Assessing Teacher Quality in India

Mehtabul Azam[1]
Oklahoma State University & IZA

Geeta Gandhi Kingdon[2]
IoE, University of London & IZA

Abstract

Using administrative data from linked private schools from one Indian district that matches 8,319 pupils to their subject specific teachers at the senior secondary level, we estimate the importance of individual teachers for student outcomes in the high-stake senior secondary exam (at the end of twelfth-grade) controlling for prior achievement at the secondary level (at the end of tenth-grade). In addition to controlling for prior achievement, we exploit the fact that students took exams in multiple subjects during their senior secondary exam to control for pupil fixed effects. We find a considerable variability in teacher effectiveness over a two year course—a one standard deviation improvement in teacher quality adds 0.366 standard deviation points in students score. Furthermore, consistent with studies in the US, we find that although teacher quality matters, the observed characteristics explain little of the variability in teacher quality.

JEL Codes: I21, O15
Keywords: teacher value-added, teacher quality, student achievement, India

[1] Corresponding author. 326 Business Building, Oklahoma State University, Stillwater, OK, USA. Email: mazam@okstate.edu.
[2] UCL Institute of Education, University College London, 20 Bedford Way, London WC1H 0AL, UK. Email: g.kingdon@ioe.ac.uk

1. Introduction

One of the important debates in education policy has been how to improve educational achievement within schools. In this context, various issues such as reducing class sizes, providing more school inputs, incentive-based policies, or increasing the quality of teachers are routinely discussed. It has been increasingly recognized that one of the most important factors determining student achievement is teacher quality.[3] Research over the past decade in the US confirms that the most important determinant of education quality is teacher quality (Rivkin et al., 2005; Rockoff 2004). Hence, identifying the relative effectiveness of individual teachers is of significant policy relevance as policymakers explore the idea of rewarding individual teachers for good performance, as measured by their ability to raise test scores.

It is natural to ask how one defines a good teacher, or how to recognize a good teacher.[4] In recent times, increasing attention has been focused on the direct relationship between teachers and student outcomes. This outcome-based approach, now commonly called value-added analysis, takes the perspective that a good teacher is simply one who consistently gets higher achievement from students (after controlling for other determinants of student achievement such as family influences and prior teachers) (Hanushek and Rivkin, 2012). Several recent papers in the US have sought to identify and reliably measure teacher value-added (discussed in Section 1.1).

[3]The importance of highly qualified teachers is reflected in the public policy. For example, in the US, there exists government regulations at many levels including standards for highly qualified teachers as mandated by the Federal No Child Left Behind Act and state level licensing requirements. In the Indian context, Minister of State for Human Resource Development, Dr. Shashi Tharoor, stated (on 20th August, 2013) that to improve the quality of school teachers, the Government of India has adopted a three-pronged strategy, which includes (i) the strengthening of Teacher Education Institutions, (ii) the revision of curriculum for teacher education in accordance with the National Curriculum Framework for Teacher Education 2009 and (iii) the laying down of minimum qualifications for Teacher Educators and their continuous professional development.
http://pib.nic.in/newsite/erelease.aspx?relid=98428

[4]No Child Left Behind in the US introduced a requirement for highly qualified teachers within schools serving disadvantaged students. This requirement was phrased in terms of qualifications as opposed to effectiveness in the classroom, and the definitions of highly qualified were left up to the separate states. As a result, most states simply inserted variants of the existing requirements for teacher certification (Hanushek and Rivkin, 2012).

Teacher value-addition, which is a statistical measure of the extent to which a teacher is able to improve student learning during the period of time they are responsible for teaching the concerned student, is also a useful measure of 'gain in student human capital'. Chetty et al. (2014) link 2.5 million children in the US to their adult outcomes to measures of teacher value-added in grades 3 to 8. They find that teacher quality measured by value addition is strongly predictive of adult outcomes including college attendance, quality of college attended, and wages. Teacher quality in school is also positively correlated with social outcomes such as reduced teenage pregnancy and improved quality of neighborhood lived in.

The importance of improving educational attainment within school has been growing in the education policy debate in India, and in developing countries per se, as the focus has been gradually shifting from providing access to education towards providing access to quality education. In the Indian context, the increasing interest in quality of education is partially driven by the realization that the rapid gains in enrollment are not translating into gains in cognitive skills as measured by test scores in reading, writing, or math. Test scores remain low compared to international benchmarks.[5] There exist few studies that examine the impacts of input based policies in India (discussed in Section 1.1), however, we are not aware of any work for India or for any developing country that directly measures teacher value added (TVA). Given the evidence in the US and UK that the teachers play a key role in improvement of student achievement, it is

[5]Two Indian states—Himachal Pradesh and Tamil Nadu—participated in the extended cycle of 2009 OECD PISA (Programme for International Student Assessment) survey of 15-year-olds' knowledge and skills in reading, mathematical and scientific literacy. In the reading literacy score, out of the 74 regions participating in PISA 2009 or (PISA 2009+), these two Indian states beat out only Kyrgyzstan, i.e. ranked 73rd out of 74 countries/regions. In the mathematics score also, out of the 74 regions participating, the two Indian states finished again second and third to last, again beating only Kyrgyzstan. In science literacy, the results were even worse, Himachal Pradesh came in dead last, behind Kyrgyzstan, while Tamil Nadu inched ahead to finish 72nd of 74 (Walker, 2011). What is more worrisome is that these two states are the better states in national rankings within India. Das and Zajonc (2008) used results from standardized math tests based on TIMSS (Trends in Mathematics and Science Study) questions from two Indian states—Orissa and Rajasthan—to create indices on mathematics performance similar to those of TIMSS and found these states to be near the bottom of the global rankings.

important to address how much teachers can play a role in improving student achievement in the Indian context.

In this paper, we use matched administrative panel data on teachers and students from a group of linked private schools from one district in India to delve into the teacher quality question. We focus on the outcome based perspective, and define a good teacher as one who consistently gets higher achievement for students. We seek to find out the teacher effectiveness using scores from twelfth-grade and tenth-grade exams. In our dataset, we observe students taking exams in multiple subjects for their twelfth grade at a point of time, and we also know their prior achievement (in tenth-grade) in those subjects. Students are matched to subject-specific teachers who taught them for two years. Effectively, we observe teachers teaching the same subject in multiple classrooms and over years. We use pupil fixed effects, which enables us to control for the possibility that teachers are not randomly assigned to students.

The findings of the paper are following. Teachers matter a great deal as far as achievement of students is concerned: being taught over a two-year course by a high quality teacher (defined as 75th percentile teacher) rather than a low quality teacher (defined as 25th percentile teacher) adds 0.476 of the standard deviation to the score. Second, there exists a great deal of variation in teachers' quality within-school. Third, although the teacher's quality matters, the observed characteristics of teachers hardly explain any of the variation in the teacher quality. These findings corroborate recent finding in the US and UK.

The findings question the emphasis put on the certification in the hiring and retaining of teachers in India. The factors (qualifications, training, years of experience) rewarded in the status quo are not the ones that matter for teacher quality. Although there is potential to improve achievement through improving average teacher quality, it is not so straightforward as good teachers are hard to identify *ex ante* based on observed teachers' characteristics. In this scenario, *ex post* evaluation of teachers based on their contributions to student achievement or "value added" may be optimal (Gordon et al., 2006). However, this requires a significant improvement and building up of administrative databases that can be used to estimate value added with some confidence. In developing countries, large teacher-matched administrative data sets like those used in the US value-added literature simply do not exist or are not accessible to researchers. The fact that our paper suggests both the importance of teacher quality and the lack of importance of the traditional measures of teacher quality makes a strong argument for such administrative data sets to assess these results on a more general population of students and teachers.

The paper contributes to the existing literature in the following ways. First, the paper is the first study (to our knowledge) in a developing country context to use administrative panel data to estimate TVA directly. In addition, this is only the second paper we are aware of that examines senior secondary school teachers.[6] Second, the paper corroborates the findings of the US and UK in a developing country context, increasing the confidence in those finding irrespective of the context. Third, the paper provides evidence that the emphasis put on teachers' qualifications, training and experience by Indian policymakers in recruiting and rewarding teachers is probably misguided, and the compensation structure can be based on the value-added of a teacher rather than based on teachers' resumè characteristics.

The paper is organized as follows. Section 1.1 presents a brief overview of the literature, Section 2 describes the data, and Section 3 describes the methodology followed in this paper.

[6]Slater et al. (2012) estimate teacher quality at the senior secondary level. The context studied by Aaronson et al. (2007) is high school teachers (ninth-grade) from Chicago public schools. The majority of the papers on teacher quality focus on elementary schools. As argued by Aaronson et al. (2007), although it is important to understand teacher effects at all points in the education process, studying high schools has the additional advantage that classrooms are subject specific. Thus, one can examine student-teacher matches at a level that plausibly corresponds with what one thinks of as a teacher effect. Furthermore, the exams at the secondary and senior secondary levels are high-stake exams.

Section 4 presents the results, and Section 5 concludes.

1.1 Literature

Several papers have examined the relationship between teacher characteristics and achievement directly. Some such studies have used experimental methods, mainly investigating the effect of teacher incentives (Duflo and Hanna, 2005; Glewwe et al., 2010; Muralidharan and Sundararaman, 2011). Other studies have used statistical approaches such as an instrumental variable approach (Hoxby, 1996; Kingdon and Teal, 2007), estimating standard cross-sectional achievement production function, or a panel data approach (Clotfelter et al. 2006, 2010).[7] They find that the variation in students' achievement cannot be predicted by most observable characteristics of teachers (including the factors that are commonly considered to be proxies for quality such as teacher experience, education, and professional training).

However, the effects of specific teacher characteristics cannot be taken as the overall contribution of teachers. The finding in these studies – namely that the commonly used indicators of quality differences are not closely related to achievement gains - led to shifts from a research design that focuses on the link between student outcome and specific teacher characteristics to a research framework that uses a less parametric approach to identify overall teacher contribution to learning as teacher value-added (Hanushek and Rivkin, 2012).

Several recent papers in the US have sought to identify and reliably measure the teacher value-added (TVA) directly. Using mainly administrative data from US schools, they have sought to measure teacher quality by the teacher fixed effect in a student achievement equation where a teacher is matched to students in the various classes of a given grade she/he taught in a year or the

[7]Hanushek (2003) provides a review of the US and international evidence on the effectiveness of input based policies.

cohorts she/he taught over various years. Rockoff (2004) uses panel data from two school districts in New Jersey over the years 1989-90 to 2000-01 covering grades 2-6 to estimate teacher ‘fixed effects’ while controlling for fixed student characteristics and classroom specific variables. He finds large and statistically significant differences in effectiveness among teachers. Hanushek et al. (2005) look at the market for teacher quality by using matched panel data on students and teachers in grade 4 through grade 8 for the school years 1995-96 to 2000-01 from a large district in Texas, to estimate variations in teacher quality. Their estimates confirm the existence of substantial variation in teacher effectiveness within schools, and they argue that this within-school heterogeneity has direct implications for the design of teacher accountability and teacher incentive programs. Rivkin et al. (2005) use data on three cohorts spanning grade 3 to grade 7, and over a half million students across 3000 schools in Texas. Their data does not match students with individual teachers, and they use grade average information on teachers. They give a lower bound estimate of standard deviation in teacher quality of 0.11 in math and 0.095 in English. Aaronson et al. (2006) use unique administrative data on Chicago public high school students and their teachers to estimate the importance of teachers in determining students' mathematical achievement. They find that teachers are educationally and statistically important.

Outside the United States, relatively little research has been carried out on the measurement of teacher effectiveness. A recent work on the UK by Slater et al. (2012) links 7,305 pupils to the individual teachers who taught them, in each of their compulsory subjects in the high-stakes GCSE exams at age 16. They find considerable variability in teacher effectiveness, a little higher than the estimates found in the few US studies. Similarly using administrative panel data from the state of Queensland, a state in Australia, Leigh (2010) finds large variation in teacher effectiveness: moving from a teacher at the 25th percentile to a teacher at the 75th percentile would raise student test scores by one-seventh of a standard deviation.

Two common findings from these papers are that teacher quality matters and that the observed characteristics of the teachers—their pay, education, training and experience—explain little of (the measures obtained of) teacher effectiveness. These findings are clearly of importance for policy. If teachers do matter—something that parents have always believed—and good teachers are hard to identify, then new thinking is required on how good teachers can be identified and rewarded. The factors that are rewarded in the status quo may not be the ones that matter for teacher quality.

Although, there is no work in India that directly measures teacher quality/ effectiveness, as measured by ability to raise test scores, the debate on how to raise the student achievement is ongoing in India, which mirrors the wider debate in the work on developed countries. Most studies in India examine the impacts of input based policies using sample data. For example, Kingdon (2006) examines the effect of teacher characteristics on pupil learning using the standard cross-section achievement production function while allowing for pupil fixed effects. She uses the test scores of tenth-grade students in different subjects from a sample of 183 schools, and supplemental information on students, their teachers and their principals. As she does not know the exact teacher who taught the student, she assigns the average characteristics of all teachers in the school that teach a given subject to grade 10, to all students of grade 10, for that subject. Thus, in Kingdon (2006) a student is matched to an average of grade 10 teachers' characteristics in the school. Moreover, she also does not have subject-specific prior achievement information, and her estimated model is thus in levels, not *value-added*. She finds that 'Masters level or higher' qualification and 'possession of pre service teacher training' both raise pupil achievement by 0.09 standard deviations. She suggests that these are upper bound estimates. Using the same data and similar empirical strategy as Kingdon (2006), Kingdon and Teal (2010) find that union membership of a teacher reduces student achievement and increases the salary costs in private schools in India. Similar to Kingdon (2006), Kingdon and Teal (2010) also do not control for prior achievement. Rawal and Kingdon (2010) explore whether having a teacher of the same caste, religion, or gender influences student outcomes in the sample primary schools. Similarly, Muralidharan and Sheth (2015) examine whether having a same gender teacher impacts student outcomes in sampled government-run primary schools in the Indian State of Andhra Pradesh. Using a large scale randomized evaluation of group and individual teacher performance pay programs implemented across a large representative sample of government-run rural primary schools in the Indian state of Andhra Pradesh, Muralidharan (2012) and Muralidharan and Sundararaman (2011) examine the impact of teachers incentives on student outcomes. Muralidharan and Sundararaman (2013) provide experimental evidence and Atherton and Kingdon (2010) provide school and pupil fixed effects evidence on the relative effectiveness of regular and contractual teachers.

2. Data

Upon agreement with a private school consortium that administer a group of linked schools in one of the districts in the north state of Uttar Pradesh, the consortium provided us administrative records of all students who took the twelfth-grade board exams between 2006 and 2010 (5 cohorts) in its ten schools.[8] These are English Medium schools, and each of the ten schools has multiple sections (classes) for twelfth-grade.

[8] According to the State Census 2011, Uttar Pradesh is the most populated state in India with a population of 199.6 million.

In each year, these ten schools together sent (an average of) around 1680 students to the grade twelfth exam, i.e. a mean of 168 students per school per year. Since each class has around 40 students, each school had an average of 4 classes of twelfth grade. The exam board these schools are affiliated to is the Council for the Indian School Certificate Examinations (CISCE) based at New Delhi, which is the former Cambridge Examination board which the British operated until 1947. It is a national exam board with around 2000 affiliated schools, and is smaller than the only other national exam board in India (the Central Board of Secondary Education, CBSE) which has about 13,000 schools affiliated to it. While both English- and Hindi-medium schools can affiliate to the CBSE board, only English Medium schools can affiliate to the CISCE board. Our 10 sample schools are fairly typical of English-medium secondary schools in India in terms of size, socio-economic characteristics of the student body, infrastructure, etc., when compared with data from a sample of 183 CISCE-affiliated schools from 16 states of India used in Kingdon and Teal (2010). While our sample of schools is small, and these schools are not representative of all Indian secondary schools (a high proportion of which are state-funded regional-language medium schools), they nevertheless represent an important part of the secondary school segment in India.

The administrative records provided by the ten schools include subject-wise performance of each student in the twelfth-grade board exam. In addition, the records also contain subject-wise performance of each student in the tenth-grade board exam. The twelfth-grade (known as Indian School Certificate, ISC) exams are typically taken at the age 17-18, and are considered very important, as universities and colleges in India use these scores for admissions into higher education. The tenth-grade (known as Indian Certificate of Secondary Education, ICSE) exams are taken at the age of 15-16.[9] Unlike exams in other grades which are typically set and graded within

[9] Both ISC and ICSE are English Medium exam boards.

schools by teachers who teach them, twelfth and tenth-grade exams are nationally set and marked outside the school, leaving little scope for manipulation. Both twelfth and tenth-grade exams are high stakes exams.[10,11]

Most importantly, the administrative records also link each classroom to subject-specific teachers who taught the particular subject to the class for two years (during 11th and 12th grade). Student-teacher data are matched in and by the school, thus ensuring a high-quality match. The administrative records also contain specifics about human capital and demographics of the teachers. However, the records do not contain any student characteristics.

The data contain the twelfth grade board exam results for 8,382 pupils. We dropped 63 pupils for whom the tenth-grade score was nor reported for any subject. Typically, a student appears for a minimum of 5 subjects and a maximum of 7 subjects for qualifying at ISC (12th grade). A student has to attempt the subjects of English and other subjects of the student' choosing, however, the choices are restricted by schools offering only a limited number of these subjects. In our sample 50 percent of the students appeared in 6 subjects, while about 25 percent each appeared in 5 or 7 subjects in grade 12. Similarly, a minimum of 7 subjects should be taken to qualify for ICSE (10th grade). Although, it is not necessary that the subjects taken by a student in 10th grade is almost same as taken in 12th grade, there is a great deal of overlap. As a result, we have information on prior achievement on most of the subjects taken by the student in grade 12. Essentially, an observation in our data is a pupil-teacher match, or equivalently a pupil-subject-teacher match as each teacher only teaches one subject. In case, if we do not have a subject specific 10th grade score

[10]The Indian School Certificate (ISC) is an examination conducted by the Council for the Indian School Certificate Examinations for Grade 12, i.e., year 12. A student usually attempts this examination after first completing the Indian Certificate of Secondary Education in Grade 10, although the completion of a recognized equivalent level of education is normally sufficient.

[11]Clotfelter et al. (2010) emphasize the importance of test being external to school, related to the material the teachers are hired to teach, and that the students are likely to take seriously.

for a particular student, that subject (observation) is dropped for that student (see appendix Table A.1 for details). In our final sample, the average number subjects for which both 12th and 10th grades scores are reported for a student is 4.4, while both the scores are reported for at least four subjects for more than 90 percent of the twelfth-grade students. The data used in the initial regression contain 38,228 pupil-subject-teacher (or pupil-teacher) matches. There are 191 teachers in the dataset, and median (average) number of classrooms observed per teacher is 5 (6.8).[12] Table 1 presents the descriptive statistics. The average age of teachers in our data is 41 years, while about half of the teachers are female. 85 percent of teachers holds a masters degree or higher, while 59 percent of the teachers have received Bachelor of Education (B.Ed) training.[13]

All subjects are marked out of 100, so given marks may be interpreted as percentages. In order to render the marks in different subjects and years comparable, we standardize the marks in each subject by year, i.e., we use the z-score of achievement as our dependent variable. The z-score is the student marks in a subject in a year less the average marks in that subject in that year, divided by the standard deviation of marks in that subject in that year. Thus, by construction, mean of the z-score in any given subject in a year is zero and its standard deviation is 1. The normalization implies that the estimated coefficient can be interpreted as a fraction of the standard deviation (SD).

Although our data is smaller in comparison with the administrative datasets used in some of the studies in the US, for a developing country it is unusual, and complements the datasets from developed countries used by Aaronson et al. (2007), Clotfelter et al. (2006, 2007, 2010), Rivkin et al. (2005), Rockoff (2004), Kane et al. (2008), and Slater et al. (2012). Like Aaronson et al. (2007),

[12]Observed characteristics are missing for 3 teachers.

[13]Bachelor of Education (B.Ed) is a one-year course offered for those interested in pursuing career in teaching. Many schools in India make the B.Ed degree mandatory for teaching in higher primary schools and high schools. The minimum qualification required for entry into B.Ed course is a bachelor degree.

Rockoff (2004) and Slater et al. (2012), but unlike Rivkin et al. (2005) and Kane et al. (2008), we can match a student to her/his actual teacher, rather than to the school-grade average teacher. Unlike Aaronson et al. (2007), Clotfelter et al. (2006, 2007, 2010), Rivkin et al. (2005), Rockoff (2004), Kane et al. (2008), and like Slater et al. (2012) our context is students taking terminal exams that are very important to them and to the school. Similar to Clotfelter et al. (2010) and Slater et al. (2012), we exploit the fact that we observe students taking exams in multiple subjects during their twelfth-grade or tenth-grade exams, allowing us to use pupil fixed effects and thus use within-student between-subject (between-teacher) variation, in addition to the subject-specific exam scores that capture subject-specific prior attainment. We believe that this allows us to control well for variations in student ability that might otherwise bias measures of teacher effectiveness if students are not randomly assigned to teachers.

3. Empirical Methodology

In our data, we observe students taking exams in multiple subjects at the same time (for twelfth-grade exam). In addition, we also have information on scores obtained in those subjects two years previously (for tenth-grade exam). We modify the traditional value-added equation:[14]

$$Y_{izjkt}^{12} = (1-\lambda)Y_{iz}^{10} + \tau_j T_j + \theta_i + \mu_t + \rho_k + \delta_z + \epsilon_{izjkt} \tag{4}$$

where Y_{izjkt}^{12} refers to the i^{th} student twelfth-grade score in subject z, taught by teacher j at school k at time t (here, t refers the cohort taking twelfth grade exam at time t), Y_{iz}^{10} is i^{th} student's tenth-grade score in subject z, and δ_z are subject dummies. T_j are dummies for teachers, hence each element τ_j refers to the effects of two year spent with teacher j, and thus

[14] Todd and Wolpin (2003) provide a discussion on the value-added model. Slater et al. (2012) implement a similar strategy as ours.

identifies the quality of the teacher. The variables θ_i, ρ_k, and μ_t measure the time-invariant characteristics of the student, the time invariant characteristics of the schools, and any secular temporal change in test performance, respectively, while ϵ_{ijkt} is the white noise.

In studies using state administrative data at the elementary level, researchers have addressed the problem of non-randomness in the assignment of students and teachers to classrooms by including pupil fixed effects in the model as their longitudinal data includes outcome measures, such as test scores in math, for each student across multiple years (Clotfelter et al., 2010). The pupil fixed effects control statistically for unobservable time-invariant characteristics of students—such as their ability or motivation—that could be correlated with the teacher.

Since we have scores for multiple subjects at the same point of time, we also address the sorting problem with the use of pupil fixed effects, which means that we identify the relevant coefficients on the T_j variable based on the within-student between-subjects variation i.e. they are based on the fact that different subjects are taught by different teachers.[15] Our identification of effects of T_j variable is analogous to Clotfelter et al's (2010) identification of effects of teachers' credentials. However, unlike Clotfelter et al's (2010) model that is in levels, our model captures value-added.[16]

Although pupil fixed effects potentially capture all the observed and (subject-invariant) unobserved factors, it fails to take account of any changes that have occurred over time (in this case, between grade ten and grade twelve). This is similar to traditional fixed effects (within-subject across time) that do not capture the changes in unobserved factors, and assume that

---

[15] However, similar to many administrative data, our data lacks students' characteristics (X_i) is not included in equation (4)).

[16] Clotfelter et al. (2010) do not control for subject-specific prior test score.

unobserved factors have same impact over time. In addition, most of the administrative data lack an extensive set of variables to fully account for changes in observed factors. The advantage of pupil fixed effects is that it can potentially account for time varying observed and unobserved factors under the assumption that they remain same across subjects. Another but related issue of concern is that of subject-specific ability (which is analogous to time-varying unobserved factors in across-time estimations). For example, a subject-specific high ability (or low ability) student matches up with a subject-specific high ability teacher, which might be a case when a school has more than one subject-specific teacher, as in our case. However, prior achievement can also work as a good proxy for subject-specific ability, and bias if any, should be minimal.

We observe teachers linked to students over a five year time period.[17] However, all teachers remain in the same school over time. This implies that it is impossible to separately identify a pure teacher effect and a school effect. School fixed effects in equation (4) controls for time-invariant school characteristics that co-vary with individual teacher quality. Hence the reported variance in the estimated values of τ_j is within-school variation in τ_j, i.e., the variance of $(\tau_j - \bar{\tau}_{j(k)})$. This provides a lower bound to the degree of variation. If schools hired teachers randomly, then this measure would reflect the true overall variation in teachers' effectiveness. But if good teachers cluster together and bad teachers clusters together, then the within school variance will be lower than the true overall variation (Slater et al., 2012).

Our identification of teacher effects arises from comparing the exam score progress of a student taught different subjects by different teachers over the same 2-year period.[18] As noted by

---

[17]Observing teachers over multiple years allows us to distinguish permanent teacher quality from idiosyncratic class-year shocks.

[18]Rivin et al. (2005), Rockoff (2004), and Aaronson et al. (2007) use exam scores that are administered annually. However, similar to Slater et al. (2012), our exam scores are separated by two-year period. Nevertheless, the same subject specific teacher teaches the class for the two years. In contrast, Leigh (2010) who also uses scores from biennially exam at the elementary school level allows for different teachers teaching each of the two years. He

Slater et al. (2011), this controls for all general attributes of the student at one point of time: intelligence, effort, motivation, imagination, and ability to learn, and it also conditions on subject-specific ability as measured by the tenth-grade score.

Across-subject differencing has an important methodological advantage over across-time differencing, namely that the former approach does not suffer from the problem of non-random attrition of teachers and students over time that can occur 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 teachers. Rivkin et al (2005) also 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.

Despite the above advantage of the across-subject pupil fixed effects approach, it is important to assess any remaining sources of bias with this approach that could threaten a causal interpretation. The main problem with identification of the teacher fixed effect is the possibility that subject-varying pupil unobservables remain in the error term and may be correlated with the teacher variable (e.g. students who are abler in a particular subject may systematically match with teachers who are stronger in that subject). However, our inclusion of prior achievement in the

estimates the value added either by ignoring the intervening year altogether, or by creating an assumed test score in the intervening year which lies at the midpoint of the other two tests. Slater et al. (2012) weigh each pupil-teacher observation by $1/n$ if a student has n teachers in the subject.

subject (at grade 10 level) means that the only thing remaining to be worried about is the possible correlation (with the teacher dummy variable) of any *change in subject-specific unobservables* over the two year period between grade 10 and grade 12, for example, if a student becomes brighter or more motivated in math during the 2 year period between grade 10 and 12 and is re-assigned (part-way between these two years) to be taught math by a math teacher who is known to be more effective in math teaching.

It seems unlikely that students' subject-varying unobservables could be correlated with the included teacher dummy variable since in the sample schools, there is no policy of assigning or re-assigning pupils or teachers to particular classes on the basis of their strength in given subjects. We believe that the inclusion of tenth-grade score and pupil fixed effects addresses the non-random sorting of students. However, under the theoretical scenario that students are assigned to teachers based on expected progress in a given subject relative to expected progress in other subjects, we would falsely attribute more of test score progress to teachers rather than students, which would bias our measures of teacher effectiveness upwards.[19] However, this seems unlikely in our case. In addition, we assumed that prior achievement will have a linear effect on students' future relative gains. In case of violation of this assumption, our teacher effects may be biased. Furthermore, we assume that teachers have no effect on the results of subjects other than their own. Any violation of this assumption will introduce downward bias in our teacher effects. However, these problems are faced by all econometric models of teacher effects (Slater et al., 2012).

[19]Rothstein (2010) notes that if students are dynamically assigned to teachers on the basis of prior unobserved shocks to student achievement and these shocks are serially correlated, then controlling for observable student characteristics or even adjusting for unobserved time-invariant student heterogeneity via student fixed effects, will not be sufficient to produce unbiased teacher effects. Using data from a single cohort of students in North Carolina, Rothstein uncovers evidence of future teacher "effects" on current achievement, suggesting value-added measures of teacher performance are indeed biased. However, Koedel and Betts (2011) find that dynamic sorting of student and teachers to classrooms is transitory and that observing teachers over multiple time periods (as we do, over 5 time periods, in the current paper) mitigates the dynamic sorting bias envisioned by Rothstein. Further, Kane and Staiger (2008), compare experimental evidence on the impacts of randomly assigned pairs of teachers to their pre-experimental value-added scores and fail to reject the equivalence of the two measures for a number of value-added models.

3.1 What explains the variation in teacher effectiveness?

As teachers' credentials are important policy levers, it is important to know how these credentials are related with the teachers' effectiveness. We have information on teachers' age, gender, teaching experience, educational qualification (whether the teacher holds a master degree), and professional training received (whether the teacher holds Bachelor of Education, B.Ed). Hence, we explore whether the observed characteristics have any explanatory power of estimated teacher effectiveness, $\hat{\tau}_j$, which we obtained using equation (4). That is, we estimate the following equation:

$$\hat{\tau}_j = \pi Z_j + u_j \tag{5}$$

where Z_j is the j^{th} teacher's characteristics.[20]

3.2 Sampling Variation

As argued by Kane and Staiger (2002), Rockoff (2004), and Aaronson et al. (2007), the variance in estimated teacher effects ($\hat{\tau}_j$) will overestimate the variation in true teacher effects as the variation in estimated teacher effects will include the sampling variation in addition to variation in true teacher effects. They show that the importance of sampling variation declines as more students are used to estimate the teacher fixed effects. To address the problem of sampling error, we have only included those teachers in our analysis who have taught at least 15 students. In addition, following Aaronson et al. (2007), we analytically adjust the variance of estimated teacher fixed effects, $\hat{\tau}_j$. Aaronson et al. (2007) assume that the variance of estimated teacher effects has two components—the true variance of teacher effects and average sampling variance, and use the mean of the square of the standard error estimates of τ as an estimate of sampling error variance

[20]We take an average of teacher's experience, which varies over time.

and subtract this from the observed variance of $\hat{\tau}_j$to get the adjusted variance, i.e., $var(\tau_j) = var(\hat{\tau}_j) - mean(se(\hat{\tau}_j)^2)$.

4. Results

First, we begin with a naïve model which includes the tenth-grade standardized score, subject dummies, year dummies, and teacher dummies as explanatory variables, while twelfth-grade score is the dependent variable.[21,22] The results are presented in Table 2. The importance of fixed teacher quality can be measured by the variation in teacher fixed effects (Rockoff, 2004). For example, one might measure the expected rise in the test score for moving up one standard deviation of teacher fixed effects. First, we find that teacher fixed effects are jointly highly significant in explaining student achievement. Second, we also find a great deal of variation in estimated teacher fixed effect. Standard deviation of estimated teacher fixed effects, $\hat{\tau}_j$, is 0.513, which is quite broad. Aaronson et al. (2007) find a standard deviation of 0.15 in teacher fixed effects using a similar model and Chicago public high school data.[23] The adjustment in sampling error reduces the standard deviation in teacher fixed effects marginally to 0.490. The adjusted standard deviation suggests that teacher quality has a large impact on student achievement. Thus a teacher who is one standard deviation above the mean of the distribution of teachers in terms of quality (i.e., roughly comparing the 84th-percentile teacher to the 50th-percentile teacher) is estimated to produce marginal learning gains of about 0.5 standard deviations of student achievement above the average teacher. In terms of the student achievement distribution, this

[21]One can argue that the tenth-grade test score may serve as good proxy for θ_i. As suggested by Guarino et al. (2012), even if, technically speaking the estimates are inconsistent, it could nevertheless can provide relatively accurate estimates for teacher fixed effects. However, we are not arguing that controlling for tenth-grade score eliminates omitted variable bias, and in later models, we control for individual fixed effects.

[22] Inclusion of teacher fixed effects control for within-teacher correlation (as we believe that is solely driven by common teaching process), but not the additional within-student correlation, which is accounted by clustering at student level.

[23]They reports 0.4 standard deviation in terms grade equivalents (the standard deviation of ninth-grade is 2.71).

would move a student from the 50th percentile to the 70th percentile.

What is more interesting is the gap in value addition between a 90th percentile teacher (a very good teacher) and a 10th percentile teacher (a bad teacher). Thus a student who is taught by a 90th percentile teacher scores 1.271 standard deviations more than a student who is taught by a 10th percentile teacher. Column (2) of Table 2 presents the estimated teacher fixed effects weighted by the number of students taught by each teacher. Weighted standard deviation in teacher fixed effects drop to 0.453 from the un-weighted standard deviation of 0.513. Weighting the teacher fixed effects with the number of students taught by the teacher reduces the dispersion in teacher fixed effects marginally but overall conclusions remain similar.

The parsimonious model presented above may not fully capture the heterogeneity in students and family backgrounds. To capture those, we introduce pupil fixed effects.[24] Table 3 presents the results. For comparison purposes, column (1) of Table 3, repeats the results of parsimonious model presented in Table 2. Introduction of pupil fixed effects makes a huge difference, and reduces the standard deviation in estimated teacher fixed effects, $\hat{\tau}$, from 0.513 to 0.379 (column (2) of Table 3). Further adjustment of sampling variation reduces the standard deviation marginally to 0.366 from 0.379. Thus the amount of estimation error is small in the model with the pupil fixed effects. Slater et al. (2012) also find that estimation error is much greater in a model that uses the pupil characteristics than the model which use pupil fixed effects, suggesting that pupil fixed effects model is much more precise in estimating the teacher fixed effects. The teacher fixed effect estimates from model (2) that controls for the pupil fixed effects suggest that being taught by a teacher who is one standard deviation above an average teacher

[24]The STATA routine a2reg for two way fixed effects (Ouazad, 2008) is used. The standard errors of estimated teacher fixed effects used for adjustment of variance in teacher fixed effects (calculation of estimation error) are derived through bootstrapping with 100 replications.

increases the pupil achievement by 0.366 of the standard deviation, which moves an average student from the 50th to about 65th percentile of the score distribution. Slater et al. using a similar model find that the standard deviation of teacher effects is 0.610 GCSE (General Certificate of Secondary Education in UK), which is about 0.358 of the standard deviation of the score.[25] The interquartile range (IQR, 75th-25th percentile) of estimated teacher effects is 0.456 of the standard deviation. This is consistent with IQR of 0.440 of the standard deviation found by Slater et al. in UK.[26] IQR is the gain per pupil from having a good teacher (defined as being at the 75th percentile) relative to a poor teacher (defined as being at the 25th percentile). The gain per pupil from having a very good teacher (as defined as being at the 90th percentile) relative to a very bad teacher (as defined as being at the 10th percentile) is 0.934 of the standard deviation. Slater et al. report 95th-5th gap of 1.18 of the standard deviation. Thus our estimates are much in line with the estimates reported by Slater et al. for the UK (2012).

Next, we introduce school fixed effects in the model (column (3) of Table 3). Unsurprisingly, introduction of school fixed effects makes a little difference to the standard deviation of teacher fixed effects, however, there are some marginal changes in gains when teachers at different percentiles are compared.[27] As discussed in the empirical methodology section, model (3) provides us within school variation in teacher effectiveness, i.e. the variation in teacher effectiveness when compared with an average teacher in the school.

Hanushek and Rivkin (2012) helpfully summarize the results of many studies in the US (Table 1 of Hanushek and Rivkin, 2012) in terms of standard deviations of teacher effectiveness. They report an average standard deviation of 0.13 in teacher effectiveness for students' reading

[25]They report standard deviation of 1.705 in GCSE score.
[26]They report an IQR of 0.75 in terms of GCSE points (standard deviation of GCSE is 1.705).
[27]As school effects are already incorporated in the teacher fixed effects estimated without school fixed effects, introduction of school fixed effects removes the mean of teacher fixed effects (a constant) within each school from the teacher fixed effects. In addition, our sample consists of a limited number of schools.

score, while an average standard deviation of 0.17 in teacher effectiveness for students' math score. They report a highest (lowest) standard deviation of 0.18 (0.07) in teacher effectiveness for reading score, as found by Kaine and Staiger (2008) (Nye et al., 2004) using Los Angeles (Tennessee) data. They also report a highest standard deviation of 0.26 in teacher effectiveness for math score found by Jacob and Lefgran (2008) using a Midwest city data, while a lowest standard deviation of 0.11 in teacher effectiveness found by Rockoff (2004); Rivkin et al. (2005); and Hanushek and Rivkin (2010).[28]

In comparison to the US studies, our estimate of standard deviation of 0.379 (or adjusted standard deviation of 0.366) in teacher effectiveness seems much larger. However, while the US studies estimates of teacher effectiveness are based on one year spent with the teacher, our estimates of teacher effectiveness are based on two years spent with the teacher. Thus, our estimates are value addition over a 2-year course, and these estimates are about twice (or more) as high as the estimates from the US for annual progress. Our study is much closer to Slater, Davies, and Burgess (2012) for the UK in terms of context, empirical strategy, and the duration of value added. They found a standard deviation of 0.233 in teacher effectiveness within-school, which is smaller than our estimate of 0.366. Thus the variation in teacher effectiveness in India, a developing country, is larger than what is estimated in the UK, and the US.

4.1 What explains teacher effectiveness?

We have found that teacher quality is a very important determinant of achievement. However, how a teacher's quality is related to the teacher's credentials (education, training, experience, etc.) is also an important issue. For example, a credential-related policy lever might be used to raise the overall quality of teachers and to ensure an equitable distribution of high-quality teachers across

[28]These estimates are adjusted for sampling variation.

schools and classrooms (Clotfelter et al., 2010). In addition, understanding the factors that affect teacher productivity and the degree to which these determinants are measurable would also inform current policy debates over how best to evaluate and compensate teachers.

To delve into this issue further, we relate the estimated teacher fixed effects, $\hat{\tau}_j$, to measurable characteristics of the teachers available in the school administrative data. Table 4 reports the results. The dependent variable, teacher fixed effects, are the same teacher fixed effects as reported in the earlier section (Table 3). The amount of variation in teacher quality (as measured by the teacher fixed effect) explained by the teacher characteristics is very low.[29] Column (1) of Table 4 reports the results which use teacher fixed effects estimated from model (1) of Table 2 with no pupil fixed effects. Only one of the six characteristics (teacher has a master degree) is statistically significant at conventional level. Thus having a teacher who holds a master or higher degree compared with having a teacher with a bachelor degree increases the achievement by 0.353 of the standard deviation, which is a big impact. In addition, teacher having a teacher training increases student achievement by 0.14 Standard deviations, though this is significant only at 10 percent level. Since in model (1) we are not controlling for pupil fixed effects, some of those impacts might be driven by positive non-random matching.

Using the teacher fixed effects from model (2) and (3) that controls for pupil fixed effects, none of the teacher characteristics remain statistically significant (column (2) & (3) of Table 4). Moreover, except for the master's degree all other coefficients are close to zero. Although, the teachers' master's degree is not statistically significant, it suggest that having a master degree increases the students' achievement by 0.11 SD. This estimates are comparable to Kingdon (2006)

---

[29]As argued by Aaronson et al. (2007), the reported R^2 is an understatement of the explanatory power since some variation in teacher fixed effects is due to sampling.

who took a direct estimation approach and finds that `masters level or higher' qualification raise pupil achievement by 0.09 standard deviations. She suggests that these are upper bound estimates. In our sample, the coefficient on master's degree is based on a very small number of teachers (85 percent of teachers in our sample have master's degree, hence there is not much variation in the data regarding teachers' educational qualifications), and hence the imprecise estimate. However, given the magnitude of the impact, this may a relevant policy measure and hence needs to be explored with a data set with a larger variation in educational qualifications. The magnitude of the impact of the master degree is comparable to the impact found by Muralidharan and Sundararaman (2011) for providing additional input at elementary level. They provided an extra contract teacher and a cash block grant for school materials to randomly chosen elementary schools in the Indian state of Andhra Pradesh. At the end of 2 years, students in schools receiving the input programs scored 0.08 SD higher than those in comparison schools.

A few studies in the US suggest that the first few years of teaching experience seem to matter to student achievement. For example, Clotfelter et al. (2010) find that all of the gains in achievement associated with teacher experience occur in the first five years of teaching. Our estimates suggest (column (3) of Table 4) that experience does not matter. In column (4) of Table 4, we allow for nonlinearity in experience by specifying years of experience as a series of indicator variables. Given the limited number of teachers in our sample, we divided experience into four categories, with each category consisting of about 25 percent of teachers. None of the experience indicators are statistically significant. In addition we fail to reject the null that experience indicators are jointly insignificant.

Overall, the variation in teacher quality explained by teachers' characteristics is extremely low. Thus the observed factors (observed in the data) explain very little of the teacher quality variation, suggesting that it is unobserved factors such as perhaps drive, passion, connection with the students, empathy, attitude to effort, communication skills and so forth, that account for much of the variation in teacher effectiveness. The lack of explanatory power of human capital regressors and their lack of association with teacher quality is consistent with the studies in the US (Aaronson et al., 2007; Rivkin et al., 2005) and the UK (Slater et al., 2012).

4.2 Robustness

It is natural to ask whether our across-subject (i.e. pupil fixed effects) results could be driven by differences in the distribution of marks across subjects, despite using z-score. As our identification strategy exploits within-pupil between-subjects variation, using more subjects creates more variation. As pointed out in the data section (and appendix Table A1), in our sample construction, we dropped the subject for a student if no grade 10 score is reported for that subject, as a result we lose pupil-subject observation for some pupils who appeared in that subject in grade 12 but not in grade 10. To find out how losing an observation because of non-availability of the subject score in grade 10 affects the variation in teacher effectiveness, we restrict our sample to six subjects (English, Math, Physics, Chemistry, Biology, and Computer sciences). The students who appeared for these subjects in grade 12, also appeared for these subjects in grade 10. Row (2) of Table 5 report the standard deviation in teacher effectiveness, while for comparison purposes row (1) of Table 5 reports standard deviation in teacher effectiveness from entire sample as reported in Table 4. The adjusted standard deviation drops marginally from 0.366 to 0.350. Row (3) to row (12) of Table 5 reports the standard deviations of estimated teacher fixed effects estimated by dropping one subject from the sample each time. Dropping one subject from the sample does not reduce number of rows for all students except when the dropped subject is English. As English is mandatory for everyone, dropping English implies, number of rows for each pupil decreases by one. However, for other subjects decline in number of rows is applicable only to those who appeared in that subject in grade 12. Although there is some variation in the standard deviations in teacher fixed effects estimated using different subject-sub-samples, the difference is marginal when compared to standard deviation estimated using the full sample. The results of Table 5 increase the confidence that the standard deviation estimated from the full sample is not driven by the inclusion or exclusion of any one of the subjects.

5. Conclusion

In this paper, we use administrative data provided by a group of linked private secondary schools from one of the districts in the state of Uttar Pradesh, India to address the issue of teacher effectiveness, as defined by the value added. The data provides us information on subject specific scores obtained by twelfth grade students during the high stake Indian School Certificate (ISC) exam held at the end of grade twelve for cohorts taking the exam during 2006-2010. The data also provide us the scores for the same subjects obtained by the students during the Indian Certificate of Secondary Education at the end of grade ten. Furthermore, the data links the 8319 pupils to their subject specific teachers who taught them during the two years (grade eleven and grade twelve).

We address the issue of non-randomness in matching of students with teachers through controlling for prior achievement and pupil fixed effects, and estimate the value added of teacher based on the two year time spent with the student. As found in many studies in the US, we also find considerable variation in teacher effectiveness, thus confirm the findings of developed countries in an underdeveloped country settings. The standard deviation of teacher effects in India is 0.379 which is marginally more than twice of the average standard deviations reported in the US studies. However, our teacher effects capture the impact of spending two year with the teacher, while the US literature reports impacts of one year spent with the teacher.

Our findings about the importance of teacher quality for the high stakes exams suggest that students' family background is not the only thing of importance in determining learning outcomes, contrary to the pervasive belief in India that personal history and family background determines destiny. The same student can systematically score significantly different marks given different teacher quality. Teacher assignment in principle can play an important role in alleviating unequal outcomes across genders or social groups. As a teacher's quality affects the entire class, it will have a greater impact compared with any student based incentives. Similarly, improving teacher quality is less prone to substitution by households when compared to increasing school inputs. For example, Das et al. (2013) find that households in India and Zambia offset their own spending in response to anticipated school grants, and they suggest caution when interpreting estimates of school inputs on learning outcomes as parameters of an education production function.

As found in many studies in the US, we also find that observed characteristics of teachers in our data do not explain the teacher effectiveness well. This suggests that it may be hard to identify good teachers *ex ante*, but administrative data can be used to identify them *ex post*. As Slater, Davies, and Burgess (2012) suggest that in this situation, there can be a greater role for performance management and personnel policies in schools. In addition, teacher progression policies may be radically rethought if *ex ante* discrimination is hard.

However, certain caveats apply to our conclusions. First, our sample consists of secondary private schools in one of the districts in India, and hence we cannot claim that the conclusions will hold for the entire country. There might be some geographical dimensions which we are unable to capture because of data limitations. Second, it might be possible that the public schools show a different pattern than what is found using a sample of private schools.

Acknowledgements

This project greatly benefited from financial support from the International Growth Centre (IGC) at the London School of Economics for the data collection and the research. The views expressed here are those of the authors and do not necessarily reflect those of the IGC. The authors would like to thank Karthik Muralidharan as well as conference participants at the 2014 Society of Labor Economists (SOLE), the 5th IGC-ISI India Development Policy Conference, and the 2013 Southern Economic Association Conference (SEA), for their comments and suggestions.

References

Aaronson, D., Barrow, L., Sander, W., 2007. Teachers and Student Achievement in the Chicago Public High Schools. Journal of Labor Economics, 25, 95-136.

Chetty, R., Friedman, J. N., Rockoff, J. E., 2014. Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. American Economic Review, 104(9), 2633-79.

Clotfelter, C. T., H., Ladd, & J. Vigdor, 2010. Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects. Journal of Human Resources, 45(3), 655-81.

Clotfelter, C. T., H., Ladd, J., Vigdor, 2007. Teacher credentials and student achievement: Longitudanal analysis with student fixed effects. Economics of Education Review, 26, 673-682.

Clotfelter, C. T., H., Ladd, J. Vigdor, 2006. Teacher-Student Matching and the Assessment of Teacher Effectiveness. Journal of Human Resources, 41, 778-820.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., Sundararaman, V., 2013. School Inputs, Household Substitution, and Test Scores. American Economic Journal: Applied Economics, 5(2), 29-57.

Das, J., Zajonc, T., 2010. India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement. Journal of Development Economics, 92(2), 175-87.

Duflo, E., Hanna, R., 2005. Monitoring Works: Getting Teachers to Come to School. NBER Working Paper, 11880.

Glewwe, P., Ilias, N., Kremer, M., 2010. Teacher Incentives. American Economic Journal: Applied Economics, 2(3), 205-27.

Gordon, R., Kane, T.J., Staiger, D.O., 2006. Identifying Effective Teachers Using Performance on the Job. The Brookings Institution Discussion Paper, 2006-01.

Guarino, C. M., Reckase, M. D., Wooldridge, J. M., 2012. Can Value-Added Measures of Teacher Performance Be Trusted? Education Policy Center at Michigan State University, Working Paper, 18.

Hanushek, E. A., 2003. The Failure of Input-Based Schooling Policies. Economic Journal, 113(485): F64-F98.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., Rivkin, S. G., 2005. The Market for Teacher Quality. NBER Working Paper, 11154.

Hanushek, E. A., Rivkin, S G., 2012. The Distribution of Teacher Quality and Implications for Policy. Annual Review of Economics, 4, 131-157.

Hanushek, E. A., Rivkin, S. G., 2010. Generalizations about using value-added measures of teacher quality. American Economic Review, 100, 267–71.

Hoxby, C., 1996. How teachers unions affect education production. Quarterly Journal of Economics, 111(4), 671-718.

Kane, T. J., McCaffrey, D. F., Miller, T., Staiger, D. O., 2013. Have we identified effective teachers? Validating measures of effective teaching using random assignment. MET Project Research Paper, Bill & Melinda Gates Foundation.

Kane, T. J., Staiger, D. O., 2008. Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. NBER Working Paper, 14607.

Kingdon, G. G., Teal, F., 2010. Teacher Unions, teacher pay, and student performance in India: A pupil fixed effects approach. Journal of Development Economics, 91(2), 278-288. March.

Kingdon, G.G., Teal, F., 2007. Does performance related pay for teachers improve student performance? Some evidence from India. Economics of Education Review, 26(4), 473-86.

Kingdon, G. G., 2006. Teacher characteristics and student performance in India: A pupil fixed effects approach. Global Poverty Research Group (GPRG)-WPS-059, University of Oxford.

Koedel, C., Betts, J., 2011. Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique. Education Finance and Policy, 6(1), 18-42.

Leigh, A., 2010. Estimating teacher effectiveness from two-year changes in students' test scores. Economics of Education Review, 29, 480-88.

Muralidharan, K., 2013. Priorities for Primary Education Policy in India's 12th Five Year Plan. India Policy Forum 2012-13, Vol. 9, 1-46.

Muralidharan, K., 2012. Long Term Effects of Teacher Performance Pay. unpublished manuscript, UCSD, April 2012.

Muralidharan, K., Sheth, K., 2015. Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers. Journal of Human Resources, in press.

Muralidharan, K., Sundararaman, V., 2011. Teacher Performance Pay: Experimental Evidence from India. Journal of Political Economy, 119(1), 39-77.

Muralidharan, K., Sundararaman, V., 2013. Contract Teachers: Experimental Evidence from India. NBER Working Paper, 19440.

Nye, B., Konstantopoulos, S., Hedges, L. V. 2004. How large are teacher effects? Education Evaluation and Policy Analysis, 26, 237–57.

Ouazad, A., 2008. A2REG: Stata module to estimate models with two fixed effects. Statistical Software Components S456942, Boston College Department of Economics.

Rawal, S., Kingdon, G. G., 2010. Akin to my teacher: Does caste, religious or gender distance between student and teacher matter? Some evidence from India. DoQSS Working Papers, 10-18.

Rivkin, S. G., Hanushek, E. A., Kain, J. F., 2005. Teachers, schools, and academic achievement. Econometrica, Vol. 73, pp. 417-458.

Rockoff, J. E., 2004. The impact of individual teachers on student achievement: Evidence from panel data. American Economic Review, 94, 247-252.

Rockoff, J. E., Speroni, C., 2010. Subjective and Objective Evaluations of Teacher Effectiveness. American Economic Review, 100, 261-66.

Slater, H., Davies, N. M., Burgess, S., 2012. Do teachers matter? Measuring the variation in teacher effectiveness in England. Oxford Bulletin of Economics and Statistics, 74(5), 629-45.

Todd, P. E., Wolpin, K. I., 2003. On The Specification and Estimation of The Production Function for Cognitive Achievement. Economic Journal, 113(485), F3-F33.

Walker, M., 2011. PISA 2009 Plus Results: Performance of 15-year-olds in reading, mathematics, and science for 10 additional participants. Australian Council for Educational Research.

Table 1: Variable means and standard deviations

	Mean	Standard Deviation	Number of Observations
Dependent variable	-0.004	1.002	
Twelfth-grade Score			
English	78.60	8.95	8319
Chemistry	75.31	12.80	5693
Physics	73.03	14.34	5693
Mathematics	78.46	17.02	5266
Computer Science	86.10	9.33	4497
Environmental Science	83.84	10.00	3349
Hindi	81.15	9.15	1451
Biology	73.08	13.12	1198
Commerce	65.69	17.29	372
Economics	60.95	19.20	261
Teachers credentials			
Age	41.32	8.45	188
Male	0.51		188
Experience	9.97	6.93	188
1-4 years	0.25		46
5-8 years	0.25		47
9-13 years	0.26		49
9 or more years	0.25		46
Teacher have Master or higher degree	0.85		188
Teacher have received B.Ed training	0.59		188

Note: 1) The dependent variable is normalized student achievement score, normalized for each subject and cohort.

2) English is mandatory, while a student can choose other 4-6 subjects.

Table 2: Distribution of teachers fixed effects

	Un-weighted	Weighted
10th percentile	-0.491	-0.474
25th percentile	-0.196	-0.143
50th percentile	0.086	0.126
75th percentile	0.314	0.295
90th percentile	0.762	0.553
90-10 gap	1.253	1.027
90-50 gap	0.677	0.427
75-50 gap	0.228	0.169
75-25 gap	0.510	0.438
50-25 gap	0.282	0.269
Standard Deviation of TFE	0.513	0.453
Adjusted Standard Deviation of TFE	0.491	
R-Square	0.428	
P-values for F-test on:		
Teacher Fixed effects	0.000	
Tenth-grade math score	0.000	
Subject Fixed effect	0.000	
Year Fixed effects	0.263	
Score Units	Normalized	
Observations	38288	
Number of student thresholds	15	

Note: 1) Dependent variable is twelfth-grade subject specific normalized score, while X-matrix includes tenth grade subject specific normalized score, subject dummies, year dummies, teacher dummies.

2) Weighted implies percentiles are generated using numbers of students taught by that particular teacher as weight

.

Table 3: Variability in teacher effectiveness

	Model (1)	Model (2)	Model (3)
Standard Deviation	0.513	0.379	0.379
Adjusted Standard Deviation	0.491	0.366	0.366
90-10 gap	1.253	0.934	0.974
90-50 gap	0.677	0.513	0.509
75-50 gap	0.228	0.217	0.219
75-25 gap	0.510	0.456	0.476
50-25 gap	0.282	0.238	0.257
P-values of *F*-test for joint significance of teacher fixed effects	0.000	0.000	0.000
Teacher effects	Yes	Yes	Yes
Subject effects	Yes	Yes	Yes
Year effects	Yes	Yes	Yes
Pupil effects	NO	Yes	Yes
School effects	NO	NO	Yes

Note: Dependent variable is twelfth-grade subject specific normalized score, while dependent variables in model (1) includes tenth grade subject specific normalized score, subject dummies, year dummies, teacher dummies; model (2) add individual student fixed effects to model (1) independent variables, while model (3) add school fixed effects to model (2) dependent variables.

Table 4: Explaining teacher effectiveness (teacher fixed effects)

	(1)	(2)	(3)	(4)
	Model (1)	Model (2)	Model (3)	Model (3)
Age of Teacher	-0.008	-0.006	-0.006	-0.004
	(0.005)	(0.004)	(0.004)	(0.004)
Sex code as per M=1 , W=0	-0.084	-0.022	-0.028	-0.028
	(0.077)	(0.059)	(0.059)	(0.059)
Experience of teaching in that school	0.001	-0.000	0.000	
	(0.006)	(0.005)	(0.005)	
experience: 1-4 years				0.041
				(0.081)
experience: 5-8 years				Omitted
experience: 9-13 years				0.015
				(0.079)
experience: 14 or more years				-0.044
				(0.090)
Teacher is MA	0.343***	0.107	0.106	0.092
	(0.103)	(0.079)	(0.079)	(0.080)
Teacher has received B.Ed training	0.136*	-0.010	-0.011	-0.006
	(0.078)	(0.060)	(0.060)	(0.062)
Constant	0.100	0.252	0.261	0.192
	(0.228)	(0.174)	(0.174)	(0.184)
F-test for joint significance of experience indicators (*p-value*)				0.825
Observations	188	188	188	188
R-squared	0.107	0.029	0.030	0.034

Note: 1) The dependent variable is teacher fixed effects estimated from model (1), model (2), and model (3) as reported in Table 3. 2) Standard errors in parentheses. 3) *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Robustness tests: the teacher effectiveness in different samples

		Standard Deviation in Teacher fixed effects	Adjusted Standard Deviation in Teacher Fixed effects	Number of teachers
(1)	All subjects (10)	0.379	0.366	191
(2)	*Six subjects: English, Physics, Chemistry, Biology, Math, Computer Science*	*0.362*	*0.350*	*142*
	Drop one subject (9 subjects)			
(3)	Drop Biology	0.376	0.364	179
(4)	Drop Chemistry	0.382	0.368	169
(5)	Drop Computer Science	0.419	0.403	172
(6)	Drop English	0.398	0.379	150
(7)	Drop Environmental Science	0.357	0.343	167
(8)	Drop Hindi	0.387	0.375	180
(9)	Drop Math	0.394	0.383	168
(10)	Drop Physics	0.365	0.351	166
(11)	Drop Economics	0.384	0.371	184
(12)	Drop Commerce	0.381	0.368	184

Note: The estimates in the first row is based model (3), Table 3 that is estimated from a sample that include all ten subjects. Estimates in row (2) are based on a sample that include only six subjects. Estimates from row (3) - row (12) are based on a sample of nine subjects.

Appendix

Table A.1: Number of students who appeared for 12^{th} grade exam in each subjects

Subject	Number of students who took 12th Exam in each subject	No 10th score in that subject available	Number of students in each subject for analysis	% of column (1)
	(1)	(2)	(3)	(4)
English	8,320	1	8,319	100
Math	5,268	2	5,266	100
Chemistry	5,700	7	5,693	100
Physics	5,700	7	5,693	100
Biology	1,200	2	1,198	100
Computer Science	4,667	170	4,497	96
Environmental Science	6,789	3,440	3,349	49
Hindi	5,465	4,014	1,451	27
Commerce	2,319	1,947	372	16
Economics	2,418	2,157	261	11

Note: column (1) gives the number of students who appeared for 12^{th} grade exam in each subject. Column (2) counts number of students for whom no 10^{th} grade score reported in that subject in data. Column (3) reports number of students in each subject for whom both 12^{th} and 10^{th} grade score is reported (our final sample).