

Defining Teacher Quality in India

Mehtabul Azam[†]

Oklahoma State University & IZA

Geeta Gandhi Kingdon

IoE, University of London & IZA

October, 2013

Abstract

Using administrative data from linked private schools in India that matches 8,319 pupils to their subject specific teachers at the senior secondary level, we estimate the importance of individual teachers on student outcomes in the high-stake senior secondary exam (at the end of twelfth-grade). We control for prior achievement at the secondary level (at the end of tenth-grade) to estimate the value added of a teacher over a two year course, and define a good teacher as one who consistently gets higher achievement for students. In addition to the 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 0.38 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: I21, O15

Keywords: teacher value-added, teacher quality, student achievement, India

[†]Corresponding author: Mehtabul Azam, 326, Business Building, Oklahoma State University, Stillwater, OK. Email: mazam@okstate.edu.

1 Introduction

One of the important debates in education policy has been how to improve the 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 discussed. It has been increasingly recognized that one of the most important factors determining student achievement is teacher quality.¹ Hence, identifying the relative effectiveness of individual teachers is of increasing 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.² In recent times, increasing attention has been focused on direct relationship between teachers and student outcomes. This outcome-based perspective, 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 or prior teachers) (Hanushek and Rivkin, 2012). Several recent papers in the US have sought to identify and reliably measure teacher value-added. Using mainly the administrative data from the US schools, they have sought to identify the teacher quality effect by the teacher fixed effect in a student achievement equation where a teacher is matched to students in the various classes of

¹The 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>

²No 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).

a given grade she/he taught in a year or the 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, Kain, O’Brien, and Rivkin (2005) look at the market for teacher quality 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 accountability and teacher incentive programs. Rivkin, Hanushek, and Kain (2005) use data for three cohorts spanning over 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 maths and 0.095 in English. Aaronson, Barrow and Sander (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, Davies and Burgess (2012) links 7,305 pupils to the individual teachers who taught them, in each of their compulsory subjects in the high-stakes 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 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, 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.

Some papers have taken an alternative approach where they do not estimate the “teacher quality-as measured by their ability to raise test scores,” but examine the relationship between teacher characteristics and achievement directly. Some of such studies have used experimental methods, mainly investigating the effect of teacher incentives (Duflo and Hanna, 2005; Glewwe, Ilias and Kremer, 2010; Muralidharan and Sundararaman, 2011). Other studies have used statistical approaches such as an instrumental variable approach (Hoxby, 1996; Kingdon and Teal, 2005), estimating standard cross-sectional achievement production function, or a panel data approach (Clotfelter, Ladd and Vigdor, 2006, 2010; Kingdon, 2006).³ The evidence from these studies is not undisputed. In addition, the effects of specific teacher characteristics cannot be taken as overall contribution of teachers. The findings in these studies that commonly used indicators of quality differences are not closely related to achievement gains, led to shift from a research design that focuses on the link between student outcome and specific teacher characteristics to a research framework that uses less parametric approach to identify overall teacher contribution to learning as teacher value-added (Hanushek and Rivkin, 2010).

The importance of improving educational attainment within school has been growing in education policy debate in India, or in developing countries per se, as the focus has been gradually shifting from providing access to education to providing access to qual-

³Hanushek (2003) provides a review of the US and international evidence on the effectiveness of input based policies.

ity education. In the Indian context, the increasing interest in quality of education is partially driven by the realization that the rapid gains in enrollment and attendance is not translating into gains in the cognitive skills as measured by test scores in reading, writing, or math. These test scores remain low compared to international benchmarks.⁴ Given the evidence in the US and UK that the teachers play a key role in improvement of student achievement, it is important to address how much teachers can play a role in improvement of student achievement in the Indian context. However, we are not aware of any work on “Teacher quality/effectiveness-as measured by their ability to raise test scores” for India or for any developing country.

Although, there is no work on 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. There are few studies which examine the impacts of input based policies in India 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 scores of tenth-grade students in different subjects from a sample of schools, and supplemental postal questionnaire sent by the Exam Board and filled by students, teachers that teach them, and student’s principal. 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

⁴Two 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 2009+, these two states beat out only Kyrgyzstan. 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 the 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 Maurice, 2011). What is more worrisome is that these two states are better states among national rankings. 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 near the bottom of the global rankings.

10, for that subject. She finds that ‘Masters level or higher’ qualification and ‘possession of pre service teacher training’ raises 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) finds that a union membership of teacher reduces student achievement and increases the salary costs in private schools in India. However, both Kingdon (2006) and Kingdon and Teal (2010) do not control for prior achievement. Rawal and Kingdon (2010) explore whether having the same caste, religion, or gender teacher influence student outcomes in the sample primary schools. Similarly, Muralidharan and Sheth (2013) examine whether having the 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 (2010) provide experimental evidence on contractual teachers.

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 fact that teachers are not randomly assigned to students. A similar strategy to that implemented in this paper is

carried out in Slater, Davies, and Burgess (2012).

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 teacher hardly explain any of the variation in the teacher's quality. These findings basically corroborate recent finding in the US and UK. The findings question the emphasis put on the certifications in hiring or retaining teachers in India. Although there is a potential to improve achievement through improving average teacher quality, it is not as 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, Kane, and Staiger, 2006). However, this requires a significant improvement and building up administrative database, which is almost non-existence in the Indian context, that can be used to estimate value added with some confidence.

The paper contributes to the existing literature in the following ways. First, the paper is the first study (to our knowledge) in developing country context to use administrative panel data to estimate teacher quality directly. The paper provides the first estimates of teacher effectiveness in India, or in a developing country. In addition, the paper is only the second paper that we are aware of that examines senior secondary school teachers.⁵ Second, the paper corroborates the findings of the US and UK in a developing country

⁵Slater, Davies, and Burgess (2012) estimate teacher quality at the senior secondary level. The context studied by Aaronson, Barrow, and Sander (2007) is high school teachers (ninth-grade) from Chicago public schools. Majority of the papers on teacher quality focuses on elementary schools. As argued by Aaronson, Barrow, and Sander (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.

context, increasing the confidence in those finding irrespective of the context. Third, the paper provides evidence that the emphasize put on teachers training by Indian policymakers is probably misguided, and the compensation structure can be based on the value-added of a teacher rather than based on teachers characteristics.

The paper is organized as follows. Section 2 describes the data, Section 3 describes the methodology followed in this paper. Section 4 presents the results, and Section 5 concludes.

2 Data

We use administrative records provided by a group of linked private schools from one of the districts in India. The records include all students who took twelfth-grade exam between 2006 and 2010 in ten schools. Each of ten schools has multiple sections for twelfth-grade. The records include twelfth and tenth-grade scores in multiple subjects. The twelfth-grade (known as Indian School Certificate, ISC) exams are typically taken at the age 17-18, and are considered very important, as many 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. Unlike exams in other grades which are typically set and graded within 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.^{6,7} In addition to twelfth and tenth-grade scores, the school also provided information on teachers who taught those subjects for two years. Student-teacher

⁶The 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. Candidates have to attempt the subjects of English and three to five other subjects of the candidate's choosing, however, the choice are restricted by schools offering only a limited number of these subjects.

⁷Clotfelter, Ladd, and Vigdor (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.

data are matched in and by the school, thus ensuring a high-quality match. The schools also provided us some human capital and demographic characteristics of the teachers. However, the data does not contain any student characteristics.

The data contain the twelfth grade exam results for 8,382 pupils. We dropped 57 pupils as the tenth-grade score was missing for them. Typically, the average number subjects for which twelfth grades scores are reported is 5.8, while the scores are reported for at least four subjects for 99 percent of the twelfth-grade students. Essentially, an observation is a pupil-teacher match, or equivalently a pupil-subject-teacher match as each teacher only teaches one subject. We also dropped observations where we do not have information on the pupils tenth-grade score in the particular subject.⁸ The data used in the initial regression contain 38,265 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).⁹ 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 master degree or higher, while 59 percent of the teachers have received Bachelor of Education (B.Ed) training.¹⁰

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

⁸It is not necessary that a student take exactly the same set of subjects in both tenth and twelfth grade. As a result, tenth grade scores are missing for some subjects for some pupils.

⁹Observed characteristics are missing for 3 teachers

¹⁰Bachelor 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 bachelors degree.

can be interpreted as fraction of the standard deviation.

Although this 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, Barrow, and Sander (2007), Clotfelter, Ladd, and Vigdor (2006, 2007, 2010), Rivkin, Hanushek, and Kain (2005), Rockoff (2004), Kane, Rockoff, and Staiger (2008), and Slater, Davies, and Burgess (2012). Like Aaronson, Barrow, and Sander (2007), Rockoff (2004) and Slater, Davies, and Burgess (2012), but unlike Rivkin, Hanushek, and Kain (2005) and Kane, Rockoff, and Staiger (2008), we can match a student to her/his actual teacher, rather than to the school-grade average teacher. Unlike Aaronson, Barrow, and Sander (2007), Clotfelter, Ladd, and Vigdor (2006, 2007, 2010), Rivkin, Hanushek, and Kain (2005), Rockoff (2004), Kane, Rockoff, and Staiger (2008), and like Slater, Davies, and Burgess (2012) our context is students taking terminal exams that are very important to them and to the school.

Similar to Clotfelter, Ladd, and Vigdor (2010) and Slater, Davies, and Burgess (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, 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 the most general formulation of value added model, education is considered a cumulative process. Achievement in grade g involves not only educational inputs in that grade, but also the entire history of inputs that provided the basic knowledge. The model—often referred to as the generalized cumulative effects model (CEM)—includes all relevant past child, family, and school inputs (Guarino, Reckase, and Wooldridge, 2012; Todd and

Wolpin, 2003; Hanushek and Rivkin, 2012), and expressed as:

$$Y_{ig} = f(S_{ig}, S_{i,(g-1)}, \dots, S_{i,0}, X_{ig}, X_{i,(g-1)}, \dots, X_{i,0}, \theta_i, \epsilon_{ig}, \epsilon_{i,(g-1)}, \dots, \epsilon_{i,0}) \quad (1)$$

where Y_{ig} is the achievement of child i in grade g , S_{ig} represents school-related inputs, X_{ig} represents a set of relevant time-varying child and family inputs, θ_i captures the time-invariant unobserved student ability, and the ϵ_{ig} represent the idiosyncratic shock. Assuming linear form, equation (1) can be written as:

$$Y_{iG} = \sum_{g=0}^G S_{ig} \phi_g + \sum_{g=0}^G X_{ig} \gamma_g + \sum_{g=0}^G \theta_i + \sum_{g=0}^G \epsilon_{ig} \quad (2)$$

The data requirement for estimation of equation (2) is too restrictive. We can decompose the achievement determinants into current grade G and prior grade influence by imposing a structure on the parameters such as a geometrically declining pattern for inputs in the more distant past (indicating that the impact of past inputs depreciates at a constant rate λ over time). We get the following:

$$\begin{aligned} Y_{iG} = & S_{iG} \phi + X_{iG} \gamma + \theta_i + \epsilon_{iG} + \sum_{g=0}^{G-1} S_{ig} \phi (1 - \lambda)^{G-g} \\ & + \sum_{g=0}^{G-1} X_{ig} \gamma (1 - \lambda)^{G-g} + \sum_{g=0}^{G-1} \theta_i (1 - \lambda)^{G-g} + \sum_{g=0}^{G-1} \epsilon_{ig} (1 - \lambda)^{G-g} \end{aligned} \quad (3)$$

As the relationship holds across grades, the later part of equation (3) is simply $(1 - \lambda)Y_{i,(G-1)}$, and equation (3) can be written as current achievement as a function of (depreciated) past achievement plus the inputs during grade G :

$$Y_{iG} = S_{iG} \phi + X_{iG} \gamma + \theta_i + (1 - \lambda)Y_{i,(G-1)} + \epsilon_{iG} \quad (4)$$

Many data sets including administrative data have limited information on family income and other current family characteristics; however, the availability of multiple observations for students in panel data sets makes possible alternative approaches to account for such heterogeneity. Many studies have estimated the following specification:

$$Y_{ijkt} = (1 - \lambda)Y_{i,(t-1)} + \beta X_{it} + \tau T_j + \theta_i + \mu_t + \rho_k + \epsilon_{ijkt} \quad (5)$$

where i refers to the student, j to the teacher, k to the school and t to time. Thus Y_{ijkt} refers to the achievement of the i^{th} student, taught by teacher j in school k at time t . The variable $Y_{i,(t-1)}$ then refers to this achievement at an earlier period and the equation is interpreted as a value-added achievement function as the achievement at any period controls for the achievement in an earlier period. Controlling for past scores is especially important with information on family and previous schooling lacking.¹¹ The variable X_{it} refers to the time varying aspects of the student. T_j are dummies for teachers, hence each element τ_j of τ refers to the effects of one year spent with teacher j , and thus 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 change in test performance, respectively, while ϵ_{ijkt} is the white noise.

In our data, we do not observe the students multiple times; however, 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 year back (during the tenth-grade exam). We follow the modified version of equation (1):¹²

$$Y_{izjkt}^{12} = (1 - \lambda)Y_{iz}^{10} + \tau T_j + \theta_i + \mu_t + \rho_k + \delta_z + \epsilon_{ijzkt} \quad (6)$$

¹¹Inclusion of lagged achievement account for the achievement that the student brings to the classroom, before being taught by the relevant teacher in the relevant classroom.

¹²Similar strategy is also implemented in Slater, Davis, and Burgess (2012).

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 tenth-grade score in subject z , and δ_z are subject dummies. Since we have scores for multiple subjects at the same point of time, we can identify pupil fixed effects. However, similar to many administrative data, our data lacks students characteristics (X_i is not included in equation (6)). In order to allow for correlation across subjects for each pupil, we cluster standard errors at the pupil level.

Inclusion of the pupil fixed effects means, as would be the case in longitudinal studies, that effects of the T variable are estimated within students, i.e., they are based on the fact that different subjects are taught by different teachers. Pupil fixed effects powerfully nets out all the observed and unobserved factors at a point of time which is invariant across subjects. One can argue that although pupil fixed effects potentially capture all the observed and unobserved subject invariant factors, it fail to take account of any changes that have occurred over time (in this case, between grade ten and grade twelve). However, the traditional fixed effects (applied across time) also do not capture the changes in unobserved factors, and assume that unobserved factors have same impact over time. In addition, most of the administrative data lack extensive set of variables to fully account for changes in observed factors. One can also argue that pupil fixed effects can potentially account for time varying observed and unobserved factors under the assumption that they remain same across subjects. Another but related issue that might be of concern is that of the subject varying factors (mostly unobserved factors, such as subject specific ability which is analogues 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 have more than one subject specific teacher as in our case. We believe that prior achievement can also work as good proxy for subject-specific ability, and bias if any, should be minimal.

We observe teachers linked to students over time.¹³ 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 (2) 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 school 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, Davis, and Burgess, 2012).

3.1 Identification of teachers effects

To allow for non-random assignment of students to teachers, we include both prior subject-specific test score and pupil fixed effects. Hence identification arises from comparing the exam score progress of a student taught different subjects by different teachers over the same 2-year period.¹⁴ As noted by Slater, Davis, and Burgess (2011), this controls for all general attributes of the student at one point of time: intelligence, effort, motivation, imagination, and ability to learn, and also conditions on subject specific ability as measured by the tenth-grade score.

We believe that inclusion of tenth-grade score and pupil fixed effect addresses the non-random sorting of students. However, if students are assigned to teachers based on

¹³Observing teachers over multiple years allows us to distinguish permanent teacher quality from idiosyncratic class-year shocks.

¹⁴Rivin, Hanushek, and Kain (2005), Rockoff (2004), and Aaronson, Barrow, and Sander (2007) use exam scores that are administered annually. However, similar to Slater, Davies, and Burgess (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 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, Davies, and Burgess weigh each pupil-teacher observation by $1/n$ if a student has n teachers in the subject.

expected progress in a subject relative to expected progress in the 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.¹⁵ 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, Davis, and Burgess, 2012).

3.2 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 extra 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 (6). That we estimate the following equation:

$$\hat{\tau}_j = \pi Z_j + u_j \tag{7}$$

where Z_j is j ’s teacher’s characteristics.¹⁶

¹⁵Rothstein (2010) notes, 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 evidence that dynamic sorting of student and teachers to classrooms is transitory and that observing teachers over multiple time periods mitigates the dynamic sorting bias envisioned by Rothstein.

¹⁶We take an average of teacher’s experience, which varies over time.

3.3 Sampling Variation

As argued by Kane and Staiger (2002), Rockoff (2004), and Aaronson, Barrow, and Sander (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. Aaronson, Barrow, and Sander (2007) shows 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 only included those teachers in our analysis who have taught at least 15 students. In addition, following Aaronson, Barrow, and Sander (2007), we analytically adjust the variance of estimated teacher fixed effects, $\hat{\tau}$. Aaronson, Barrow, and Sander (2007) assumes 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 and subtract this from the observed variance of $\hat{\tau}$ 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.¹⁷ 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

¹⁷One can argue that the tenth-grade test score may serve as good proxy for θ_i . As suggested by Guarino, Reckase, and Wooldridge (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.

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, $\hat{\tau}$, is 0.511, which is quite broad. Aaronson, Barrow, and Sander (2007) find a standard deviation of 0.15 in teacher fixed effects using a similar model and Chicago public high school data.¹⁸ 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 would move a student from the 50th percentile to the 70th percentile.

What is more interesting is gap in value addition between a 90th (very good teacher) and 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 1 present the estimated teacher fixed effects weighted by the number of students taught by each teacher. Weighted standard deviation in teacher fixed effects drop to 0.449 from the un-weighted standard deviation of 0.511. 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.¹⁹ Table 3 presents the results. For comparison purposes, column (1) of Table 3, repeat the

¹⁸They reports 0.4 standard deviation in terms grade equivalents (the standard deviation of ninth-grade is 2.71)

¹⁹The 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.

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 0.511 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, Davies, and Burgess (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 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 increases the pupil achievement by 0.379 of the standard deviation, which moves an average student to about 65th percentile of the score distribution. Slater, Davies and Burgess 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.²⁰ 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, Davies, and Burgess in UK.²¹ 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, Davies, and Burgess report 95th-5th gap of 1.18 of the standard deviation. Thus our estimates are much in line with the estimates reported by Slater, Davies, and Burgess.

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

²⁰They report standard deviation of 1.705 in GCSE score.

²¹They report an IQR of 0.75 in terms of GCSE points (standard deviation of GCSE is 1.705).

deviation of teacher fixed effects, however, there are some marginal changes in gains when teachers at different percentiles are compared.²² 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. Figure 1 and Figure 2 presents the un-weighted and weighted kernel density of teacher fixed effects as estimated from three models reported in Table 3, respectively.

Hanushek and Rivkin (2012, 2010) summarize the results of many studies in the US (Table 1 of Hanushek and Rivkin, 2012, 2010) in terms of standard deviations. They report an average standard deviation of 0.13 in teacher effectiveness for reading score, while an average standard deviation of 0.17 in teacher effectiveness for math score. They report a highest (lowest) standard deviation of 0.18 (0.07) in teacher effectiveness for reading score found by Kaine and Staiger (2008) (Nye, Konstantopoulos, and Hedges, 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, Hanushek, and Kain (2005); and Hanushek and Rivkin (2010).²³

In comparison to the US studies, our estimate of standard deviation of 0.379 (or adjusted standard deviation of 0.336) in teacher effectiveness seems much larger. However, while the US studies estimates of teacher effectiveness is based on one year spent with the teacher, our estimates of teacher effectiveness is based on two year 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) in terms of context, empirical

²²As 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.

²³These estimates are adjusted for sampling variation.

strategy, 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 0.379. 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

In the earlier section, we have found that teacher quality remains an important determinant of achievement. However, how a teacher’s quality is related to teacher’s credentials is also an important issue. For example, how 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 schools and classrooms (Clotfelter, Ladd, and Vigdor, 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, $\hat{\tau}$, 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 explained by the teacher characteristics is very low.²⁴ Column (1) of Table 4 reports the results which uses teacher fixed effects estimated from model (1) with no pupil fixed effects. Only one of the six characteristics (teacher has a master degree) is significant. 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. However, this impact is mainly because of positive matching of high ability students with the teachers with master degree. Using the teacher fixed effects from

²⁴As argued by Aaronson, Barrow, and Sander (2007), the reported R^2 is an understatement of the explanatory power since some variation in teacher fixed effects is due to sampling.

model (2) and (3) that controls for pupil fixed effects; none of the teacher characteristics remain statistically significant. This is in contrast to Kingdon (2006) who took a direct estimation approach and finds that ‘masters level or higher’ qualification and possession of ‘pre service teacher training’ each raise pupil achievement by 0.09 standard deviations. She suggests that these are upper bound estimates. Moreover, the variation in teacher quality explained by teachers characteristics is extremely low. Thus the observed factors (observed in the data) explains very little of the teacher variation, while the unobserved factors such as drive, passion, connection with the students, and so forth, account for the rest of the variation in teacher effectiveness. The lack of explanatory power of human capital regressors and no association of these regressors with teacher quality is consistent studies in the US (Aaronson, Barrow, and Sander, 2007; Rivkin, Hanushek, and Kain, 2005) and UK (Slater, Davies, and Burgess, 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 distribution of marks across subjects, despite using z -scores. As our identification strategy exploits within pupil, across subjects variation, using more subjects creates more variation.²⁵ 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 chose that subject. Although there is some variation in the standard deviations in teacher fixed effects estimated using different sub-samples, the difference is marginal when compared to standard

²⁵Note that although our sample consists of ten subjects, for most of the pupils we have either five or six subjects (five or six rows). However, as there is a choice over subjects, those subjects will not be necessarily the same across all pupils.

deviation estimated using the full sample. The results of Table 5 increase the confidence in the standard deviation estimated from full sample, i.e. it is not driven by 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 schools from one of the districts in 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 family background is not everything, which is of great importance in the Indian context, as there is a pervasive belief in India that the personal history determines destiny in India.

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 teacher's effectiveness 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) finds that households in India and Zambia offset their own spending in response to anticipated 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 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.

References

- [1] Angrist, J. and Lavy, V. (1999), “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement,” *Quarterly Journal of Economics*, 114(2): 533-75.
- [2] Aaronson, D., Barrow, L. and Sander, W. (2007), “Teachers and Student Achievement in the Chicago Public High Schools,” *Journal of Labor Economics*, 25, 95-136.
- [3] Clotfelter, C. T., Helen, L. F., and Jacob, V. L. (2006), “Teacher-Student Matching and the Assessment of Teacher Effectiveness,” *Journal of Human Resources*, 41, 778-820.
- [4] Clotfelter, C. T., Helen, L. F., and Jacob, V. L. (2007), “Teacher credentials and student achievement: Longitudinal analysis with student fixed effects,” *Economics of Education Review*, 26, 673-682.
- [5] Clotfelter, C. T., Helen, L. F., and Jacob, V. L. (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.
- [6] Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., and Sundararaman, V. (2013), “School Inputs, Household Substitution, and Test Scores,” *American Economic Journal: Applied Economics*, 5(2), 29-57.
- [7] Das, J. and 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.
- [8] Duflo, E. and Hanna, R. (2005), “Monitoring Works: Getting Teachers to Come to School,” *NBER Working Paper*, 11880.

- [9] Glewwe, P., Ilias, N., and Kremer, M. (2010), "Teacher Incentives," *American Economic Journal: Applied Economics*, 2(3), 205-27.
- [10] Gordon, R., Kane, T.J., and Staiger, D.O. (2006), "Identifying Effective Teachers Using Performance on the Job," The Brookings Institution.
- [11] Goldhaber, D., and Hansen, M. (2010), "Using Performance on the Job to Inform Teacher Tenure Decisions," *American Economic Review*, 100, 250-55.
- [12] Guarino, C. M., Reckase, M. D., and Wooldridge, J. M. (2012), "Can Value-Added Measures of Teacher Performance Be Trusted?" Education Policy Center at Michigan State University, Working Paper 18.
- [13] Hanushek, E. A. (2003), "The Failure of Input-Based Schooling Policies," *Economic Journal*, 113(485): F64-F98.
- [14] Hanushek, E. A. and Rivkin, S G. (2012), "The Distribution of Teacher Quality and Implications for Policy," *Annual Review of Economics*, 4, 131-157.
- [15] Hanushek, E. A., Kain, J. F., O'Brien, D. M., and Rivkin, S. G. (2005), "The Market for Teacher Quality." NBER Working Paper, 11154.
- [16] Hoxby, C. (1996), "How teachers unions affect education production," *Quarterly Journal of Economics*, 111(4), 671-718.
- [17] Kane, T. J., McCaffrey, D. F., Miller, T., and 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.
- [18] Kingdon, G. G. (2006), "Teacher characteristics and student performance in India: A pupil fixed effects approach," *Global Poverty Research Group (GPRG)-WPS-059*.

- [19] Kingdon, G. and Teal, F. (2010), “Teacher Union, teacher pay, and student performance in India: A pupil fixed effects approach,” *Journal of Development Economics*, 91, 278-288.
- [20] Kingdon, G.G. and Teal, F. (2007), “Does performance related pay for teachers improve student performance? Some evidence from India,” *Economics of Education Review*, 26(4), 473-86.
- [21] Muralidharan, K. (2012), “Long Term Effects of Teacher Performance Pay,” *unpublished manuscript*, UCSD, April 2012.
- [22] Muralidharan, K. and Sheth, K. (2013), “Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers,” *NBER Working Paper*, 19341.
- [23] Muralidharan, K., and Sundararaman, V. (2011), “Teacher Performance Pay: Experimental Evidence from India,” *Journal of Political Economy*, 119(1), 39-77.
- [24] Muralidharan, K., and Sundararaman, V. (2010), “Contract Teachers: Experimental Evidence from India,” *unpublished manuscript*, UCSD, May 2010.
- [25] Koedel, C. and 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.
- [26] Leigh, A. (2010), “Estimating teacher effectiveness from two-year changes in students test scores,” *Economics of Education Review*, 29, 480-88.
- [27] Ouazad, A. (2008), “A2REG: Stata module to estimate models with two fixed effects,” Statistical Software Components S456942, Boston College Department of Economics.
- [28] Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005), “Teachers, schools, and academic achievement,” *Econometrica*, Vol. 73, pp. 417-458.

- [29] Rockoff, J. E. (2004), “The impact of individual teachers on student achievement: Evidence from panel data,” *American Economic Review*, 94, 247252.
- [30] Rockoff, J. E, and Speroni, C. (2010), “Subjective and Objective Evaluations of Teacher Effectiveness,” *American Economic Review*, 100, 261-66.
- [31] Slater, H., Davies, N. M., and Burgess, S. (2012), “Do teachers matter? Measuring the variation in teacher effectiveness in England,” *Oxford Bulletin of Economics and Statistics*, 74(5), 629-45.
- [32] 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.
- [33] Todd, P. E. and Wolpin, K. I. (2003), “On The Specification and Estimation of The Production Function for Cognitive Achievement,” *Economic Journal*, 113(485), F3-F33.

Table 1: Variable means and standard deviations

| | Mean | Standard Deviation | Number of Observations |
|--------------------------------------|--------|--------------------|------------------------|
| Dependent variable | -0.004 | 1.002 | 38288 |
| 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 | 5260 |
| 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 | 1196 |
| Commerce | 65.69 | 17.29 | 348 |
| Economics | 60.95 | 19.20 | 221 |
| Teachers credentials | | | |
| Age | 41.324 | 8.445 | 188 |
| Male | 0.511 | 0.501 | 188 |
| Experience | 9.968 | 6.928 | 188 |
| Teacher have Master or higher degree | 0.851 | 0.357 | 188 |
| Teacher have received B.Ed training | 0.590 | 0.493 | 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 five subjects.

Table 2: Distribution of teachers fixed effects

| | Un-weighted | Weighted |
|---|-------------|----------|
| 10th percentile | -0.498 | -0.461 |
| 25th percentile | -0.213 | -0.134 |
| 50th percentile | 0.088 | 0.131 |
| 75th percentile | 0.323 | 0.308 |
| 90th percentile | 0.773 | 0.568 |
| 90-10 gap | 1.271 | 1.029 |
| 90-50 gap | 0.685 | 0.437 |
| 75-50 gap | 0.235 | 0.177 |
| 75-25 gap | 0.537 | 0.442 |
| 50-25 gap | 0.302 | 0.265 |
| Standard Deviation of TFE | 0.511 | 0.449 |
| Adjusted Standard Deviation of TFE | 0.490 | |
| R-Square | 0.429 | |
| <i>P-values</i> for F-test on: | | |
| Teacher Fixed effects | 0.000 | |
| Tenth-grade math score or year dummies or subject dummies | 0.000 | |
| 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 | 0.511 | 0.379 | 0.379 |
| Adjusted Standard Deviation | 0.490 | 0.366 | 0.366 |
| 90-10 gap | 1.271 | 0.934 | 0.974 |
| 90-50 gap | 0.685 | 0.513 | 0.509 |
| 75-50 gap | 0.235 | 0.217 | 0.219 |
| 75-25 gap | 0.537 | 0.456 | 0.476 |
| 50-25 gap | 0.302 | 0.238 | 0.257 |
| <i>P</i> -values of <i>F</i> -test for joint significance of teacher fixed effects | 0.00 | 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)

| | Model (1) | Model (2) | Model (3) |
|-------------------------------------|---------------------|-------------------|-------------------|
| Age of Teacher | -0.008 (0.005) | -0.006 (0.004) | -0.006 (0.004) |
| Teacher is Male | -0.087 (0.077) | -0.016 (0.059) | -0.023 (0.059) |
| Teaching experience in that school | 0.009 (0.018) | -0.019 (0.013) | -0.019 (0.013) |
| Experience Square/100 | -0.023 (0.058) | 0.067 (0.044) | 0.068 (0.044) |
| Teacher is MA | 0.353*** (0.104) | 0.092 (0.079) | 0.091 (0.079) |
| Teacher has received B.Ed. training | 0.127 (0.079) | -0.002 (0.060) | -0.003 (0.060) |
| Constant | 0.059 (0.240) | 0.339* (0.183) | 0.348* (0.183) |
| Observations | 188 | 188 | 188 |
| R-squared | 0.107 | 0.041 | 0.042 |

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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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

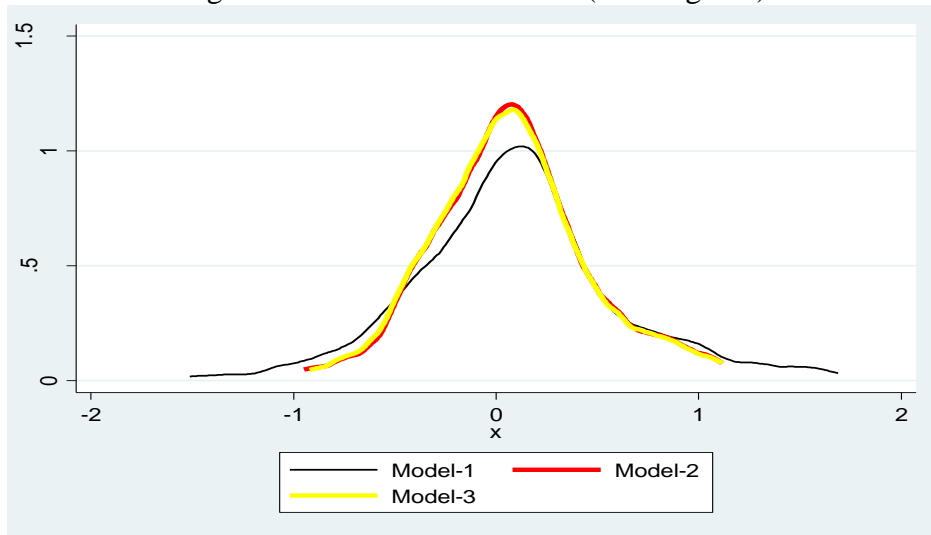
| | Standard error in | |
|----------------------------|-----------------------|--------------------|
| | Teacher fixed effects | Number of teachers |
| All subjects | 0.379 | 191 |
| Drop Biology | 0.376 | 179 |
| Drop Chemistry | 0.380 | 169 |
| Drop Computer Science | 0.419 | 172 |
| Drop English | 0.395 | 150 |
| Drop Environmental Science | 0.355 | 167 |
| Drop Hindi | 0.384 | 180 |
| Drop Mathematics | 0.394 | 168 |
| Drop Physics | 0.365 | 166 |
| Drop Economics | 0.380 | 184 |
| Drop Commerce | 0.378 | 184 |

Note: 1) The estimate in the first row is based on model (3) from Table 3, which include all ten subjects.

2) A student takes English as a mandatory subject while he/she can choose four or five subjects from rest of the nine.

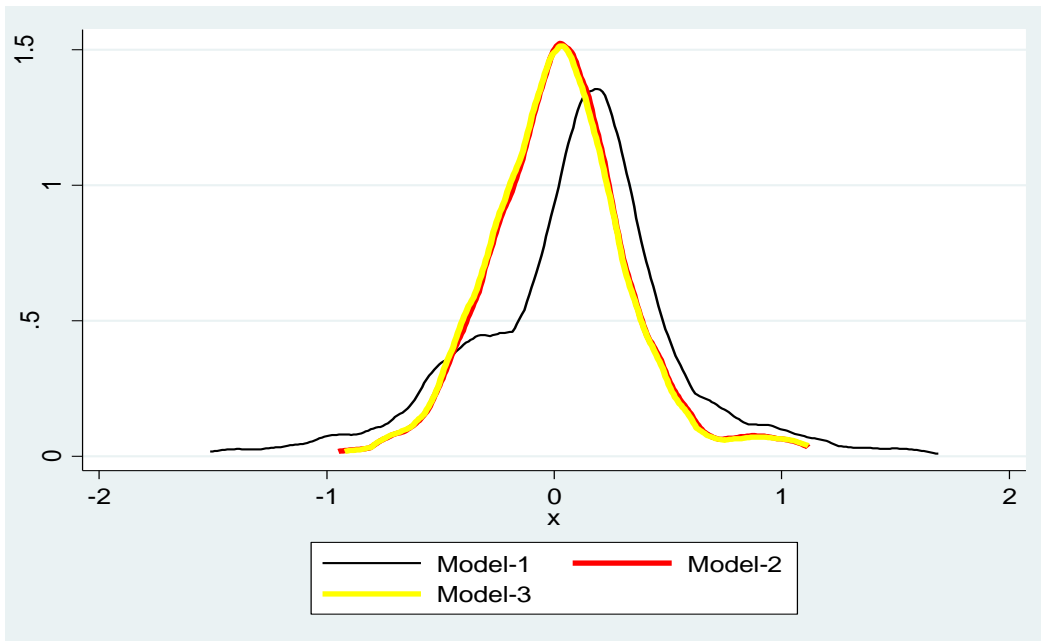
3) The estimates in the other rows (excluding first row) is also based on model (3) from Table 3, however one subject is dropped from the sample i.e. sample include all observation except observations from one subject.

Figure-1: Teacher fixed effects (un-weighted)



Note: The kernel density plots the teacher fixed effects of 191 teachers estimated from Model (1): No pupil effects, Model (2) Pupil fixed effects, Model (3) Pupil fixed effect and school fixed effect. See Table 3 for more details.

Figure-2: Teacher fixed effects (weighted by number of students)



Note: The kernel density plots the teacher fixed effects of 191 teachers estimated from Model (1): No pupil effects, Model (2) Pupil fixed effects, Model (3) Pupil fixed effect and school fixed effect. See Table 3 for more details.