Final report

Identifying an effective teacher in public schools in Delhi

International Growth Centre

Deepti Goel
Bidisha Barooah

DIRECTED BY

LSE

UNIVERSITY OF OXFORD

FUNDED BY

UKaid
from the British people

Identifying an Effective Teacher in Public Schools in Delhi

Deepti Goel and Bidisha Barooah[*]
August 2017

Abstract

We quantify the influence of teacher assignment on student test scores at the high-stakes grade twelve examination in state-run schools in Delhi, India. We find that being assigned to a one standard deviation better than average teacher moves a student at the middle of the achievement distribution to the 72nd percentile. The impact size is much higher than documented in existing literature, indicating that teachers play a much bigger role in the context we study. Consistent with earlier studies, standard observable characteristics such educational qualifications are unable to predict an effective teacher. For the first time we examine whether personality traits matter, and find that 'Openness', as opposed to being closed to experiences, positively correlates with a high quality teacher.

1 Introduction

Cognitive skills are an important component of human capital.[1] Higher cognitive skills lead to higher individual earnings (Hanushek, Schwerdt, Wiederhold and Woessmann 2015), and also boost a country's economic growth (Ciccone and Papaioannou 2009; Hanushek and Woessmann 2012), making it important to understand how they are formed. Such an understanding is especially crucial for India as Das and Zajonc (2010) find that the median Indian child failed to meet an international low benchmark in a standardized mathematics test, and the variance of test scores of Indian students was amongst the highest in the world.[2] Given that gaps in test scores explain corresponding gaps in college attainment rates and earnings (Chay, Guryan and Mazumder 2014; Johnson and Neal 1998), it is conceivable that differences in skill acquisition are partly responsible for India's economic inequality which has been rising since the early nineties (Basole and Basu 2015; Subramanian and Jayaraj 2015). Against this background, we study the process of cognitive skill formation at the higher secondary level (grade 12) in public (government) schools in Delhi, India. Specifically, we focus on the role of teachers in improving students' test scores, and also identify characteristics that make an effective teacher. We measure teacher effectiveness using the value added approach (Todd and Wolpin 2003), and accordingly define a high quality teacher as one who brings about higher test scores among her students after accounting for other current and past inputs, including prior teachers.[3]

A large body of research has confirmed that teachers matter in the sense that variation in teacher assignment explains a large part of the variation in students' test scores (Hanushek and Rivkin 2012 and Ladd 2008 review this literature). Most of this work is from high income countries, predominantly the United States:[4] Aaronson, Barrow and Sander 2007; Chetty, Friedman and Rockoff 2014; Clotfelter, Ladd and Vigdor 2010; Hanushek, Kain, O'Brien and Rivkin 2005; Rivkin, Hanushek and Kain 2005; and Rockoff 2004 are selected studies from the United States; Slater, Davies and Burgess 2012 from the United Kingdom; and Leigh 2010 from Australia. Of late, new research is emerging from middle income countries: Araujo, Carneiro, Cruz-Aguayo and Schady 2016 from Ecuador, Azam and Kingdon 2015 from India, Bau and Das 2017 and Talanché 2016 from Pakistan, and Metzler and Woessmann 2012 from Peru. We add to this nascent literature by being the first to examine the role of teachers in *public* schools in India.

Our work complements that by Azam and Kingdon (2015) who also examined teacher quality at the higher secondary level, but did so for a *private* consortium of schools in the Indian state of Uttar Pradesh. Even though the share of private schools in student enrolment is increasing in

---

[1] These skills are to be understood as raw problem solving abilities for abstract problems.

[2] Their study is based on 2005 data for ninth graders from two Indian states, namely, Orissa and Rajasthan. The Annual Status of Education Reports (ASER) provide a more recent picture of poor learning levels among 3-16-year olds in rural India. For example in 2016, only 43 percent of eight graders could correctly carry out a 3-digit by 1-digit division problem (ASER 2017).

[3] We define teacher quality solely in terms of his/ her ability to raise test scores. In doing so we are ignoring other important contributions that a teacher may make such as instilling curiosity and imparting a sense of civic responsibility.

[4] One reason for this is the existence of and easier access to large administrative datasets in these countries.

India, it is important to study government schools as they still account for a larger share,[5] and being subsidized by the state, cater mostly to students from weaker family backgrounds. Furthermore, government schools in India are not as productive as private schools as they result in test scores that are either the same (Chudgar or Quin 2012) or worse (Azam, Kingdon and Wu 2016; Muralidharan and Sundararaman 2015; Singh 2015), while their per-student expenditure is larger, mainly on account of higher teacher salaries (Pritchett and Aiyar 2015).[6, 7] Moreover, teachers in these schools have very different characteristics compared to their private counterparts: They are less likely to have a college degree, more likely to have a formal teacher training certificate, more likely to be absent, and less likely to be teaching when present (Desai, Dubey, Vanneman and Banerji 2008; Muralidharan and Kremer 2009). For these reasons, it is plausible that the two types of schools function very differently. It is therefore important to study them separately. The only other paper that has looked at teacher quality in public schools at the higher secondary level is Slater et al. 2012 for England. In the concluding section we compare our findings with both these related studies.

Finally, we go beyond the existing literature to examine whether certain characteristics, hitherto not looked at, can predict teacher effectiveness. The broad consensus is that most observable individual characteristics such as gender, experience, educational qualifications and training are not strongly associated with value added measures of teacher quality (Hanushek and Rivkin 2010).[8] For the first time we examine whether specific personality traits matter for teacher effectiveness. We do so by administering *The Big Five* test to teachers. The Big Five has emerged as the most widely accepted taxonomy of personality traits in psychology (John, Naumann and Soto 2008). Its personality dimensions are extraversion, agreeableness, conscientiousness, neuroticism and openness, each of which summarizes a large number of more specific personality traits.[9] As highlighted by Borghans, Duckworth, Heckman and ter Weel (2008a), these traits capture how people actually think, feel and behave and are sufficiently stable across situations to support the claim that latent traits exist. Research has shown that the Big Five predict important life outcomes: Conscientiousness predicts good health habits and

---

[5] At the all-India level, in 2014, 56 percent of students at the secondary and higher secondary levels (grades 10 through 12) studied in government schools. The corresponding figure for Delhi is much higher at 71 percent (National Sample Survey 2016).

[6] Pritchett and Aiyar (2015) estimate that, at the elementary level (grades 1 through 9), in 2011-12, the 'accounting cost' per student in a public school was Rs. 14,615 (USD 223), while in a private school it was only Rs. 5,961 (USD 91). Additionally, between 2009 and 2013, 70 to 80 percent of the elementary education budget in 6 states was spent on teacher salaries.

[7] All the studies cited here are based on analyses at the primary (grades 1 to 5) or secondary (grade 10) levels. We do not know of any work that compares public with private schools at the higher secondary level.

[8] A notable exception to this conclusion is Clotfelter et al. 2010. Additionally, some studies (Hanushek et al. 2005; Leigh 2010) have found that early experience adds to teacher effectiveness.

[9] Select trait adjectives associated with each dimension are: Talkative and assertive with extraversion; sympathetic and kind with agreeableness; organized and thorough with conscientiousness; tense and anxious with neuroticism; wide interests and imaginative with openness. See Table 4.4 in John et al. 2008 for an extensive list of trait adjectives. While there is broad consensus among psychologists over the five dimensions, there is still a wide disagreement over which traits correspond to each dimension.

longevity (Hampson and Friedman 2008), while low agreeableness and low conscientiousness predict delinquency in adolescents (Widiger and Smith 2008). Turning to work outcomes, psychologists have shown that conscientiousness is a general predictor of performance across a wide range of jobs, while other dimensions relate to specific jobs, for example, extraversion predicts success in sales and management positions (Barrick and Mount 1991; Barrick, Mount and Judge 2001). More recently, economists have begun to conceptualize (Almlund, Duckworth, Heckman and Kautz 2011; Borghans et al. 2008a) and test (Lindqvist and Vestman 2011; Mueller and Plug 2006) the importance of personality traits for labor market outcomes. Lindqvist and Vestman show that, compared to cognitive ability, personality traits are stronger predictors of labor force participation and earnings at the low end of the earnings distribution, while Mueller and Plug find that, in a group of highly educated individuals, the Big Five were as closely associated with earnings as was cognitive ability. Motivated by this evidence, and an intuitive expectation that traits such as gregariousness, empathy, self-discipline, irritability and idealism should influence the relationship between teachers and their students and thereby affect student test scores, we examine whether the Big Five can predict teacher effectiveness. We make an important contribution to the literature at the interface of psychology and economics by studying the importance of personality traits for job performance in the public sector where the link between performance and pay is tenuous at best.

Using data collected from 31 public higher secondary schools in Delhi for the academic year 2015-16 we find that, on average, being assigned to a one standard deviation better than average teacher improves student performance by 0.592 (or 0.639, depending on the specification chosen) standard deviations per subject. The subject-specific effects for English, Hindi and political science are 0.431, 0.391 and 0.419, respectively. Additionally, standard observational characteristics such as gender, educational qualifications, and experience do not correlate with value added measures of teacher effectiveness. On the other hand, 'Openness', one of the five personality dimensions of the Big Five taxonomy, positively predicts an effective teacher.

2 Empirical Framework

We model cognitive skill formation using the value added approach which mitigates the need to have data on historical inputs and endowments. Consider the following production function for test scores:

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

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

Estimating Teacher Effects

Following Todd and Wolpin (2003), we assume that there is a linear relationship between each input (T, X and θ) and the corresponding test score, the marginal impact of each (denoted below by β, γ, and δ, respectively) is age-invariant, and the impacts of all past inputs decay at a constant annual rate, $1 - \alpha$. Under these assumptions equation (1) can be written as:

$$A_{izt} = \beta T_{izt} + \alpha\beta T_{izt-1} + \cdots + \alpha^t \beta T_{iz0} + \gamma X_{izt} + \alpha\gamma X_{izt-1} + \cdots + \alpha^t \gamma X_{izo}$$
$$+\delta\theta_{it} + \alpha\delta\theta_{it-1} + \cdots + \alpha^t \delta\theta_{i0} + \varepsilon_{izt} \qquad (2)$$

Combining terms we get,

$$A_{izt} = \alpha A_{izt-1} + \beta T_{izt} + \gamma X_{izt} + \delta\theta_{it} + u_{izt} \qquad (3)$$

where $u_{izt} = \varepsilon_{izt} - \alpha\varepsilon_{izt-1}$.[10] Equation (3) is the value added specification wherein the effect of all past inputs is captured by the lagged test score, A_{izt-1}. Value addition due to current teacher quality, βT_{izt}, and current student ability, $\delta\theta_{it}$, are estimated using subject-specific teacher fixed effects,[11] and subject invariant student fixed effects, respectively. Introducing subscripts j and k for teacher and school, respectively, and suppressing the time subscript,[12] the fixed effects specification can be written as follows:

$$A^{12}_{izjk} = \alpha A^{10}_{iz} + \{\tau_{jz}\} + \gamma X_{iz} + \{\omega_i\} + u_{izjk} \qquad (4)$$

where $\{\tau_{jz}\}$, and $\{\omega_i\}$ are teacher and student fixed effects, respectively.

The standard deviation of teacher fixed effects is an estimator of the importance of teacher quality. Since teachers (and students) do not shift schools within the school year, it is not possible to separate teacher (student) effects from school effects. We therefore report within school variation in estimated teacher fixed effects, i.e. the standard deviation of $(\tau_{jz} - \bar{\tau}_{k(jz)})$, where $\bar{\tau}_{k(jz)}$ is the average over all teachers in school k to which teacher jz belongs to.[13] As pointed out by Slater et al. (2012), the within school estimator is a lower bound for the actual degree of teacher variation if, as is likely, teachers cluster in schools on the basis of their quality. Additionally, we assume that a teacher does not influence subject results other than her own. If this is violated, then our estimator of teacher effectiveness would be biased downwards.

---

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

[11] Note that any teacher who teaches multiple subjects has a distinct effect for each subject. We refer to our estimated effects as teacher effects, although they are more precisely subject-teacher effects.

[12] We have data for only a single year. Lack of multi-year data would be a serious limitation if our aim was to arrive at a ranking of individual teachers that was persistent over time as such a ranking has been shown to be unstable from one year to the next (McCaffrey, Sass, Lockwood and Mihaly 2009). However, assuming stationarity of the distribution of teacher effects, data for a single year is sufficient to characterize the overall distribution of teacher effects in any given year.

[13] We implement equation (4) in STATA using the `felsdvregdm' command. felsdvregdm deals with over parameterization by imposing sum to zero constraints within each school. School means are not separately identified in the presence of student fixed effects.

The inclusion of *contemporaneous* or point-in-time student fixed effects implies that individual teacher quality is derived from within student across subject variation in test scores at a point in time. To a large extent this addresses the concern of non-random matching of students and teachers. To fully address it we require that, having conditioned for lagged subject-specific scores, students should not be matched with teachers according to their *subject-specific* abilities.[14]

The vector X in equation (4) includes the following subject-specific inputs applied by the student: Interest (captured using a self-reported ranking of subjects according students' liking), study time outside school, and whether or not tuitions were taken. Some of these inputs may be responses to perceived or actual teacher quality. If this is the case and our interest is in the overall (policy) effect of a change in teacher quality, then these inputs should not be included as separate controls. Alternatively, we may be interested in production function parameters i.e. the effect of teacher quality keeping other inputs fixed; or it may be that the inputs in X are driven by unobserved student subject-specific heterogeneity such as parental motivation to excel in particular subjects. In both these alternative scenarios we would like to include student inputs as separate controls to avoid biased estimates. We therefore present two separate specifications, with and without X, which bound our estimator of teacher quality.[15]

Correcting for Sampling Variation

The variance of estimated teacher effects consists of the true variation in teacher effects plus sampling variation. We use bootstrap estimating procedure to obtain non-parametric estimates of the standard deviation of the variability of teacher effects.

Identifying an Effective Teacher

To identify who is an effective teacher, estimates of teacher fixed effects are regressed on the standard observable teacher characteristics and also on scores for each of the Big Five personality dimensions. Observable characteristics that we consider are gender, whether post graduate with a teaching degree, whether permanent/tenured, total teaching experience and teaching experience at the higher secondary level.

---

[14] Unlike student fixed effects in panel data models, the inclusion of point-in-time student fixed effects allows a student's subject invariant ability to vary over time. It also allows for dynamic tracking of students (Rothstein 2010) wherein students are matched with teachers according to the students' most recent ability measure. What we require is for this ability measure to be *subject-invariant* or, if it is subject-specific, then it be based on grade 10 subject scores. It would be problematic if students were matched to teachers according to unobserved expectation of subject specific performance, or according to revealed subject specific performance in grade 11. In our schools, the section of a student determines teacher assignments. Students either do not change sections between 11 and 12, or if they do, it is mostly because smaller sections are merged into larger ones. There does not seem to be selective matching of individual students with subject specific teachers in our schools.

[15] It is not possible to predict which specification would lead to a larger standard deviation of teacher effects. If the inputs in X are substitutes for (complements to) teacher quality, then controlling for them would give larger (smaller) estimates.

3 Data

Data come from 31 higher secondary schools that fall under the purview of the Directorate of Education (DOE), the education ministry of the Delhi Government. In the academic year 2015-16, there were 871 higher secondary schools under the DOE. In Appendix 1 we present some descriptive statistics for these schools. We also examine whether our sample is representative of all schools and do not find evidence to suggest otherwise.

We study a single cohort of students and their teachers, the former appeared for their higher secondary exams in March 2016.[16] Recall that our unit of observation for estimating teacher effects is a student-subject-grade-teacher match (see equation (4)).[17] Consequently, we restrict the analysis to only those student-subject combinations for whom we have both their grade 10 and grade 12 results, as well as a complete assignment of teachers who taught them in both grades 11 and 12. We lose observations due to missing section information, missing tenth scores, and incomplete time-tables for a small number of sections.[18] Consequently, while a total of 3,594 students appeared for the higher secondary exams from our study schools, we study a subset of 2,089 students. The analysis sample comprises of 20,682 student-subject-grade-teacher matches, arising from 394 subject-teachers, whose quality we intend to estimate. Note that these subject-teachers stem from 365 individual persons teaching in these schools.

At the higher secondary level there is a wide variety of subjects offered across Arts, Science, Commerce and Vocational streams. We restrict the analysis to 18 subjects which were taken by at least 30 students and for which we could find a matching subject in grade 10.[19]

The dependent variable, namely, a student's subject score in grade 12, is measured as a z-score. The z-score for a particular subject is constructed using that subject's mean and its standard deviation for all students in our study schools who appeared for the higher secondary examination in that subject.

As mentioned above we are unable to study all students from our study schools. It is therefore important to characterize who we are studying. Our analysis sample consists of students who did better than average at the higher secondary exams: As seen in Table 1, the overall mean (all

---

[16] These exams are offered at the all-India level by the Central Board of Secondary Education (CBSE), the largest national level board in the country. They are typically taken by 17-18 year olds and are ideal to study teacher effectiveness for the following reasons: (a) These are high-stakes exams that students (and their parents) take very seriously as they determine entry into institutions of higher education, including several elite institutions that offer highly sought after professional degrees in medicine and engineering; (b) They are based on material that the teachers are hired to teach; (c) They are set and graded outside the schools, thus minimizing the scope for manipulation of results.

[17] In equation (4) we did not use a separate subscript for grade, and clubbed grades 11 and 12 into one time period. For estimations, however, for each student-subject we have at least 2 observations, one each for grades 11 and 12.

[18] We are trying to recover at least some of these observations by re-visiting schools.

[19] Compared to the number of subjects offered grade 12, there are fewer subjects in grade 10. Consequently, it is not possible to have an exact subject match for each of the 18 subjects in grade 12. We matched each subject in grade 12 to what we thought was the closest offered subject in grade 10. These subject matches are listed in Table 1.

subjects clubbed together) z-score is 0.113, and it is statistically different from 0 at the 1 percent level. Table 1 also shows the subject-wise mean and standard deviation of z-scores for our analysis sample. It also lists the closest matched subject in grade 10, and the subject-wise mean score in grade 10.[20]

Table 1: Subject-wise Grade 12 Performance (Z-scores), Analysis Sample					
Subject in XIIth	Mean	Std. Dev.	No. of Students	Match in X[th]	Mean in X[th]
Accounting	0.077	0.992	1354	Mathematics	51.9
Agriculture	0.253	1.041	158	Science	43.3
Biology	0.219	0.862	235	Science	71.0
Business Studies	0.057	0.986	1319	Social Science	65.3
Chemistry	0.104	0.963	408	Science	70.7
Economics	0.040	0.993	2210	Social Science	63.6
English	0.133	0.970	4189	English	56.6
Engineering Graphics	-0.605	0.368	15	Mathematics	65.5
Geography	0.219	0.957	800	Social Science	51.0
Hindi	0.117	1.000	2971	Hindi	52.9
History	0.127	0.974	1833	Social Science	51.3
Home Science	0.079	0.926	764	Social Science	52.2
Mathematics	0.167	0.970	922	Mathematics	64.3
Physics	0.714	1.013	464	Science	71.8
Political Science	0.129	0.938	2414	Social Science	51.6
Psychology	-0.007	1.072	54	Social Science	62.6
Sanskrit	0.142	1.042	232	Sanskrit	46.2
Socio	0.126	1.009	340	Social Science	48.0
All 18 Subjects	0.113	0.977	20682		

Finally, the data on individual teacher characteristics is from a primary survey of teachers. The survey was voluntary and was completed by 261 out of the 365 teachers that make our analysis sample. Later on we check for selection by surveyed status of teachers.

The survey included the 44-item Big Five Inventory (BFI) constructed by John, Donahue and Kentle 1991 (see page 70 in John and Srivastava 1999 for the BFI instrument and its scoring sheet).[21] Against each item, the respondent is required to choose between one and five on a Likert scale indicating the extent to which he/she agrees or disagrees with the statement made. The respondent gets a score for each personality dimension where a higher score indicates

---

[20] We do not have numerical scores for grade 10. Instead we have letter grades which we converted to a numerical score using the mid-point of the range that each letter grade represents.
[21] Several variations of the BFI are available. The version we use has the advantage of retaining the brevity of single word adjectival items, while avoiding their ambiguity through the use of elaborative phrases (John et al. 2008).

greater strength of that dimension. A teacher's personality can evaluated only when he/she gave a response to all 44 items.

4   Results

Table 2 presents our estimates of teacher quality, namely, the standard deviation of estimated teacher effects. The first row present the naïve estimate where we do not include student fixed effects. The estimate is 0.463 and is not reliable as it may confound student effects with teacher effects. The remaining specifications include both teacher and student fixed effects. When none of the subject-specific student inputs are included, the estimate is 0.592 and when all three are included, it is 0.639. These are to be interpreted as follows: Having a one standard deviation better than average teacher, adds 0.592 (or 0.639 depending on the chosen specification) standard deviation per student per subject. This would move a student at the middle of the test score distribution to the 72nd percentile (74th percentile). Looking at the last two columns, it is clear that student-specific factors are relatively more important as they explain more than 5 times the variation in z-scores that is explained by variation in teacher assignments.

Another important observation from Table 2 is that inclusion of subject-specific student inputs increases the estimate of teacher effectiveness: From 0.592 to 0.639. Although this suggests that student inputs may be substitutes for teacher quality (see discussion in footnote 15), the magnitude of change is small in real terms as it makes a difference of 2 percentiles to student performance (72nd versus 74th).

Table 2: Estimates of Teacher Quality					
Specification	**Std. Dev. of TFE**	**Bootstrap Std. Err.**	**Number of Subject Teachers**	**Share explained by TFE**	**Share explained by SFE**
Without student fixed effects	0.463	0.009	394	14.2	
Both Teacher and Student FE; without X	0.592	0.009	394	12.1	58.3
Both Teacher and Student FE; only Tuitions	0.614	0.012	379	13.9	59.0
Both Teacher and Student FE; only Study time	0.626	0.013	379	14.2	59.2
Both Teacher and Student FE; only Interest	0.629	0.012	378	13.5	60.1
Both Teacher and Student FE; All three	0.639	0.013	378	13.9	60.2
Notes: TFEs and SFEs stand for Teacher and Student Fixed Effects, respectively					
Share (in %) explained by TFE refers to Cov(zscore12, TFE) / Var(Z-Score). Similarly, for SFE					

Although we have used z-scores for each subject, research shows that teacher effects differ by subject: Hanushek and Rivkin (2012) collate evidence from multiple studies for the United States and document an average standard deviation of teacher effectiveness of 0.13 for reading, and 0.17 for math. We therefore examine whether we also find subject-wise differences. For this we collect teacher effects by subject for the top three subjects by teacher count. Table 3 provides the estimates. Looking at the specification with all student inputs, having a one standard deviation better than average teacher increases the test scores by 0.479, 0.375 and 0.469 standard deviations for English, Hindi and political science, respectively. This would move a student at

the middle of the test score distribution in English, Hindi and political science to the 68th, 65th, and 68th percentiles, respectively.

Table 3: Subject-wise Estimates of Teacher Quality			
Specification	Std. Dev. of TFE	Bootstrap Std. Err.	Number of Teachers
English			
Without student fixed effects	0.461	0.017	51
Both Teacher and Student FE; without X	0.431	0.012	51
Both Teacher and Student FE; only Tuitions	0.453	0.017	51
Both Teacher and Student FE; only Study time	0.461	0.017	51
Both Teacher and Student FE; only Interest	0.473	0.014	51
Both Teacher and Student FE; All three	0.479	0.018	51
Hindi			
Without student fixed effects	0.401	0.025	49
Both Teacher and Student FE; without X	0.391	0.015	49
Both Teacher and Student FE; only Tuitions	0.367	0.021	49
Both Teacher and Student FE; only Study time	0.383	0.023	49
Both Teacher and Student FE; only Interest	0.362	0.020	49
Both Teacher and Student FE; All three	0.375	0.023	49
Political Science			
Without student fixed effects	0.406	0.037	45
Both Teacher and Student FE; without X	0.419	0.018	45
Both Teacher and Student FE; only Tuitions	0.448	0.031	43
Both Teacher and Student FE; only Study time	0.469	0.032	43
Both Teacher and Student FE; only Interest	0.436	0.030	43
Both Teacher and Student FE; All three	0.469	0.031	43
Note: TFE refers to Teacher Fixed Effects			
Bootstrap standard errors calculated using 84 replications			

Next, we examine whether we are able to predict an effective teacher using individual teacher characteristics. As mentioned earlier, we have data on these characteristics only for a subset of 261 out of 365 teachers who *volunteered* to fill our survey questionnaire. It is therefore important to check for systematic selection of teacher types by survey status. We find that there is no difference in mean value added between surveyed and non-surveyed teachers when using teacher effects arising from the specification without student inputs (mean surveyed: -0.025; mean non-surveyed: 0.050; p-value of 0.233). When using teacher effects from the specification with student inputs, the mean is statistically different at the 10 percent level of significance (mean surveyed: -0.040; mean non-surveyed: 0.080; p-value of 0.086). In sum, there is weak evidence that surveyed teachers are of lower quality. With this caveat in mind we examine whether we can predict an effective teacher among surveyed teachers.

Table 4 presents descriptive statistics on select characteristics for surveyed teachers in our analysis sample. Personality scores are only available for 202 teachers who completed all 44-items in the Big Five test.

Table 4: Descriptive Statistics, Surveyed Teachers		
	Mean (Std. Dev.)	No. of Teachers
Female	0.52	261
Post Graduate Trained	0.89	261
Permanent	0.82	261
Tot. Experience	20.4 (10.3)	258
Higher Sec. Experience	7.5 (5.6)	253
Extraversion	21.9 (3.4)	202
Agreeableness	28.9 (3.5)	202
Conscientiousness	25.0 (2.7)	202
Neuroticism	19.8 (3.0)	202
Openness	33.6 (3.8)	202

Table 5 presents the results on the correlates of an effective teacher.[22] Consistent with existing literature none of the observable characteristics are able to predict value added measures of teacher effectiveness. When we look at the five personality dimensions, 'openness' is positively associated with a higher quality teacher. Openness is contrasted against being closed to experiences. It captures the following personality traits: curious, imaginative, artistic, wide interests, excitable and unconventional. Given that children in public schools come from vulnerable backgrounds given their weaker economic status, it is not surprising that an unconventional teacher does better. Surprisingly, conscientiousness, which has been shown to be a strong predictor of performance in a wide variety of jobs, is not a significant trait in predicting an effective teacher. It could be that it is harder for a rule-based teacher to negotiate the complicated personal backgrounds of students in government schools and therefore such a teacher is not as effective in improving student performance.

---

[22] This analysis while having predictive power for teacher quality is not causal in nature. To estimate causal relationships one needs to address reverse causality between teacher quality and some of the characteristics such as personality traits.

Table 5: Predicting an Effective Teacher, Surveyed Teachers								
	Dependent var. TFEs, w/o Student Inputs				Dependent var. TFEs, w/ Student Inputs			
	Coeff.	OLS Std. Err.	Coeff.	OLS Std. Err.	Coeff.	OLS Std. Err.	Coeff.	OLS Std. Err.
Female	-0.090	0.057	-0.020	0.063	-0.091	0.060	-0.027	0.067
Post Graduate	-0.031	0.091	0.016	0.099	-0.034	0.096	0.047	0.106
Total Exp. 2-5	0.071	0.160	-0.007	0.178	0.001	0.168	-0.095	0.187
Total Exp. 5-15	0.144	0.168	0.157	0.189	0.022	0.176	-0.007	0.201
Total Exp. 15-30	0.009	0.173	0.013	0.193	-0.145	0.182	-0.168	0.205
Total Exp. > 30	0.233	0.174	0.168	0.197	0.054	0.183	-0.025	0.210
High. Sec. Exp. 2-5	0.034	0.108	0.092	0.121	0.033	0.114	0.082	0.129
High. Sec. Exp. 5-15	-0.099	0.112	-0.078	0.129	-0.103	0.118	-0.140	0.138
High. Sec. Exp. > 15	0.095	0.134	0.137	0.151	0.086	0.141	0.103	0.160
Permanent	0.055	0.131	0.066	0.144	0.190	0.138	0.218	0.151
Extraversion			-0.010	0.010			-0.015	0.010
Agreeableness			-0.008	0.010			-0.009	0.011
Conscientiousness			-0.021	0.013			-0.010	0.014
Neuroticism			0.008	0.011			-0.003	0.012
Openness			0.029	0.010			0.029	0.010
Subject dummies	Yes		Yes		Yes		Yes	
No. of Obs. (Teachers)	253		197		245		189	
Adjusted R squared	0.519		0.554		0.540		0.561	

5   Conclusions

We examined whether the teachers a student was assigned to matter for his/her test scores at the higher secondary level in government schools in Delhi. We find that teachers matter a lot. Over a two year period, namely, grades 11 and 12, being taught by a one standard deviation better teacher raises test score by 0.592 (or 0.639 when controlling for subject specific student inputs) standard deviations per subject. This would move a student at the middle of the achievement distribution to the 72nd (or 74th) percentile. These estimates measure *within* school variation in teacher quality and are therefore underestimates as they do not account for between school variations. We also estimated subject-specific teacher effects: 0.431 for English, 0.391 for Hindi and 0.419 for political science. To the best of our knowledge these are the first quantitative estimates of the importance of teachers in state-run schools in India.

Azam and Kingdon (2015) undertook a similar exercise for a consortium of private schools in the adjoining state of Uttar Pradesh, while Slater et al (2012) did so in public schools in England. Like ours, both these studies examined students at the higher secondary level. They find the standard deviation of teacher effects to be 0.366 and 0.233, respectively. They did not control for student inputs, so these estimates are to be compared with 0.592 in our study. A student at the middle of the achievement distribution would move to the 64th and the 59th percentile in their

contexts, compared to the 72nd percentile in ours. Thus, we have shown that teachers in Delhi's public schools play a much bigger role in improving student test scores than do teachers in other comparable contexts studied so far. That teachers play such an important role in state run schools is heartening. Good quality teachers can perhaps overcome some of the shortcomings that students face on account of their weaker socio-economic status. Given that test scores predict future earnings, this could in turn raise intergenerational economic mobility and reduce economic inequality at the macro-level. If this is a policy objective, greater emphasis must be placed on improving teacher quality in state run schools.

We also find that the estimate of teacher effectiveness drops from 0.639 to 0.592 when we do not account for subject-specific student inputs. While this indicates that, taken together, student-inputs tend to substitute teacher quality, the size of the effect is not large in real terms: It is a matter of moving from the 74th to the 72nd percentile of the student achievement distribution. This is again important from a policy perspective: We need not be concerned that students and their parents may cut back on their own inputs if teacher quality is raised due to state interventions.

Finally, ours is the first study to examine whether a person's personality matters for performing well as a teacher. We used the 'Big Five' to assess personalities. We find that an effective teacher is one who is 'Open', as opposed to someone who is conventional or closed to experiences. Consistent with earlier literature, standard observable characteristics such as teacher qualifications and experience do not predict teacher effectiveness. This questions the importance laid on these characteristics while hiring and promoting teachers in India. The factors that are rewarded at present are not the ones that matter for teacher quality. An important policy implication is to re-work compensation structures in public schools. The existing pay structure in public schools is consistent with a standard micro-economic model where a single wage results in a pooling equilibrium with good and bad types being hired. One solution to employ only the 'Open' types could be that part of a teacher's remuneration be based on his/her measured value addition to test scores. This would require the state to collect and maintain large administrative datasets, as is being done in some states in the United States. An important caveat is to avoid a system where the entire compensation is based on value addition as then there is the danger of teachers only 'teaching to the test'. Although we find that 'Openness' predicts teacher performance, it is hard to identify personality traits. Greater scrutiny is warranted at the time of hiring teachers, perhaps by involving psychologists during the interview process. This may help identify individuals with desirable personality traits.

References

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

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

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

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

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

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

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

Barrick, Murray R. and Michael K. Mount (1991): "The Big Five Personality Dimensions and Job Performance: A Meta-Analysis," *Personnel Psychology,* Vol 44, pp. 1-26.

Barrick, Murray R., Michael K. Mount, Timothy A. Judge (2001): "Personality and Performance at the beginning of the New Millennium: What do we know and where do we go next?" *Personality and Performance,* Vol 9, No 1/2, pp. 9-30.

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

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

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

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

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

Ciccone, Antonio and Elias Papaioannou (2009): "Human Capital, The Structure of Production, and Growth," *The Review of Economics and Statistics*, Vol 91, No 1, pp 66-82.

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

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

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

Hampson, Sarah E. and Howard S. Friedman (2008): "Personality and Health: A Lifespan Perspective, " In Oliver P. John, Richard W. Robins and Lawrence A. Pervin (Eds.) Handbook of Personality: Theory and Research 3rd Edition, pp 770-794, New York, NY, US: Guilford Press.

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

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

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

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

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

John, Oliver P., Donahue, E. M. and Kentle, R. L. (1991): "The Big Five Inventory-Versions 4a and 54," Berkeley: University of California at Berkeley, Institute of Personality and Social Research.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Subramanian, S and D Jayaraj (2015): "Growth and Inequality in the Distribution of India's Consumption Expenditure," Working paper 2015/025, UNU-WIDER.

Talanché, Marine de (2016): "Better Teachers, Better Results? Evidence from Rural Pakistan," *The Journal of Development Studies,* doi:10.1080/00220388.2016.1265944

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

Widiger, Thomas A. and Gregory T. Smith (2008): "Personality and Psychopathology," In Oliver P. John, Richard W. Robins and Lawrence A. Pervin (Eds) Handbook of Personality: Theory and Research 3rd Edition, pp 743-769, New York, NY, US: Guilford Press.

Appendix 1

School Selection

In this section we examine whether our sample of 31 schools is representative of the population of 871 higher secondary schools under the DOE. Figure 1 presents a map of Delhi showing the sampled schools. We followed convenience sampling based on school location and as seen in the map there is some clustering of schools. It is therefore important to check for selectivity.
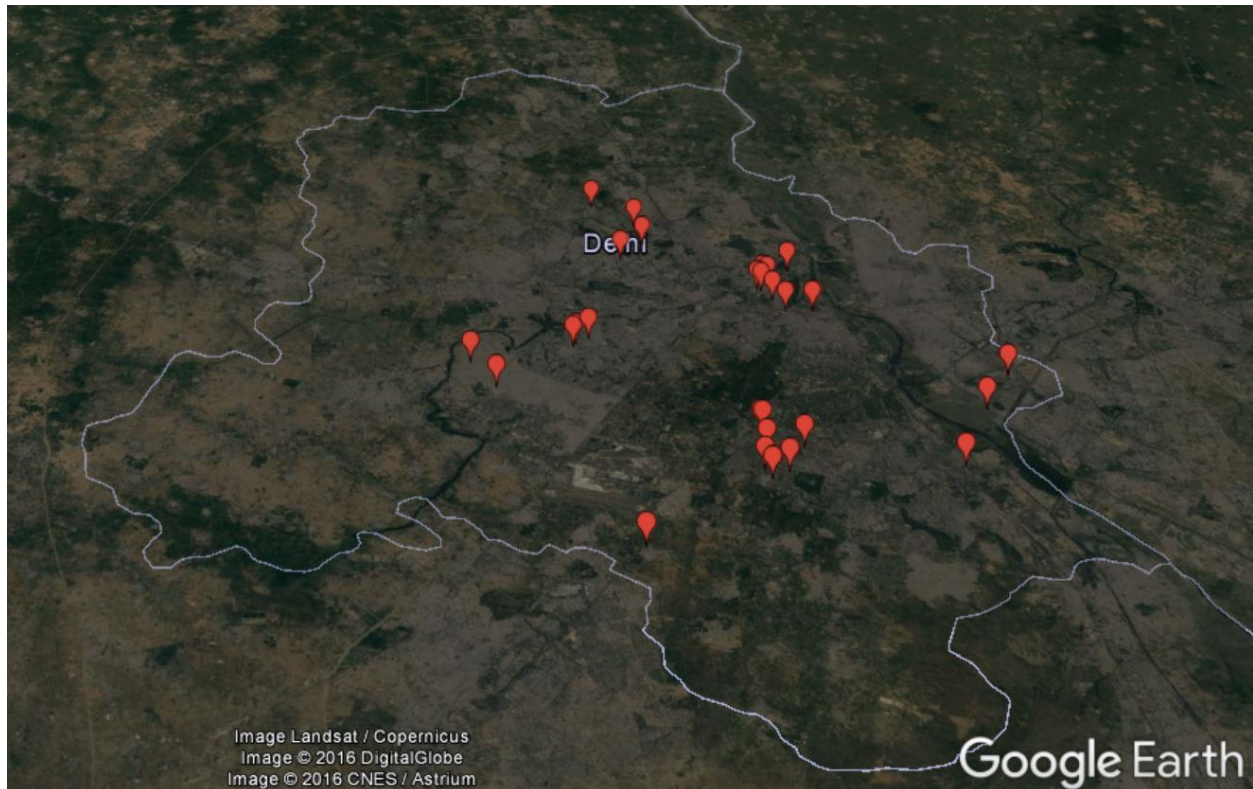

Table A1 shows select descriptive statistics by sampling status. At 10 percent level of significance, it would seem that we oversampled morning shift schools, smaller schools, those offering the science stream, and those with better test scores in the vocational stream. We test for selection using 30 characteristics. The Bonferroni corrected critical value for this number of hypotheses at the 10 percent level of significance is 0.0033. All the p-values in Table A1 are above this cut off. Thus, the evidence cannot reject the claim that our sample is representative.

Table A1.1: Descriptive Statistics by Sampling Status: Higher Secondary Public Schools, 2015-16					
	All Schools	Sampled	Not Sampled	Difference	P-value[1]
Number of Schools	871	31	840		
Gender; Shares in Percent					
Girls	43.7	35.5	44.1	-8.6	0.3363
Boys	41.2	45.2	41.1	4.1	0.6583
Co-ed	15.0	19.4	14.9	4.5	0.5411
Shift; Shares in Percent					
Morning	69.9	87.1	69.3	17.8	0.0050
Evening	30.1	12.9	30.7	-17.8	
Mean Number of Students	150.8 (111.6)	118.1 (65.7)	152.0 (112.8)	-33.9	0.096
Streams Taught; Shares in Percent					
Has Arts	99.1	96.8	99.2	-2.4	0.4606
Has Commerce	69.4	77.4	69.1	8.3	0.2834
Has Science	30.8	48.4	30.1	18.3	0.0488
Has Vocational	28.9	41.9	28.5	13.4	0.1407
Pupil Teacher Ratio	26.5 (14.7)	24.2 (10.5)	26.6 (14.8)	-2.4	0.383
Pupil Classroom Ratio	39.4 (17.6)	36.9 (15.9)	39.5 (17.6)	-2.6	0.425
Mean Pass Percentage					
Arts	89.6 (12.8)	91.0 (11.6)	89.5 (12.9)	1.5	0.540
Commerce	91.6 (12.1)	93.2 (7.8)	91.6 (12.3)	1.6	0.532
Science	95.2 (9.9)	93.6 (12.1)	95.3 (9.8)	-1.7	0.524
Vocational	85.1 (17.6)	87.0 (22.6)	85.0 (17.3)	2.0	0.693
Mean Marks per Student					
Arts	285.6 (32.0)	294.9 (28.4)	285.2 (32.1)	9.7	0.106
Commerce	296.2(37.6)	303.4 (29.1)	295.9 (37.9)	7.5	0.339
Science	336.3 (37.4)	338.2 (35.7)	336.2 (37.5)	2.0	0.845
Vocational	291.9 (29.2)	306.8 (32.5)	291.1 (29.6)	15.7	0.065
Marks per Student, 25th Percentile					
Arts	265.4	276.0	265.2	10.8	0.193
Commerce	275.1	280.6	274.9	5.7	0.462
Science	316.5	335.1	315.2	19.9	0.179
Vocational	276.6	287.3	275.7	11.6	0.290
Marks per Student, Median					
Arts	285.3	294.0	284.4	9.6	0.103
Commerce	295.2	303.5	295.1	8.4	0.345
Science	338.2	341.5	337.6	3.9	0.670
Vocational	293.6	312.7	293.1	19.6	0.034
Marks per Student, 75th Percentile					
Arts	304.4	312.2	303.8	8.4	0.301
Commerce	317.6	325.2	316.7	8.5	0.427
Science	355.7	360.3	355.6	4.7	0.720
Vocational	308.7	331.6	307.2	24.4	0.004
Source: Directorate of Education, Government of NCT of Delhi					
Standard Deviation in Parentheses. P-values are for the null hypotheses that the difference is zero.					

The International Growth Centre
(IGC) aims to promote sustainable
growth in developing countries
by providing demand-led policy
advice based on frontier research.

Find out more about
our work on our website
www.theigc.org

For media or communications
enquiries, please contact
mail@theigc.org

Subscribe to our newsletter
and topic updates
www.theigc.org/newsletter

Follow us on Twitter
@the_igc

Contact us
International Growth Centre,
London School of Economic
and Political Science,
Houghton Street,
London WC2A 2AE

IGC

**International
Growth Centre**