Measuring the Impacts of Teachers I:
Evaluating Bias in Teacher Value-added Estimates

By Raj Chetty, John N. Friedman, and Jonah E. Rockoff

Online Appendix

Online Appendix A: Value-Added Estimation Methods

In this appendix, we provide a step-by-step guide to implementing our method of estimating VA in the presence of drift. In practice, we cannot follow exactly the method described in Section 2.2 because data availability varies across teachers. For instance, there are different numbers of students per class and teachers have a different number of past and future classes from which to construct value-added in any given year. We calculate value-added in three steps, separately for each subject (math and English) and school level (elementary and middle).

Step 1 [Residualization of Test Scores]: We begin by residualizing student scores A_{it}^* with respect to controls X_{it} by running an OLS regression with teacher fixed effects of the form

$$A_{it}^* = \alpha_j + \beta X_{it}$$

and constructing residuals

$$A_{it} = A_{it}^* - \hat{\beta} X_{it}.$$

Step 2 [Estimation of Variance Components]: Next, we estimate the individual-level variance of residual test scores, $\sigma_\varepsilon^2 = Var(\varepsilon_{it})$, as

$$\hat{\sigma}_\varepsilon^2 = MSE * \left(\frac{N-1}{N-K-C+1} \right)$$

where MSE is the variance of the within-classroom deviations of A_{it}, N is the total number of students, C is the total number of classrooms, and K is the number of control variables in the X_{it} control vector. The scaling term is required to correct the degrees of freedom for the fact that we have already estimated K parameters to form the residual A_{it}. We also estimate $Var(A_{it})$, the total variance of A_{it}, again accounting for the prior estimation of $\hat{\beta}$ when calculating the degrees of freedom correction.

At this point, we collapse the data to the classroom-level by constructing the average residualized score $\bar{A}_{ct}$ for each classroom c and proceed to use class-level means for the remaining steps. In middle school, teachers teach more than one class per year. We handle such cases by collapsing the data to the teacher-year level. We do so by constructing precision-weighted averages of classroom-average scores within a teacher-year. The weight for classroom c in year t is

$$h_{ct} = \frac{1}{\hat{\sigma}_\theta^2 + \frac{\hat{\sigma}_\varepsilon^2}{n_{ct}}},$$

where $\hat{\sigma}_\theta^2$ is an estimate of the class-level variance component and n_{ct} denotes the number of students in the classroom. We construct this estimate as $\hat{\sigma}_\theta^2 = Var(A_{it}) - \hat{\sigma}_\varepsilon^2 - \hat{\sigma}_{A0}$, where $\hat{\sigma}_{A0}$ is our estimate of the within-teacher-year between-class covariance in average scores, reflecting the teacher-level component of the variance. To simplify computation, we follow Kane and Staiger (2008) and randomly sort classrooms within each teacher-year cell; we then estimate the covariance $\hat{\sigma}_{A0}$ based on the covariance of the test scores of adjacent classrooms in each teacher-year cell, weighting each pair of classrooms by the sum of students taught.

We next estimate the covariances between mean scores across years within teacher, denoted $\hat{\sigma}_{As}$, in both elementary and middle schools. We allow a separate covariance for each possible time lag $s \in \{1, 2, \ldots\}$ denoting the separation between the two years in which the classes were taught. We weight each teacher-year pair by the sum of students taught. We set all covariances for lags greater than 7 to $\hat{\sigma}_{A7}$, the estimated covariance for the 7th lag.

Step 3: [Construction of VA Estimates] In this step, we use the parameter estimates to construct a VA estimate for each teacher j in each year t that she appears in the data. We depart from the method described in Section 2.2 by using data from all other years – not just years before year t – to increase the precision of our VA estimates.[42] Let $\overrightarrow{A}_j^{-t}$ denote the vector of teacher-year-mean scores used to predict teacher j's VA in year t. Let N_{jt} denote the length of this vector, so that we are using N_{jt} other years to project scores in year t. We construct the best linear predictor of teacher quality in year t as

$$\widehat{\mu}_{jt} = \left(\Sigma_{A_{jt}}^{-1}\gamma_{jt}\right)' \overrightarrow{A}_j^{-t}$$

where γ_{jt} is a $N_{jt} \times 1$ vector and $\Sigma_{A_{jt}}$ is a $N_{jt} \times N_{jt}$ matrix. We denote the weights on scores $\overrightarrow{A}_j^{-t}$ by $\psi_{jt} = \Sigma_{A_{jt}}^{-1}\gamma_{jt}$. If the mth and nth element of the scores vector $\overrightarrow{A}_j^{-t}$ are A_{js} and $A_{js'}$, the mth element of the diagonal of $\Sigma_{A_{jt}}$ in middle school is

$$\left[\Sigma_{A_{jt}}\right]_{mm} = \hat{\sigma}_{A0} + \frac{1}{\sum_{c\in\{c:j(c)=j\}} h_{cs}},$$

where the denominator of the second term is the sum of precisions for the classes taught by a teacher in year s, which is the precision of the teacher-year mean in year s. In elementary school, where teachers teach one class per year, we cannot estimate $\hat{\sigma}_{A0}$ but we can estimate $\hat{\sigma}_{A0} + \hat{\sigma}_\theta^2 = Var(A_{it}) - \hat{\sigma}_\varepsilon^2$. Here, the mth element of the diagonal of $\Sigma_{A_{jt}}$ is

$$\left[\Sigma_{A_{jt}}\right]_{mm} = (\hat{\sigma}_{A0} + \hat{\sigma}_\theta^2) + \frac{\hat{\sigma}_\varepsilon^2}{n_{ct}}.$$

In both elementary and middle school, the mnth off-diagonal element of $\Sigma_{A_{jt}}$ is

$$\left[\Sigma_{A_{jt}}\right]_{mn} = \hat{\sigma}_{A,|s-s'|}$$

and the mth element of γ_{jt} is

$$\left[\gamma_{jt}\right]_m = \hat{\sigma}_{A,|t-s|}.$$

Because the distribution of other years in which data are available varies both across teachers j and across the years t within a teacher, both the matrix $\Sigma_{A_{jt}}$ and the vector γ_{jt} will vary across j and t. We therefore construct these elements separately for each teacher-year in the data. Note that we can use this algorithm even if data on test scores for teacher j's students are missing in year t, since those data are not required to estimate $\widehat{\mu}_{jt}$.

Online Appendix B: Teacher-Level Bias

In this appendix, we define an alternative notion of bias in VA estimates, which we term "teacher-level bias," and characterize its relationship to the concept of forecast bias that we focus on in the text.[43] For simplicity, we follow Rothstein (2009) and focus on the case without drift in teacher value-added.[44] In this case, $\mu_{jt} = \mu_j$ in all periods and our estimator for teacher VA

[42] Using data from other years increases precision not just by increasing sample size but also because we have more data from nearby years. For example, data from year $t+1$ are more informative for VA in year t than data from year $t-2$ in the presence of drift.

[43] We thank Jesse Rothstein for drawing our attention to the distinction between teacher-level bias and forecast bias.

[44] Drift complicates the asymptotics because additional information from prior years does not eliminate estimation error in expected VA.

simplifies to

$$\widehat{\mu}_{jt} = \bar{A}_j^{-t} \frac{\sigma_\mu^2}{\sigma_\mu^2 + (\sigma_\theta^2 + \sigma_{\tilde{\varepsilon}}^2/n)/(t-1)}, \tag{17}$$

as shown in (9). Let $\widehat{\mu}_j^* = \lim_{t\to\infty} \widehat{\mu}_{jt}$ denote the value to which the VA estimate for teacher j converges as the number of classrooms observed approaches infinity. The asymptotic bias in the estimate of teacher j's VA is

$$\omega_j = \widehat{\mu}_j^* - \mu_j. \tag{18}$$

Definition 2. Value-added estimates are *unbiased at the teacher-level* if $Var(\omega_j) = 0$.

VA estimates are biased at the teacher-level if they are inconsistent, i.e. if we systematically mispredict a given teacher's performance when estimation error in VA vanishes. Such teacher-level bias is relevant for determining whether a value-added model treats all teachers equitably.

Forecast vs. Teacher-Level Bias. To see the connection between teacher-level bias and forecast bias, consider an experiment in which students are randomly assigned to teachers in year t. By the definition of forecast bias,

$$\begin{aligned} 1 - B\left(\widehat{\mu}_j^*\right) &= \frac{Cov\left(A_{it}, \widehat{\mu}_j^*\right)}{Var(\widehat{\mu}_j^*)} \\ &= \frac{Var\left(\mu_j\right) + Cov\left(\mu_j, \omega_j\right)}{Var\left(\mu_j\right) + Var\left(\omega_j\right) + 2Cov\left(\mu_j, \omega_j\right)} \end{aligned}$$

where the second step follows because $Cov\left(A_{it} - \mu_j, \widehat{\mu}_j^*\right) = 0$ under random assignment in year t. Hence,

$$B\left(\widehat{\mu}_j^*\right) = 0 \Leftrightarrow Var(\omega_j) + Cov\left(\mu_j, \omega_j\right) = 0.$$

This identity has two implications. First, if VA estimates are unbiased at the teacher-level, they must also be forecast-unbiased: $Var(\omega_j) = 0 \Rightarrow B(\widehat{\mu}_j^*) = 0$. Second, and more importantly for our application, forecast-unbiased VA estimates can be biased at the teacher level only if $Cov\left(\mu_j, \omega_j\right) = -Var(\omega_j)$. Intuitively, if the teacher-level bias ω_j is negatively correlated with true value-added, then the covariance of VA estimates with true scores is reduced, but the variance of VA estimates also falls. If the two forces happen to cancel out exactly, $B(\widehat{\mu}_j^*)$ could be 0 even if $Var(\omega_j) > 0$. In this sense, if a pre-specified value-added model produces VA estimates $\widehat{\mu}_{jt}$ that exhibit no forecast bias, the existence of teacher-level bias is a measure-zero (knife-edge) case.

Note that estimating the degree of forecast bias is simpler than teacher-level bias because forecast bias can be directly estimated using finite-sample estimates of $\widehat{\mu}_{jt}$ without any additional inputs. In contrast, estimating teacher-level bias requires accounting for the impacts of estimation error on $\widehat{\mu}_{jt}$ to construct the limit $\widehat{\mu}_j^*$, which is a non-trivial problem, particularly in the presence of drift.

Online Appendix C: Matching Algorithm

We follow the matching algorithm developed in Chetty et al. (2011) to link the school district data to tax records. The algorithm was designed to match as many records as possible using variables that are not contingent on ex post outcomes. Date of birth, gender, and last name in the tax data are populated by the Social Security Administration using information that is not contingent on ex post outcomes. First name and ZIP code in tax data are contingent on observing some ex post outcome. First name data derive from information returns, which are typically generated after an adult outcome like employment (W-2 forms), college attendance (1098-T forms), or mortgage interest payment (1098 forms). The ZIP code on the claiming parent's 1040 return is typically from 1996 and is thus contingent on the ex post outcome of the student not having moved far from her elementary school for most students in our analysis sample.

Chetty et al. (2011) show that the match algorithm outlined below yields accurate matches for approximately 99% of cases in a school district sample that can be matched on social security number. Note that identifiers were used solely for the matching procedure. After the match was completed, the data were de-identified (i.e., individual identifiers such as names were stripped) and the statistical analysis was conducted using the de-identified dataset.

Step 1 [Date of Birth, Gender, Last Name]: We begin by matching each individual from the school-district data to Social Security Administration (SSA) records. We match individuals based on exact date of birth, gender, and the first four characters of last name. We only attempt to match individuals for which the school records include a valid date of birth, gender, and at least one valid last name. SSA records all last names ever associated in their records with a given individual; in addition, there are as many as three last names for each individual from the school files. We keep a potential match if any of these three last names match any of the last names present in the SSA file.

Step 2 [Rule Out on First Name]: We next check the first name (or names) of individuals from the school records against information from W2 and other information forms present in the tax records. Since these files reflect economic activity usually after the completion of school, we use this information in Step 2 only to "rule out" possible matches in order to minimize selection bias. In particular, we disqualify potential matches if none of the first names on the information returns match any of the first names in the school data. As before, we use only the first four characters of a first name. For many potential matches, we find no first name information in the tax information records; at this step we retain these potential matches. After removing potential matches that are mismatched on first name, we isolate students for whom only one potential match remains in the tax records. We declare such cases a match and remove them from the match pool. We classify the match quality (MQ) of matches identified at this stage as $MQ = 1$.

Step 3 [Dependent ZIP code]: For each potential match that remains, we find the household that claimed the individual as a dependent (if the individual was claimed at all) in each year. We then match the location of the claiming household, identified by the 5-digit ZIP code, to the home address ZIP code recorded in the school files. We classify potential matches based on the best ZIP code match across all years using the following tiers: exact match, match within 10 (e.g., 02139 and 02146 would qualify as a match), match within 100, and non-match. We retain potential matches only in the highest available tier of ZIP code match quality. For example, suppose there are 5 potential matches for a given individual, and that there are no exact matches on ZIP code, two matches within 10, two matches within 100, and one non-match. We would retain only the two that matched within 10. After this procedure, we isolate students for whom only one potential match remains in the tax records. We declare such cases a match and remove them from the match pool. We classify the match quality of matches identified at this stage as $MQ = 2$.

Step 4 [Place of Birth]: For each potential match that remains, we match the state of birth from the school records with the state of birth as identified in SSA records. We classify potential matches into three groups: state of birth matches, state of birth does not match but the SSA state is the state where the school district is, and mismatches. Note that we include the second category primarily to account for the immigrants in the school data for whom the recorded place of birth is outside the country. For such children, the SSA state-of-birth corresponds to the state in which they received the social security number, which is often the first state in which they lived after coming to the country. We retain potential matches only in the best available tier of place-of-birth match quality. We then isolate students for whom only one potential match remains in the tax records. We declare such cases a match and remove them from the match pool. We classify the match quality of matches identified at this stage as $MQ = 3$.

Step 5 [Rule In on First Name]: After exhausting other available information, we return to the first name. In step 2 we retained potential matches that either matched on first name or for which there was no first name available. In this step, we retain only potential matches that match on first name, if such a potential match exists for a given student. We also use information on first name present on 1040 forms filed by potential matches as adults to identify matches at this stage. We then isolate students for whom only one potential match remains in the tax records. We declare such cases a match and remove them from the match pool. We classify the match quality of matches identified at this stage as $MQ = 4$.

Step 6 [Fuzzy Date-of Birth]: In previous work (Chetty et al. 2011), we found that 2-3% of individuals had a reported date of birth that was incorrect. In some cases the date was incorrect only by a few days; in others the month or year was off by one, or the transcriber transposed the month and day. To account for this possibility, we take all individuals for whom no eligible matches remained after step 2. Note that if any potential matches remained after step 2, then we would either settle on a unique best match in the steps that follow or find multiple potential matches even after step 5. We then repeat step 1, matching on gender, first four letters of last name, and fuzzy date-of-birth. We define a fuzzy DOB match as one where the absolute value of the difference between the DOB reported in the SSA and school data was in the set $\{1, 2, 3, 4, 5, 9, 10, 18, 27\}$ in days, the set $\{1, 2\}$ in months, or the set $\{1\}$ in years. We then repeat steps 2 through 5 exactly as above to find additional matches. We classify matches found using this fuzzy-DOB algorithm as $MQ = 5.X$, where X is the corresponding MQ from the non-fuzzy DOB algorithm. For instance, if we find a unique fuzzy-DOB match in step 3 using dependent ZIP codes, then $MQ = 5.2$.

The following table shows the distribution of match qualities for all students. We match 88.6% of students and 89.8% of student-subject observations in the analysis sample used to calculate VA. Unmatched students are split roughly evenly among those for whom we found multiple matches and those for whom we found no match.

Match Quality (MQ)	Frequency	Percent	Cumulative Match Rate
1	650002	47.55%	47.55%
2	511363	37.41%	84.95%
3	24296	1.78%	86.73%
4	10502	0.77%	87.50%
5.1	14626	1.07%	88.57%
5.2	779	0.06%	88.63%
5.3	96	0.01%	88.63%
5.4	31	0.01%	88.64%
Multiple Matches	75010	5.49%	
No Matches	80346	5.88%	
Total	1367051		88.64%

Online Appendix D: Unconditional Sorting of Students to Teachers

In this appendix, we assess the unconditional relationship between teacher VA and student observables to determine whether high VA teachers are systematically assigned to certain types of students (see Section 5.3). We are able to study such unconditional sorting because our method of constructing student test score residuals in (5) only exploits within-teacher variation. Prior studies that estimate VA typically construct test score residuals using both between- and within-teacher variation and thus do not necessarily obtain a global ranking. For example, suppose schools with higher SES students have better teachers. By residualizing test scores with respect to student SES before computing teacher VA, one would attribute the differences in outcomes across these schools to differences in student SES rather than teacher quality. As a result, one only obtains a relative ranking of teachers conditional on student SES and cannot compare teacher quality across students with different characteristics. Using within-teacher variation to estimate the coefficients on the control vector X_{it} resolves this problem and yields a global ranking of teachers across the school district.

To estimate unconditional sorting of students to teachers based on observable characteristics X_{it}, one would ideally regress teacher VA μ_{jt} on X_{it}:

$$\mu_{jt} = \alpha + \rho X_{it} + \eta_{it}. \tag{19}$$

Since true VA is unobserved, we substitute VA estimates $\widehat{\mu}_{jt}$ for μ_{jt} on the left hand side of (19). This yields an attenuated estimate of ρ because $\widehat{\mu}_{jt}$ is shrunk toward 0 to account for estimation error (see Section 2.2). If all teachers taught the same number of classes and had the same number of students, the shrinkage factor would not vary across observations. In this case, we could identify ρ by using $\widehat{\mu}_{jt}$ as the dependent variable in (19) and multiplying the estimate of ρ by $SD(\mu_{jt})/SD(\widehat{\mu}_{jt})$. In the sample for which we observe lagged test scores, the standard deviation of teacher VA estimates is $SD(\mu_{jt})/SD(\widehat{\mu}_{jt}) = 1.56$. We therefore multiply the estimate of ρ obtained from estimating (19) with $\widehat{\mu}_{jt}$ as the dependent variable by 1.56. This simple approach to correcting for the attenuation bias is an approximation because the shrinkage factor does vary across observations. However, our estimates of the magnitudes of unconditional sorting are small and hence further adjusting for the variation in shrinkage factors is unlikely to affect our conclusions.

We report estimates of unconditional sorting in Appendix Table 2. Each column reports estimates of an OLS regression of VA estimates $\widehat{\mu}_{jt}$ on various observables (multiplied by 1.56), with standard errors clustered at the teacher level to account for correlated errors in the assignment process of classrooms to teachers.

We begin in Column 1 by regressing $\widehat{\mu}_{jt}$ on lagged test scores $A^*_{i,t-1}$. Better students are assigned slightly better teachers: students who score 1 unit higher in the previous grade get a teacher whose VA is 0.0122 better on average. The tracking of better students to better teachers magnifies gaps in achievement, although the magnitude of this amplification effect is small relative to other determinants of the variance in student achievement.

Column 2 shows that special education students are assigned teachers with 0.003 lower VA on average. Again, this effect is statistically different from zero, but is quantitatively small. Relative to other students with similar prior test scores, special education students receive slightly *higher* VA teachers (not reported).

In Column 3, we regress $\widehat{\mu}_{jt}$ on parent income. A \$10,000 (0.3 SD) increase in parent income raises teacher VA by 0.00084, with the null hypothesis of 0 correlation rejected with $p < 0.0001$. Column 4 demonstrates that controlling for a student's lagged test score $A^*_{i,t-1}$ entirely eliminates the correlation between teacher VA and parent income.

Column 5 analyzes the correlation between teacher VA and ethnicity. Mean teacher quality is no different on average across minority (Hispanic or Black) vs. non-minority students.

Finally, Columns 6 and 7 analyze the relationship between teacher value-added and school-level demographics. The relationship between mean parent income in a school and teacher quality remains quite small (Column 6) and there is no relationship between fraction minority and school quality.

Finally, we assess the extent to which differences in teacher quality contribute to the gap in achievement by family income. In the sample used to estimate VA, a $10,000 increase in parental income is associated with a 0.065 SD increase in 8th grade test scores (averaging across math and English). To calculate how much smaller this gradient would be if teacher VA did not vary with parent income, we must take a stance on how teachers' impacts cumulate over time. In our companion paper, we estimate that 1 unit improvement in teacher VA in a given grade raises achievement by approximately 0.53 units after 1 year, 0.36 after 2 years, and stabilizes at approximately 0.25 after 3 years (Chetty, Friedman, and Rockoff 2014, Appendix Table 10). Under the assumption that teacher effects are additive across years, these estimates of fade-out imply that a 1 unit improvement in teacher quality in all grades K-8 would raise 8th grade test scores by 3.4 units. Using the estimate in Column 3 of Appendix Table 2, it follows that only $\frac{3.4\times 0.00084}{0.065} = 4\%$ of the income-score gradient can be attributed to differences in teacher quality from grades K-8. However, a sequence of good teachers can close a significant portion of the achievement gap. If teacher quality for low income students were improved by 0.1 units in all grades from K-8, 8th grade scores would rise by 0.34, enough to offset more than a $50,000 difference in family income.

Online Appendix E: Quasi-Experimental Estimator of Forecast Bias

This appendix shows that estimating (15) using OLS identifies the degree of forecast bias under Assumption 3. Recall that we define the degree of forecast bias $B = 1 - \lambda$ based on the best linear predictor of A_{it} in a randomized experiment in year t:

$$E^*[A_{it}|1, \widehat{\mu}_{jt}] = \alpha_t + \lambda\widehat{\mu}_{jt}.$$

In the observational data, we can decompose test scores in year t into the effect of teacher VA $E^*[A_{it}|1, \widehat{\mu}_{jt}]$ and student-level errors χ_{it}:

$$A_{it} = \alpha_t + \lambda\widehat{\mu}_{jt} + \chi_{it},$$

where χ_{it} may be correlated with $\widehat{\mu}_{jt}$ because of non-random assignment. Taking averages over all the students in a school-grade cell and first-differencing gives the quasi-experimental specification in (15):

$$\Delta\overline{A}_{sgt} = \alpha + \lambda\Delta Q_{sgt} + \Delta\chi_{sgt},$$

where Q_{sgt} is the mean of $\widehat{\mu}_{jt}$ in school s in grade g in year t. It follows immediately that estimating (15) using OLS yields an unbiased estimate of λ under Assumption 3 ($\Delta\chi_{sgt}$ orthogonal to ΔQ_{sgt}).

References

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126 (4): 1593–1660.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* forthcoming.

Kane, Thomas J., and Douglas O. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." National Bureau of Economic Research Working Paper 14607.

Rothstein, Jesse. 2009. "Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables." *Education Finance and Policy* 4 (4): 537–71.

APPENDIX TABLE 1
Structure of Analysis Dataset

Student	Subject	Year	Grade	Class	Teacher	Test Score	Matched to Tax Data?	Parent Income
Bob	Math	1992	4	1	Jones	0.5	1	$95K
Bob	English	1992	4	1	Jones	-0.3	1	$95K
Bob	Math	1993	5	2	Smith	0.9	1	$95K
Bob	English	1993	5	2	Smith	0.1	1	$95K
Bob	Math	1994	6	3	Harris	1.5	1	$95K
Bob	English	1994	6	4	Adams	0.5	1	$95K
Nancy	Math	2002	3	5	Daniels	0.4	0	.
Nancy	English	2002	3	5	Daniels	0.2	0	.
Nancy	Math	2003	4	6	Jones	-0.1	0	.
Nancy	English	2003	4	6	Jones	0.1	0	.

Notes: This table illustrates the structure of the core sample, which combines information from the school district database and the tax data. There is one row for each student-subject-school year. Students who were not linked to the tax data have missing data on parent characteristics. The values in this table are not real data and are for illustrative purposes only.

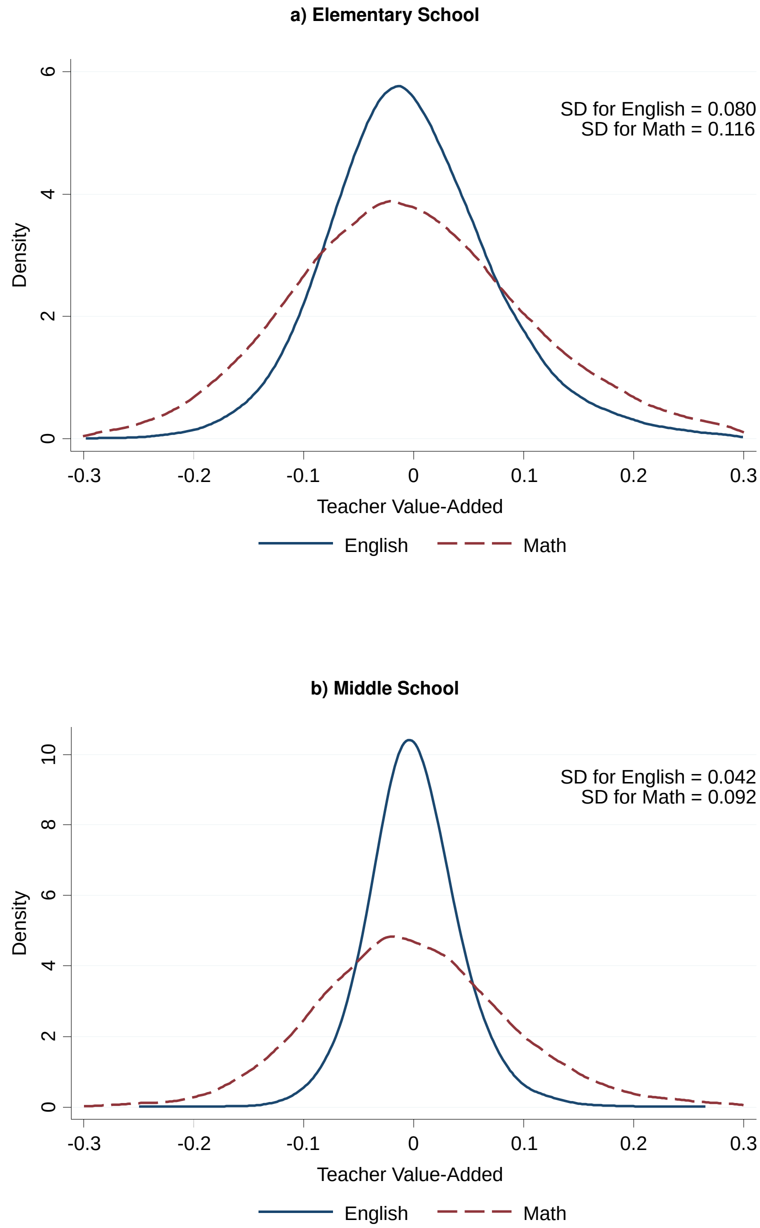
APPENDIX TABLE 2
Differences in Teacher Quality Across Students and Schools

Dependent Variable:	Teacher Value-Added						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lagged Test Score	0.0122			0.0123			
	(0.0006)			(0.0006)			
Special education student		-0.003					
		(0.001)					
Parent Income ($10,000s)			0.00084	0.00001			
			(0.00013)	(0.00011)			
Minority (black or hispanic) student					-0.001		
					(0.001)		
School Mean Parent Income ($10,000s)						0.0016	
						(0.0007)	
School Fraction Minority							0.003
							(0.003)
Observations	6,942,979	6,942,979	6,094,498	6,094,498	6,942,979	6,942,979	6,942,979

Notes: Each column reports coefficients from an OLS regression, with standard errors clustered by teacher in parentheses. Teacher VA, which is the dependent variable in all columns, is scaled in units of student test score standard deviations. Teacher VA is estimated using data from classes taught by the same teacher in other years, following the procedure in Sections II.B and 4 and using the baseline control vector (see notes to Table 3 for more details). The regressions are run at the student-subject-year level on the sample used to estimate the baseline VA model. We multiply the resulting regression coefficients by 1.56 to account for the attenuation bias due to using VA estimates instead of true VA as the dependent variable (see Appendix D for details). Columns 3 and 4 restrict the sample to students whom we are able to link to parents in the tax data. Each specification includes the student-level covariate(s) listed at the left hand side of the table and no additional control variables. See notes to Table 1 for definitions of these independent variables. In Columns 6-7, the independent variable is the school-mean of the independent variables in Columns 3 and 5, respectively. We calculate these means as the unweighted mean across all student-subject-year observations with non-missing data for the relevant variable in each school.

APPENDIX FIGURE 1

Empirical Distributions of Teacher VA Estimates



Notes: This figure plots kernel densities of the empirical distribution of teacher VA estimates $\hat{\mu}_{jt}$ for each subject (math and English) and school-level (elementary and middle school). The densities are weighted by the number of student test score observations used to construct the teacher VA estimate and are estimated using a bandwidth of 0.01. We also report the standard deviations of these empirical distributions of VA estimates. Note that these standard deviations are smaller than the standard deviation of true teacher effects reported in Table 2 because VA estimates are shrunk toward the mean to account for noise and obtain unbiased forecasts.