socio-economic background. This would be a serious limitation if surveyed students were fundamentally different from non-surveyed ones. Here we examine whether there is a difference between the two groups according to estimated student effects. We use student effects estimated from the specification without student inputs as we have values for all 1,733 students in this case.

Figure A4.1 shows the estimated kernel densities of student fixed effects according to surveyed status. The densities almost overlap each other. Further, using the Kolmogorov-Smirnov test for equality of distribution functions we cannot reject that the two distributions are identical (p-value 0.835).

Figure A4.1 Student Fixed Effects Density Functions by Survey Status

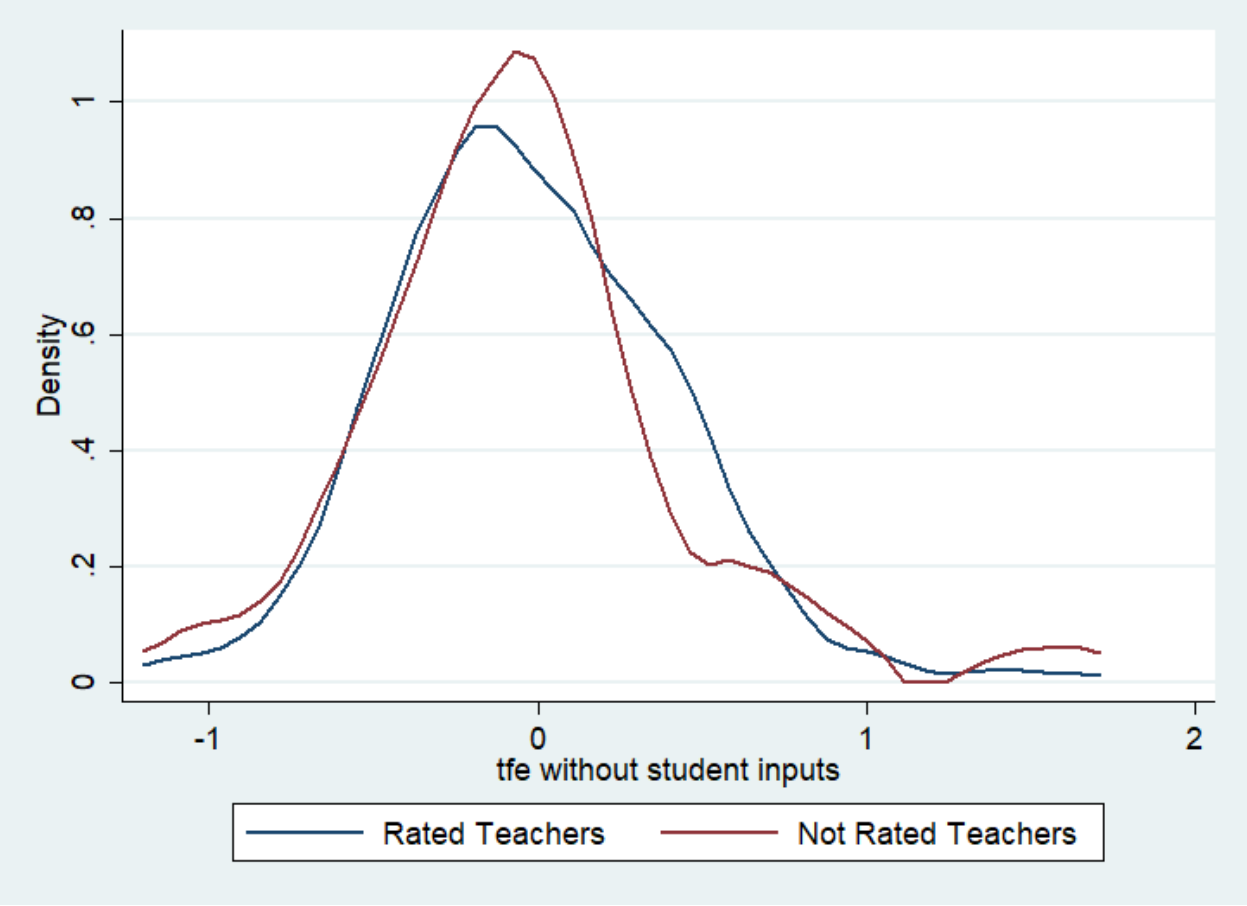


Appendix 5: Teacher Effects by whether or not Principal Rated the Teacher

Out of a total of 196 subject teachers in the 16 schools, principals rated 159 of them. Here we examine whether or not teachers who were rated are more or less effective compared to those who were not.

Figure A5.1 shows the estimated kernel densities of teacher fixed effects (from the specification without controlling for student inputs) by rating status. Using the Kolmogorov-Smirnov test for equality of distribution functions we cannot reject that the two distributions are identical (p-value 0.605).

Figure A5.1 Teacher Fixed Effects Density Functions by Rated Status



Appendix 6: Pairwise Correlations between Multiple Dimensions of Principal Ratings

Principals were required to rate teachers on a scale of 1 (lacking a lot) to 5 (exceptional), along six performance dimensions: dedication and work ethics; increasing students' understanding of the subject;

liked by students; maintaining discipline among students; positive relationship with colleagues; and organizational/administration skills.

Table A6.1 presents the pairwise correlations between these six dimensions.

Table A6.1: Pairwise Correlations between Performance Ratings

	D1	D2	D3	D4	D5	D6
D1	1					
D2	0.560	1				
D3	0.635	0.764	1			
D4	0.545	0.659	0.625	1		
D5	0.602	0.440	0.601	0.592	1	
D6	0.641	0.541	0.587	0.521	0.605	1

D1: Dedication and Work Ethic

D2: Increasing Students' Understanding of the Subject

D3: Liked by Students

D4: Maintaining Discipline Among Students

D5: Positive Relationship with Colleagues

D6: Organizational/Administration Skills

All correlations are significant at the 1 percent level