Measuring the Impacts of Teachers I:
Evaluating Bias in Teacher Value-Added Estimates[†]

By RAJ CHETTY, JOHN N. FRIEDMAN, AND JONAH E. ROCKOFF*

Are teachers' impacts on students' test scores (*value-added*) *a good measure of their quality? One reason this question has sparked debate is disagreement about whether value-added* (*VA*) *measures provide unbiased estimates of teachers' causal impacts on student achievement. We test for bias in VA using previously unobserved parent characteristics and a quasi-experimental design based on changes in teaching staff. Using school district and tax records for more than one million children, we find that VA models which control for a student's prior test scores provide unbiased forecasts of teachers' impacts on student achievement.* (*JEL* H75, I21, J24, J45)

How can we measure and improve the quality of teaching in primary schools? One prominent but controversial method is to evaluate teachers based on their impacts on students' test scores, commonly termed the *value-added* (VA) approach.[1] School districts from Washington, DC to Los Angeles have begun to calculate VA measures and use them to evaluate teachers. Advocates argue that selecting teachers on the basis of their VA can generate substantial gains in achievement (e.g., Gordon, Kane, and Staiger 2006; Hanushek 2009), while critics contend that VA measures are poor proxies for teacher quality (e.g., Baker et al. 2010; Corcoran 2010). The debate about teacher VA stems primarily from two questions. First, do the differences in test score

[1] Value-added models of teacher quality were pioneered by Hanushek (1971) and Murnane (1975). More recent examples include Rockoff (2004); Rivkin, Hanushek, and Kain (2005); Aaronson, Barrow, and Sander (2007); and Kane and Staiger (2008).

gains across teachers measured by VA capture causal impacts of teachers or are they biased by student sorting? Second, do teachers who raise test scores improve their students' outcomes in adulthood or are they simply better at teaching to the test?

This paper addresses the first of these two questions.[2] Prior work has reached conflicting conclusions about the degree of bias in VA estimates (Kane and Staiger 2008; Rothstein 2010; Kane et al. 2013). Resolving this debate is critical for policy because biased VA measures will reward or penalize teachers systematically based on the mix of students in their classrooms.

We develop new methods to estimate the degree of bias in VA estimates and implement them using information from two administrative databases. The first is a dataset on test scores and teacher assignments in grades 3–8 from a large urban school district in the United States. These data cover more than 2.5 million students and include over 18 million tests of math and English achievement spanning 1989–2009. We match 90 percent of the observations in the school district data to selected data from US tax records spanning 1996–2011. These data contain information on parent characteristics such as household income, retirement savings, and mother's age at child's birth.

We begin our analysis by constructing VA estimates for the teachers in our data. We predict each teacher's VA in a given school year based on the mean test scores of students she taught in other years. We control for student characteristics such as prior test scores and demographic variables when constructing this prediction to separate the teacher's impact from (observable) student selection. Our approach to estimating VA parallels closely that used by school districts currently, except in one respect. Typically, existing value-added models assume that each teacher's quality is fixed over time and thus place equal weight on test scores in all classes taught by the teacher when forecasting teacher quality. In practice, test scores from more recent classes are better predictors of current teacher quality, indicating that teacher quality fluctuates over time. We account nonparametrically for such *drift* in teacher quality by estimating the autocovariance of scores across classrooms taught by a given teacher; intuitively, we regress scores in year t on average scores in other years, allowing the coefficients to vary across different lags. Our VA model implies that a 1 standard deviation (SD) improvement in teacher VA raises normalized test scores by approximately 0.14 SD in math and 0.1 SD in English, slightly larger than the estimates in prior studies which do not account for drift.

Next, we turn to our central question: are the VA measures we construct *unbiased* predictors of teacher quality? To define *bias* formally, consider a hypothetical experiment in which we assign students to teachers randomly, as in Kane and Staiger (2008). Let λ denote the mean test score impact of being randomly assigned to a teacher who is rated one unit higher in VA based on observational data from prior school years. We define the degree of *forecast bias* in a VA model as $B = 1 - \lambda$. We say that VA estimates are forecast unbiased if $B = 0$: i.e., if teachers whose estimated VA is one unit higher do in fact cause students' test scores to increase by one unit on average.

---

[2] We address the second question in our companion paper (Chetty, Friedman, and Rockoff 2014). There are also other important concerns about VA. Most importantly, as with other measures of labor productivity, the signal in value-added measures may be degraded by behavioral responses if high-stakes incentives are put in place (Barlevy and Neal 2012).

We develop two methods to estimate the degree of forecast bias in VA estimates. First, we estimate forecast bias based on the degree of selection on observable characteristics excluded from the VA model. We generate predicted test scores for each student based on parent characteristics from the tax data, such as family income, and regress the predicted scores on teacher VA. For our baseline VA model—which controls for a rich set of prior student-, class-, and school-level scores and demographics—we find that forecast bias from omitting parent characteristics is at most 0.3 percent at the top of the 95 percent confidence interval. Using a similar approach, we find that forecast bias from omitting twice-lagged scores from the VA model is at most 2.6 percent.

In interpreting these results, it is important to note that children of higher-income parents do get higher-VA teachers on average. However, such sorting does not lead to biased estimates of teacher VA for two reasons. First, and most importantly, the correlation between VA estimates and parent characteristics vanishes once we control for test scores in the prior school year. Second, even the unconditional correlation between parent income and VA estimates is small: we estimate that a $10,000 increase in parent income is associated with less than a 0.0001 SD improvement in teacher VA (measured in student test score SDs).[3] One explanation for why sorting is so limited is that 85 percent of the variation in teacher VA is within rather than between schools. Since most sorting occurs through the choice of schools, parents may have little scope to steer their children toward higher-VA teachers.

While our first approach shows that forecast bias due to sorting on certain observable predictors of student achievement is minimal, bias due to other unobservable characteristics could still be substantial. To obtain a more definitive estimate of forecast bias which accounts for unobservables, we develop a quasi-experimental analog to the ideal experiment of random student assignment. Our quasi-experimental design exploits teacher turnover at the school-grade level for identification. To understand the design, suppose a high-VA fourth grade teacher moves from school A to another school in 1995. Because of this staff change, fourth graders in school A in 1995 will have lower-VA teachers on average than the previous cohort of students in school A. If VA estimates have predictive content, we would expect fourth grade test scores for the 1995 cohort to be lower on average than the previous cohort.

Using event studies of teacher arrivals and departures, we find that mean test scores change sharply across cohorts as predicted when very high- or low-VA teachers enter or exit a school-grade cell. We estimate the amount of forecast bias by comparing changes in average test scores across consecutive cohorts of children within a school to changes in the mean value-added of the teaching staff. The forecasted changes in mean scores closely match observed changes. The point estimate of forecast bias in our preferred specification is 2.6 percent and is not statistically distinguishable from 0. The upper bound on the 95 percent confidence interval for the degree of bias is 9.1 percent.

---

[3] An auxiliary implication of this result is that differences in teacher quality explain a small share of the achievement gap between high- and low-SES (socioeconomic status) students. This is not because teachers are unimportant—one could close most of the achievement gap by assigning highly effective teachers to low-SES students—but rather because teacher VA does not differ substantially across schools in the district we study.

Our quasi-experimental design rests on the identification assumption that high-frequency teacher turnover within school-grade cells is uncorrelated with student and school characteristics. This assumption is plausible insofar as parents are unlikely to switch their children to a different school immediately simply because a single teacher leaves or arrives. Moreover, we show that changes in mean teacher quality in a given subject (e.g., math) are uncorrelated with both prior scores in that subject and *contemporaneous* scores in the other subject (e.g., English), supporting the validity of the research design.

We investigate which of the controls in our baseline VA model are most important to account for student sorting by estimating forecast bias using our quasi-experimental design for several commonly used VA specifications. We find that controlling simply for a student's own lagged test scores generates a point estimate of forecast bias of 5 percent that is not significantly different from 0. In contrast, models which omit lagged test score controls generate forecast bias exceeding 40 percent. Thus, most of the sorting of students to teachers relevant for future test achievement is captured by prior test scores. This result is reassuring for the application of VA models because virtually all value-added models used in practice control for prior scores.

Our quasi-experimental method provides a simple, low-cost tool for assessing bias in various settings. For instance, Kane, Staiger, and Bacher-Hicks (2014) apply our method to data from the Los Angeles Unified School District (LAUSD). They find that VA estimates which control for lagged scores also exhibit no forecast bias in LAUSD, even though the dispersion in teacher VA is much greater in LAUSD than in the district we study. More generally, the methods developed here could be applied to assess the accuracy of personnel evaluation metrics in a variety of professions beyond teaching.

Our results reconcile the findings of experimental studies (Kane and Staiger 2008; Kane et al. 2013) with Rothstein's (2010) findings on bias in VA estimates. We replicate Rothstein's finding that there is small but statistically significant grouping of students on lagged test score gains and show that this particular source of selection generates minimal forecast bias. Based on his findings, Rothstein warns that selection on unobservables *potentially* could generate substantial bias. We evaluate directly the degree of forecast bias due to unobservables using a quasi-experimental analog of Kane and Staiger's (2008) experiment. Like Kane and Staiger, we find no evidence of forecast bias due to unobservables. Hence, we conclude that VA estimates which control for prior test scores exhibit little bias despite the grouping of students on lagged gains documented by Rothstein.

The paper is organized as follows. In Section I, we formalize how we construct VA estimates and define concepts of bias in VA estimates. Section II describes the data sources and provides summary statistics. We construct teacher VA estimates in Section III. Sections IV and V present estimates of forecast bias for our baseline VA model using the two methods described above. Section VI compares the forecast bias of alternative VA models. Section VII explains how our results relate to findings in the prior literature. Section VIII concludes.

I. Conceptual Framework and Methods

In this section, we develop an estimator for teacher VA and define formally our measure of bias in VA estimates. We begin by setting up a simple statistical model of test scores.

A. *Statistical Model*

School principals assign each student i in school year t to a classroom $c = c(i, t)$. Principals then assign a teacher $j(c)$ to each classroom c. For simplicity, assume that each teacher teaches one class per year, as in elementary schools. Let $j = j(c(i, t))$ denote student i's teacher in year t and μ_{jt} represent the teacher's value-added in year t (i.e., teacher j's impact on test scores). We scale teacher VA so that the average teacher has value-added $\mu_{jt} = 0$ and the effect of a one-unit increase in teacher VA on end-of-year test scores is 1.

Student i's test score in year t, A_{it}^*, is given by

$$(1) \qquad\qquad A_{it}^* = \beta \mathbf{X}_{it} + \nu_{it},$$

$$(2) \qquad \text{where} \qquad \nu_{it} = \mu_{jt} + \theta_c + \tilde{\varepsilon}_{it}.$$

Here, $\mathbf{X}_{it}$ denotes observable determinants of student achievement, such as lagged test scores and family characteristics. We decompose the error term ν_{it} into three components: teacher value-added μ_{jt}, exogenous class shocks θ_c, and idiosyncratic student-level variation $\tilde{\varepsilon}_{it}$. Let $\varepsilon_{it} = \theta_c + \tilde{\varepsilon}_{it}$ denote the unobserved error in scores unrelated to teacher quality. Student characteristics $\mathbf{X}_{it}$ and ε_{it} may be correlated with μ_{jt}. Accounting for such selection is the key challenge in obtaining unbiased estimates of μ_{jt}.

The model in (1) permits teacher quality μ_{jt} to fluctuate stochastically over time. We do not place any restrictions on the stochastic processes that μ_{jt} and ε_{it} follow except for the following assumption.

ASSUMPTION 1 (Stationarity): *Teacher VA and student achievement follow a stationary process:*

$$(3) \quad E[\mu_{jt}|t] = E[\varepsilon_{it}|t] = 0, \quad \mathrm{cov}(\mu_{jt}, \mu_{j,t+s}) = \sigma_{\mu s}, \quad \mathrm{cov}(\varepsilon_{it}, \varepsilon_{i,t+s}) = \sigma_{\varepsilon s}$$

$$\textit{for all} \quad t.$$

Assumption 1 requires that (i) mean teacher quality does not vary across calendar years and (ii) the correlation of teacher quality, class shocks, and student shocks across any pair of years depends only on the amount of time which elapses between those years. This assumption simplifies the estimation of teacher VA by reducing the number of parameters to be estimated. Note that the variance of teacher effects, $\sigma_\mu^2 = \mathrm{var}(\mu_{jt})$, is constant across periods under stationarity.

B. *Estimating Teacher Value-Added*

We develop an estimator for teacher value-added in year t (μ_{jt}) based on mean test scores in prior classes taught by teacher j.[4] Our approach parallels closely existing estimators for value-added (e.g., Kane and Staiger 2008), except that it accounts for drift in teacher quality over time. To simplify exposition, we derive the estimator for the case in which data on test scores is available for t years for all teachers, where all classes have n students, and where each teacher teaches one class per year. In online Appendix A, we provide a step-by-step guide to implementation (along with corresponding Stata code) which accounts for differences in class size, multiple classrooms per year, and other technical issues which arise in practice.

We construct our estimator in three steps. First, we regress test scores A_{it}^{*} on $\mathbf{X}_{it}$ and compute test score residuals adjusting for observables. Next, we estimate the best linear predictor of mean test score residuals in classrooms in year t based on mean test score residuals in prior years, using a technique analogous to an OLS (ordinary least squares) regression. Finally, we use the coefficients of the best linear predictor to predict each teacher's VA in year t. We now describe these steps formally.

Let the residual student test score after removing the effect of observable characteristics be denoted by

$$(4) \qquad\qquad A_{it} \;=\; A_{it}^{*} \;-\; \boldsymbol{\beta}\mathbf{X}_{it} \;=\; \mu_{jt} \;+\; \varepsilon_{it}.$$

We estimate β using variation across students taught by the same teacher using an OLS regression of the form

$$(5) \qquad\qquad A_{it}^{*} \;=\; \alpha_{j} \;+\; \boldsymbol{\beta}\mathbf{X}_{it},$$

where α_{j} is a teacher fixed effect. Our approach of estimating $\boldsymbol{\beta}$ using within-teacher variation differs from prior studies, which typically use both within- and between-teacher variation to estimate $\boldsymbol{\beta}$ (e.g., Kane, Rockoff, and Staiger 2008; Jacob, Lefgren, and Sims 2010). If teacher VA is correlated with $\mathbf{X}_{it}$, estimates of $\boldsymbol{\beta}$ in a specification without teacher fixed effects overstate the impact of the $\mathbf{X}$s because part of the teacher effect is attributed to the covariates.[5] For example, suppose $\mathbf{X}$ includes school fixed effects. Estimating $\boldsymbol{\beta}$ without teacher fixed effects would attribute all the test score differences across schools to the school fixed effects, leaving mean teacher quality normalized to be the same across all schools. With school fixed effects, estimating $\boldsymbol{\beta}$ within teacher requires a set of teachers to teach in multiple schools, as in Mansfield (2013). These switchers allow us to identify the school

---

[4]To maximize statistical precision, we use data from all other years—both in the past and future—to predict VA in year t in our empirical implementation. To simplify notation, we present the derivation in this section for the case in which we only use prior data to predict VA.

[5]Teacher fixed effects account for correlation between $\mathbf{X}_{it}$ and mean teacher VA. If $\mathbf{X}_{it}$ is correlated with fluctuations in teacher VA across years due to drift, then one may still understate teachers' effects even with fixed effects. We show in Table 6 below that dropping teacher fixed effects when estimating (5) yields VA estimates which have a correlation of 0.98 with our baseline estimates because most of the variation in $\mathbf{X}_{it}$ is within classrooms. Since sorting to teachers based on their average impacts turns out to be quantitatively unimportant in practice, sorting based on fluctuations in those impacts is likely to have negligible effects on VA estimates.

fixed effects independent of teacher effects and obtain a cardinal global ranking of teachers across schools.

Let $\overline{A}_{jt} = \frac{1}{n} \sum_{i \in \{i : j(i,t)=j\}} A_{it}$ denote the mean residual test score in the class teacher j teaches in year t. Let $\mathbf{A}_j^{-t} = (\overline{A}_{j1}, \ldots, \overline{A}_{j,t-1})'$ denote the vector of mean residual scores prior to year t in classes taught by teacher j. Our estimator for teacher j's VA in year t $(E[\mu_{jt} | \mathbf{A}_j^{-t}])$ is the best linear predictor of $\overline{A}_{jt}$ based on prior scores $(E^*[\overline{A}_{jt} | \mathbf{A}_j^{-t}])$, which we write as

$$\hat{\mu}_{jt} = \sum_{s=1}^{t-1} \psi_s \overline{A}_{js}.$$

We choose the vector of coefficients $\psi = (\psi_1, \ldots, \psi_{t-1})'$ to minimize the mean-squared error of the forecasts of test scores:

$$(6) \qquad \psi = \underset{\{\psi_1, \ldots, \psi_{t-1}\}}{\arg \min} \sum_j \left(\overline{A}_{jt} - \sum_{s=1}^{t-1} \psi_s \overline{A}_{js} \right)^2.$$

The resulting coefficients ψ are equivalent to those obtained from an OLS regression of $\overline{A}_{jt}$ on $\mathbf{A}_j^{-t}$. In particular, $\psi = \Sigma_A^{-1} \gamma$ where $\gamma = (\text{cov}(\overline{A}_{jt}, \overline{A}_{j1}), \ldots, \text{cov}(\overline{A}_{jt}, \overline{A}_{j,t-1}))'$ is the vector of auto-covariances of mean test scores for classes taught by a given teacher and Σ_A is the variance-covariance (VCV) matrix of $\mathbf{A}_j^{-t}$. The diagonal elements of Σ_A are the variance of mean class scores σ_A^2. The off-diagonal elements are the covariance of mean test scores of different classes taught by a given teacher $\text{cov}(\overline{A}_{jt}, \overline{A}_{j,t-s})$. Note that $\text{cov}(\overline{A}_{jt}, \overline{A}_{j,t-s}) = \sigma_{As}$ depends only on the time lag s between the two periods under the stationarity assumption in (3).

Finally, using the estimates of ψ, we predict teacher j's VA in period t as

$$(7) \qquad \hat{\mu}_{jt} = \psi' \mathbf{A}_j^{-t}.$$

Note that the VA estimates in (7) are leave-year-out (jackknife) measures of teacher quality, similar to those in Jacob, Lefgren, and Sims (2010), but with an adjustment for drift. This is important for our analysis below because regressing student outcomes in year t on teacher VA estimates without leaving out the data from year t when estimating VA would introduce the same estimation errors on both the left- and right-hand side of the regression and produce biased estimates of teachers' causal impacts.[6]

Special Cases.—The formula for $\hat{\mu}_{jt}$ in (7) nests two special cases which help build intuition for the general case. First, consider predicting a teacher's impact

[6] This is the reason that Rothstein (2010, p. 199) finds that, "...fifth grade teachers whose students have above average fourth grade gains have systematically lower estimated value-added than teachers whose students underperformed in the prior year." Students who had unusually high fourth grade gains due to idiosyncratic, nonpersistent factors (e.g., measurement error) will tend to have lower than expected fifth grade gains, making their fifth grade teacher have a lower VA estimate.

in year t using data from only the previous year $t - 1$. In this case, $\gamma = \sigma_{A,1}$ and $\Sigma_A^{-1} = \frac{1}{\sigma_A^2}$. Hence, ψ simplifies to $\frac{\sigma_{A,1}}{\sigma_A^2}$ and

$$\hat{\mu}_{jt} = \psi \overline{A}_{j,t-1}. \tag{8}$$

The shrinkage factor ψ in this equation incorporates two forces which determine ψ in the general case. First, because past test scores are a noisy signal of teacher quality, the VA estimate is shrunk toward the sample mean ($\mu_{jt} = 0$) to reduce mean-squared error. Second, because teacher quality drifts over time, the predicted effect differs from past performance. For instance, if teacher quality follows a mean-reverting process, past test scores are further shrunk toward the mean to reduce the influence of transitory shocks to teacher quality.

Second, consider the case where teacher quality is fixed over time ($\mu_{jt} = \mu_j$ for all t) and the student- and class-level errors are i.i.d. This is the case considered by most prior studies of value-added. Here, $\text{cov}(\overline{A}_{jt}, \overline{A}_{j,t-s}) = \text{cov}(\mu_j, \mu_j) = \sigma_\mu^2$ for all $s \neq t$ and $\sigma_A^2 = \sigma_\mu^2 + \sigma_\theta^2 + \sigma_{\tilde{\varepsilon}}^2/n$. In this case, (7) simplifies to

$$\hat{\mu}_{jt} = \overline{A}_j^{-t} \frac{\sigma_\mu^2}{\sigma_\mu^2 + (\sigma_\theta^2 + \sigma_{\tilde{\varepsilon}}^2/n)/(t-1)}, \tag{9}$$

where $\overline{A}_j^{-t}$ is the mean residual test score in classes taught by teacher j in years prior to t and $\psi = \dfrac{\sigma_\mu^2}{\sigma_\mu^2 + (\sigma_\theta^2 + \sigma_{\tilde{\varepsilon}}^2/n)/(t-1)}$ is the *reliability* of the VA estimate. This formula coincides with equation (5) in Kane and Staiger (2008) in the case with constant class size.[7] Here, the signal-to-noise ratio ψ does not vary across years because teacher performance in any year is equally predictive of performance in year t. Because years are interchangeable, VA depends purely on mean test scores over prior years, again shrunk toward the sample mean to reduce mean-squared error.[8]

Importantly, $\hat{\mu}_{jt} = E^*[\overline{A}_{jt} | \mathbf{A}_j^{-t}]$ represents simply the best linear predictor of the future test scores of students assigned to teacher j in observational data. This prediction does not measure necessarily the expected causal effect of teacher j on students' scores in year t, $E[\mu_{jt} | \mathbf{A}_j^{-t}]$, because part of the prediction could be driven by systematic sorting of students to teachers. We now turn to evaluating the degree to which $\hat{\mu}_{jt}$ measures a teacher's causal impact.

---

[7] Kane and Staiger (2008) derive (9) using an Empirical Bayes approach instead of a best linear predictor. If teacher VA, class shocks, and student errors follow independent Normal distributions, the posterior mean of μ_{jt} coincides with (9). Analogously, (7) can be interpreted as the posterior expectation of μ_{jt} when teacher VA follows a multivariate Normal distribution whose variance-covariance matrix controls the drift process.

[8] In our original working paper (Chetty, Friedman, and Rockoff 2011b), we estimated value-added using (9). Our qualitative conclusions were similar, but our out-of-sample forecasts of teachers' impacts were attenuated because we did not account for drift.

C. *Definition of Bias*

An intuitive definition of bias is to ask whether VA estimates $\hat{\mu}_{jt}$ predict differences accurately in the mean test scores of students who are randomly assigned to teachers in year t, as in Kane and Staiger (2008). Consider an OLS regression of residual test scores A_{it} in year t on $\hat{\mu}_{jt}$ (constructed from observational data in prior years) in such an experiment:

$$(10) \qquad A_{it} = \alpha_t + \lambda \hat{\mu}_{jt} + \zeta_{it}.$$

Because $E[\varepsilon_{it} | \hat{\mu}_{jt}] = 0$ under random assignment of students in year t, the coefficient λ measures the relationship between true teacher effects μ_{jt} and estimated teacher effects $\hat{\mu}_{jt}$:

$$(11) \qquad \lambda \equiv \frac{\mathrm{cov}\,(A_{it},\ \hat{\mu}_{jt})}{\mathrm{var}\,(\hat{\mu}_{jt})} = \frac{\mathrm{cov}\,(\mu_{jt},\ \hat{\mu}_{jt})}{\mathrm{var}\,(\hat{\mu}_{jt})}.$$

We define the degree of bias in VA estimates based on this regression coefficient as follows.

DEFINITION 1: *The amount of forecast bias in a VA estimator $\hat{\mu}_{jt}$ is $B(\hat{\mu}_{jt}) = 1 - \lambda$.*

Forecast bias determines the mean impact of changes in the estimated VA of the teaching staff. A policy that increases estimated teacher VA $\hat{\mu}_{jt}$ by 1 SD raises student test scores by $(1 - B)\,\sigma(\hat{\mu}_{jt})$, where $\sigma(\hat{\mu}_{jt})$ is the standard deviation of VA estimates scaled in units of student test scores. If $B = 0$, $\hat{\mu}_{jt}$ provides an unbiased forecast of teacher quality in the sense that an improvement in estimated VA of $\Delta\hat{\mu}_{jt}$ has the same causal impact on test scores as an increase in true teacher VA $\Delta\mu_{jt}$ of the same magnitude.[9]

Two issues should be kept in mind when interpreting estimates of forecast bias. First, VA estimates can be forecast-unbiased even if children with certain characteristics (e.g., those with higher ability) are assigned systematically to certain teachers. Forecast unbiasedness requires only that the observable characteristics $\mathbf{X}_{it}$ used to construct test score residuals A_{it} are sufficiently rich that the remaining unobserved heterogeneity in test scores ε_{it} is balanced across teachers with different VA estimates. Second, even if VA estimates $\hat{\mu}_{jt}$ are forecast unbiased, potentially one can improve forecasts of μ_{jt} by incorporating other information beyond past test scores. For example, data on principal ratings or teacher characteristics potentially could reduce the mean-squared error of forecasts of teacher quality. Hence, the fact that a VA estimate is forecast-unbiased does not necessarily imply that it is the optimal forecast of teacher quality given all the information which may be available to school districts.

---

[9] Ex post, any estimate of VA can be made forecast-unbiased by redefining VA as $\hat{\mu}'_{jt} = (1 - B)\hat{\mu}_{jt}$. However, the causal effect of a policy which raises the estimated VA of teachers is unaffected by such a rescaling, as the effect of a 1 SD improvement in $\hat{\mu}'_{jt}$ is still $(1 - B)\,\sigma(\hat{\mu}_{jt})$. Hence, given the standard deviation of VA estimates, the degree of forecast bias is informative about the potential impacts of improving estimated VA.

An alternative definition of bias, which we term *teacher-level bias*, asks whether the VA estimate for each teacher converges to her true quality as estimation error vanishes, as in Rothstein (2009). We define teacher-level bias and formalize its connection to forecast bias in online Appendix B. Teacher-level unbiasedness is more challenging to estimate and is a stronger requirement than forecast unbiasedness: some teachers could be systematically overrated relative to others even if forecasts based on VA estimates are accurate on average. However, forecast-unbiased VA estimates can be biased at the teacher level only in the knife-edge case in which the covariance between the bias in each teacher's VA estimate and true VA offsets perfectly the variance in true VA (see online Appendix B). Hence, if VA estimates obtained from a prespecified value-added model turn out to be forecast-unbiased, they are unlikely to be biased at the teacher level.[10] We therefore focus on estimating B and testing if VA estimates from existing models are forecast-unbiased in the remainder of the paper.

II. Data

We draw information from two databases: administrative school district records and federal income tax records. This section describes the two data sources and the structure of the linked analysis dataset and then provides descriptive statistics.

A. *School District Data*

We obtain information on students' test scores and teacher assignments from the administrative records of a large urban school district. These data span the school years from 1988–1989 to 2008–2009 and cover roughly 2.5 million children in grades 3–8. For simplicity, we refer below to school years by the year in which the spring term occurs (e.g., the school year 1988–1989 is 1989).

Test Scores.—The data include approximately 18 million test scores. Test scores are available for English language arts and math for students in grades 3–8 in every year from the spring of 1989 to 2009, with the exception of seventh grade English scores in 2002.[11]

The testing regime varies over the 20 years we study. In the early- and mid-1990s, all tests were specific to the district. Starting at the end of the 1990s, the tests in grades 4 and 8 were administered as part of a statewide testing system, and all tests in grades 3–8 became statewide in 2006 as required under the No Child Left Behind law. All tests were administered in late April or May during the early- and mid-1990s. Statewide tests were sometimes given earlier in the school year (e.g., February) during the latter years of our data.

---

[10] This logic relies on having a pre-specified VA model which is not changed ex post after observing estimates of forecast bias. If we were to redefine our VA estimates after estimating forecast bias as $\hat{\mu}'_{jt} = (1 - B)\hat{\mu}_{jt}$, then $\hat{\mu}'_{jt}$ would be forecast-unbiased by construction. However, teacher-level bias in $\hat{\mu}'_{jt}$ is no longer a knife-edge case because we have introduced a free parameter into the VA estimation procedure that is mechanically chosen to offset the excess variance created by teacher-level bias.

[11] We also have data on math and English test scores in grade 2 from 1991 to 1994, which we use only when evaluating sorting on lagged test score gains. Because these observations constitute a very small fraction of our sample, excluding them has little impact on our results.

Because of this variation in testing regimes, we follow prior work by normalizing the official scale scores from each exam to have mean zero and standard deviation one by year and grade. The within-grade variation in achievement in the district we study is comparable to the within-grade variation nationwide, so our results can be compared to estimates from other samples.[12]

Demographics.—The dataset contains information on ethnicity, gender, age, receipt of special education services, and limited English proficiency for the school years from 1989 to 2009. The database used to code special education services and limited English proficiency changed in 1999, creating a break in these series which we account for in our analysis by interacting these two measures with a post-1999 indicator. Information on free and reduced price lunch is available starting in school year 1999. These missing data issues are not a serious problem because our estimates of forecast bias are insensitive to excluding demographic characteristics from the VA model entirely.

Teachers.—The dataset links students in grades 3–8 to classrooms and teachers from 1991 to 2009.[13] This information is derived from a data management system which was phased in over the early 1990s, so not all schools are included in the first few years of our sample. In addition, data on course teachers for middle school and junior high school students—who, unlike students in elementary schools, are assigned different teachers for math and English—are more limited. Course teacher data are unavailable prior to the school year 1994, then grow in coverage to roughly 60 percent by 1998, and stabilize at approximately 85 percent after 2003. To ensure that our estimates are not biased by the missing data, we show that our conclusions remain very similar in a subsample of school-grade-subject cells with no missing data (see Table 5).

Sample Restrictions.—Starting from the raw dataset, we make a series of restrictions which parallel those in prior work to obtain our primary school district sample. First, because our estimates of teacher value-added always condition on prior test scores, we restrict our sample to grades 4–8, where prior test scores are available. Second, we exclude the 6 percent of observations in classrooms where more than 25 percent of students are receiving special education services, as these classrooms may be taught by multiple teachers or have other special teaching arrangements. We also drop the 2 percent of observations where the student is receiving instruction at home, in a hospital, or in a school serving disabled students solely. Third, we drop classrooms with less than 10 students or more than 50 students as well as teachers linked with more than 200 students in a single grade, because such students are likely to be mislinked to classrooms or teachers (0.5 percent of observations).

---

[12]The standard deviation of fourth and eighth grade English and math achievement in this district ranges from roughly 95 percent to 105 percent of the national standard deviation on the National Assessment of Educational Progress (NAEP), based on data from 2003 and 2009, the earliest and most recent years for which NAEP data are available. Mean scores are significantly lower than the national average, as expected given the urban setting of the district.

[13]Five percent of students switch classrooms or schools in the middle of a school year. We assign these students to the classrooms in which they took the test to obtain an analysis dataset with one observation per student-year-subject. However, when defining class and school-level means of student characteristics (such as fraction eligible for free lunch), we account for such switching by weighting students by the fraction of the year they spent in that class or school.

Finally, when a teacher is linked to students in multiple schools during the same year, which occurs for 0.3 percent of observations, we use only the links for the school where the teacher is listed as working according to human resources records and set the teacher as missing in the other schools.

B. *Tax Data*

We obtain information on parent characteristics from US federal income tax returns spanning 1996–2011.[14] The school district records were linked to the tax data using an algorithm based on standard identifiers (date of birth, state of birth, gender, and names) described in online Appendix C, after which individual identifiers were removed to protect confidentiality. In all, 88.6 percent of the students and 89.8 percent of student-subject-year observations in the sample used to estimate value-added were matched to the tax data (i.e., uniquely linked to a social security record or tax form as described in online Appendix C). Students were then linked to parents based on the earliest 1040 form filed between tax years 1996–2011 on which the student was claimed as a dependent. We identify parents for 97.6 percent of the observations in the analysis dataset conditional on being matched to the tax data.[15] We are able to link almost all children to their parents because almost all parents file a tax return at some point between 1996 and 2012 to obtain a tax refund on their withheld taxes and the Earned Income Tax Credit.

In this paper, we use the tax data to obtain information on five time-invariant parent characteristics, defined as follows. We define parental household income as mean Adjusted Gross Income (capped at $117,000, the ninety-fifth percentile in our sample) between 2005 and 2007 for the primary filer who first claimed the child; measuring parent income in other years yields very similar results (not reported). For years in which parents did not file a tax return, they are assigned an income of zero. We measure income in 2010 US$, adjusting for inflation using the Consumer Price Index.

We define marital status, home ownership, and 401(k) saving as indicators for whether the first primary filer who claims the child ever files a joint tax return, makes a mortgage interest payment (based on data from 1040s for filers and 1099s for non-filers), or makes a 401(k) contribution (based on data from W-2s) between 2005 and 2007.

We define mother's age at child's birth using data from Social Security Administration (SSA) records on birth dates for parents and children. For single parents, we define the mother's age at child's birth using the age of the filer who first claimed the child, who is typically the mother but is sometimes the father or another relative.[16] When a child cannot be matched to a parent, we define all parental characteristics as zero, and we always include a dummy for missing parents in regressions that include parent characteristics.

---

[14] Here and in what follows, the year refers to the tax year: i.e., the calendar year in which income is earned. In most cases, tax returns for tax year t are filed during the calendar year $t + 1$.

[15] Consistent with this statistic, Chetty et al. (2014, Appendix Table I) show that virtually all children in the US population are claimed as dependents at some point during their childhood. Note that our definition of parents is based on who claims the child as a dependent, and thus may not reflect the biological parent of the child.

[16] We set the mother's age at child's birth to missing for 78,007 observations in which the implied mother's age at birth based on the claiming parent's date of birth is below 13 or above 65, or where the date of birth is missing entirely from SSA records.

C. *Summary Statistics*

The linked school district and tax record analysis dataset has one row per student per subject (math or English) per school year, as illustrated in online Appendix Table 1. Each observation in the analysis dataset contains the student's test score in the relevant subject test, demographic information, teacher assignment, and time-invariant parent characteristics. We organize the data in this format so that each row contains information on a treatment by a single teacher conditional on prede-termined characteristics. We account for the fact that each student appears multiple times in the dataset by clustering standard errors as described in Section III.

After imposing the sample restrictions described above, the linked dataset contains 10.7 million student-year-subject observations. We use this *core sample* of 10.7 mil-lion observations to construct quasi-experimental estimates of forecast bias, which do not require any additional controls; 9.1 million records in the core sample have information on teacher assignment and 7.6 million have information on teachers, lagged test scores, and the other controls needed to estimate our baseline VA model.

Table 1 reports summary statistics for the 7.6 million observation sample used to estimate value-added (which includes those with missing parent characteristics). Note that the summary statistics are student-school year-subject means and thus weight students who are in the district for a longer period of time more heavily, as does our empirical analysis. There are 1.37 million students in this sample and each student has 5.6 subject-school year observations on average.

The mean test score in the sample used to estimate VA is positive and has a stan-dard deviation below 1 because we normalize the test scores in the full population that includes students in special education classrooms and schools (who have lower test scores typically). Eighty percent of students are eligible for free or reduced price lunches. For students whom we match to parents, mean parent household income is $40,800, while the median is $31,700. Though our sample includes more low-income households than would a nationally representative sample, it still includes a substantial number of higher-income households, allowing us to analyze the impacts of teachers across a broad range of the income distribution. The standard deviation of parent income is $34,300, with 10 percent of parents earning more than $100,000.

III. Value-Added Estimates

We estimate teacher VA using the methodology in Section IB in three steps: (i) construct student test score residuals, (ii) estimate the autocovariance of scores across classes taught by a given teacher, and (iii) predict VA for each teacher in each year using test score data from other years.

Test Score Residuals.—Within each subject (math and English) and school-level (elementary and middle), we construct test score residuals A_{it} by regressing raw standardized test scores A_{it}^* on a vector of covariates $\mathbf{X}_{it}$ and teacher fixed effects, as in (5). Our baseline control vector $\mathbf{X}_{it}$ is similar to Kane, Rockoff, and Staiger (2008). We control for prior test scores using a cubic polynomial in prior-year scores in math and a cubic in prior-year scores in English, and we interact these cubics with the student's grade level to permit flexibility in the persistence of test

Table 1—Summary Statistics for Sample Used to Estimate Value-Added Model

Variable	Mean (1)	SD (2)	Observations (3)
Panel A. Student data			
Class size (not student-weighted)	27.3	5.6	391,487
Number of subject-school years per student	5.6	3.0	1,367,051
Test score (SD)	0.2	0.9	7,639,288
Female	50.8%		7,639,288
Age (years)	11.4	1.5	7,639,288
Free lunch eligible (1999–2009)	79.6%		5,021,163
Minority (Black or Hispanic)	71.6%		7,639,288
English language learner	4.8%		7,639,288
Special education	1.9%		7,639,288
Repeating grade	1.7%		7,639,288
Matched to parents in tax data	87.7%		7,639,288
Panel B. Parent characteristics			
Household income	40,773	34,270	6,695,982
Owned a house	32.9%		6,695,982
Contributed to a 401(k)	30.9%		6,695,982
Married	42.7%		6,695,982
Age at child birth	29.2	8.0	6,617,975
Predicted score	0.17	0.26	7,639,288

Notes: All statistics reported are for the sample used in estimating the baseline value-added model, as described in Section IIIC. This sample includes only students who have non-missing lagged test scores and other requisite controls to estimate the VA model. Number of observations is number of classrooms in the first row, number of students in the second row, and number of student-subject-year observations in all other rows. Student data are from the administrative records of a large urban school district in the United States. Parent characteristics are measured between the years 2005–2007 from federal income tax data. All monetary values are expressed in real 2010 US$. All ages refer to the age of an individual as of December 31 within a given year. Test score is based on standardized scale scores, as described in Section IIA. Free lunch is an indicator for receiving free or reduced-price lunches. We link students to their parents by finding the earliest 1040 form from 1996 to 2011 on which the student is claimed as a dependent. Conditional on being matched to the tax data, we are unable to link 2.4 percent of students to their parents; the summary statistics for parents exclude these observations. Parent income is average adjusted gross income during the tax-years 2005–2007. For parents who do not file, household income is defined as zero. Home ownership is defined as reporting mortgage interest payments on a 1040 or 1099 form in any year between 2005 and 2007. Contributed to a 401(k) is an indicator for ever contributing to a 401(k) between 2005 and 2007. Marital status is measured by whether the claiming parent files a joint return at any point between 2005 and 2007. Parent age at child birth is the difference between the age of the mother (or father if single father) and the student. Predicted score is predicted from a regression of scores on parent characteristics using the estimating equation in Section IV.

scores as students age.[17] We also control for students' ethnicity, gender, age, lagged suspensions and absences, and indicators for special education, limited English proficiency, and grade repetition. We also include the following class- and school-level controls: (i) cubics in class and school-grade means of prior-year test scores in math and English (defined based on those with non-missing prior scores) each interacted with grade; (ii) class and school-year means of all the other individual covariates; (iii) class size and class-type indicators (honors, remedial); and (iv) grade and year

---

[17] We exclude observations with missing data on current or prior scores in the subject for which we are estimating VA. We also exclude classrooms which have fewer than seven observations with current and lagged scores in the relevant subject (2 percent of observations) to avoid estimating VA based on very few observations. When prior test scores in the other subject are missing, we set the other subject prior score to zero and include an indicator for missing data in the other subject interacted with the controls for prior own-subject test scores.

dummies. Importantly, our control vector $\mathbf{X}_{it}$ consists entirely of variables from the school district dataset.[18]

Auto-Covariance Vector.—Next, we estimate the auto-covariance of mean test score residuals across classes taught in different years by a given teacher, again separately for each subject and school-level. Exploiting the stationarity assumption in (3), we use all available classrooms with a time span of s years between them to estimate $\sigma_{As} = \text{cov}(\overline{A}_{jt}, \overline{A}_{jt-s})$, weighting by the total number of students in each pair of classes. In middle school, where teachers teach multiple classes per year, we estimate σ_{As} after collapsing the data to the teacher-year level by calculating precision-weighted means across the classrooms as described in online Appendix A.

Figure 1 plots the autocorrelations $r_s = \frac{\sigma_{As}}{\sigma_A^2}$ for each subject and school level; the values underlying this figure along with the associated covariances are reported in panel A of Table 2. If one were using data only from year s, the VA estimate for teacher j in year t would simply be $\hat{\mu}_{jt} = r_s A_{j,t-s}$, since $r_s = \psi = \frac{\sigma_{As}}{\sigma_A^2}$, as shown in (8). Hence, r_s represents the *reliability* of mean class test scores for predicting teacher quality s years later. The reliability of VA estimates decays over time: more recent test scores are better predictors of current teacher performance. For example, in elementary school, reliability declines from $r_1 = 0.43$ in the first year to $r_7 = 0.25$ in math after 7 years, and remains roughly stable for $s > 7$. The decay in r_s over a period of 5–7 years followed by stability beyond that point implies that teacher quality includes both a permanent component and a transitory component. The transitory component may arise from variation in teaching assignments, curriculum, or skill (Goldhaber and Hansen 2010; Jackson 2010).[19]

In middle school, we can estimate the within-year covariance of test score residuals σ_{A0} because teachers teach multiple classes per year. Under the assumption that class and student level shocks are i.i.d., this within-year covariance corresponds to the variance of teacher effects. We estimate the standard deviation of teacher effects, $\sigma_\mu = \sqrt{\sigma_{A0}}$, to be 0.098 in English and 0.134 in math in middle school (panel B of Table 2).[20] In elementary school, σ_{A0} is unidentified because teachers teach only one class per year. Conceptually, we cannot distinguish the variance of idiosyncratic class shocks from unforecasted innovations in teacher effects when teachers teach only one class per year. However, we can obtain a lower bound on σ_μ of $\sqrt{\sigma_{A1}}$ because $\sigma_{A0} \geq \sigma_{A1}$. This bound implies that σ_μ is at least 0.113 in English and at least 0.149 in math in elementary school (second-to-last row of panel B of Table 2). To obtain a point estimate of σ_μ, we fit a quadratic function to the log

---

[18] We do not control for teacher experience in our baseline specification because doing so complicates our quasi-experimental teacher switching analysis, as we would have to track changes in both teacher experience and VA net of experience. However, we show in Table 6 below that the correlation between our baseline VA estimates and VA estimates which control for experience is 0.99.

[19] Prior studies, which do not account for drift, estimate reliability typically as the correlation between mean scores for a random pair of classes in different years (e.g., Kane and Staiger 2008; Chetty, Friedman, and Rockoff 2011b). That method yields an estimate of the mean value of r_s over the years available in the dataset. A recent summary (McCaffrey et al. 2009) finds reliability in the range of 0.2–0.5 for elementary school and 0.3–0.7 for middle school teachers, consistent with the estimates in Figure 1.

[20] As is standard in the literature, in this paper we scale value-added in units of *student* test scores: i.e., a one-unit increase in value-added refers to a teacher whose VA is predicted to raise student test scores by 1 SD.

Panel A. Autocorrelation vector in elementary school for English and math scores
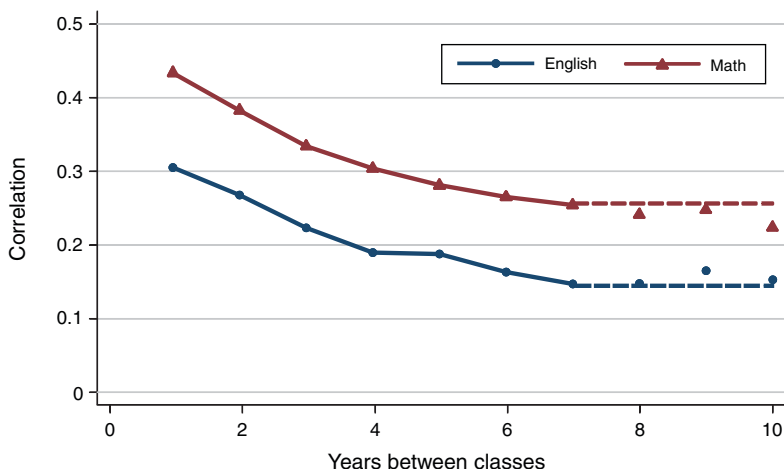
Panel B. Autocorrelation vector in middle school for English and math scores
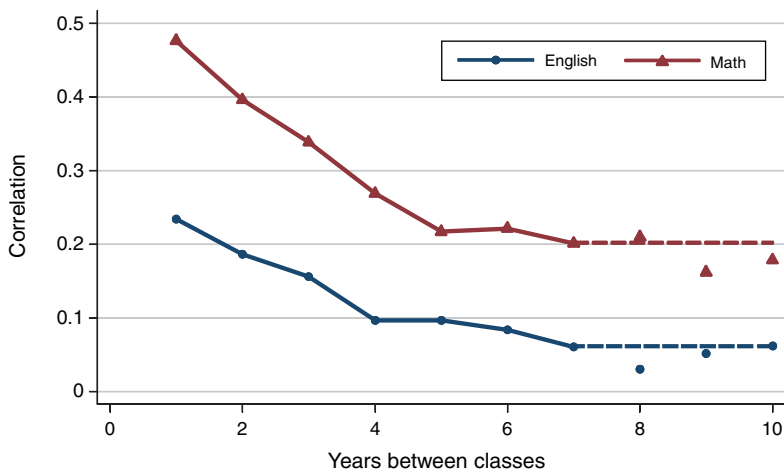
FIGURE 1. DRIFT IN TEACHER VALUE-ADDED ACROSS YEARS

Notes: These figures show the correlation between mean test-score residuals across classes taught by the same teacher in different years. Panels A and B plot autocorrelation vectors for elementary and middle school. To calculate these vectors, we first residualize test scores using within-teacher variation with respect to our baseline control vector (see notes to Table 3). We then calculate a (precision-weighted) mean test score residual across classrooms for each teacher-year. Finally, we calculate the autocorrelation coefficients as the correlation across years for a given teacher, weighting by the sum of students taught in the two years. See online Appendix A for more details on the estimation procedure for these and other parameters of the value-added model.

of the first seven covariances within each subject in elementary school (listed in panel A of Table 2) and extrapolate to 0 to estimate σ_{A0}. This method yields estimates of σ_μ of 0.124 in English and 0.163 in math, as shown in the final row of panel B of Table 2.[21] The point estimates are close to the lower bounds because the

---

[21] Applying the same quadratic extrapolation to the middle school data yields estimates of σ_μ of 0.134 in math and 0.079 in English. These estimates are fairly close to the true values estimated from the within-year covariance, supporting the extrapolation in elementary school.

TABLE 2—TEACHER VALUE-ADDED MODEL PARAMETER ESTIMATES

	Elementary school		Middle school	
	English	Math	English	Math
Sample	(1)	(2)	(3)	(4)
Panel A. Autocovariance and autocorrelation vectors				
Lag 1	0.013	0.022	0.005	0.013
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
	[0.305]	[0.434]	[0.234]	[0.476]
Lag 2	0.011	0.019	0.004	0.011
	(0.0003)	(0.0003)	(0.0002)	(0.0003)
	[0.267]	[0.382]	[0.186]	[0.396]
Lag 3	0.009	0.017	0.003	0.009
	(0.0003)	(0.0004)	(0.0003)	(0.0003)
	[0.223]	[0.334]	[0.156]	[0.339]
Lag 4	0.008	0.015	0.002	0.007
	(0.0004)	(0.0004)	(0.0003)	(0.0004)
	[0.190]	[0.303]	[0.097]	[0.269]
Lag 5	0.008	0.014	0.002	0.006
	(0.0004)	(0.0005)	(0.0004)	(0.0005)
	[0.187]	[0.281]	[0.096]	[0.217]
Lag 6	0.007	0.013	0.002	0.006
	(0.0004)	(0.0006)	(0.0004)	(0.0006)
	[0.163]	[0.265]	[0.084]	[0.221]
Lag 7	0.006	0.013	0.001	0.005
	(0.0005)	(0.0006)	(0.0005)	(0.0006)
	[0.147]	[0.254]	[0.060]	[0.201]
Lag 8	0.006	0.012	0.001	0.005
	(0.0006)	(0.0007)	(0.0005)	(0.0007)
	[0.147]	[0.241]	[0.030]	[0.210]
Lag 9	0.007	0.013	0.001	0.004
	(0.0007)	(0.0008)	(0.0006)	(0.0008)
	[0.165]	[0.248]	[0.051]	[0.162]
Lag 10	0.007	0.012	0.001	0.005
	(0.0008)	(0.0010)	(0.0007)	(0.0012)
	[0.153]	[0.224]	[0.062]	[0.179]

(*Continued*)

rate of drift across one year is small. Our estimates of the standard deviations of teacher effects are slightly larger than those in prior studies (e.g., Kane, Rockoff, and Staiger 2008; Chetty, Friedman, and Rockoff 2011b) because the earlier estimates were attenuated by drift.

Prediction of VA.—We use the estimated autocovariance vectors in Table 2 to predict teacher VA. Since there is no trend in reliability after seven periods and because the precision of the estimates falls beyond that point, we fix $\sigma_{As} = \sigma_{A7}$ for $s > 7$ in all subject and school levels when estimating VA. We predict each teacher's VA in each year t using test score residuals from all years (both past and the future) *except* year t. For example, when predicting teachers' effects on student outcomes in year $t = 1995$ ($\hat{\mu}_{j, 1995}$), we estimate VA based on all years excluding 1995. We construct these estimates using the formula in (7) with an adjustment for the variation in the

TABLE 2—TEACHER VALUE-ADDED MODEL PARAMETER ESTIMATES (*Continued*)

Sample	Elementary school		Middle school	
	English (1)	Math (2)	English (3)	Math (4)
Panel B. Within-year variance components				
Total SD	0.537	0.517	0.534	0.499
Individual-level SD	0.506	0.473	0.513	0.466
Class + teacher level SD	0.117	0.166	0.146	0.178
Class-level SD			0.108	0.116
Teacher SD			0.098	0.134
Estimates of teacher SD:				
Lower bound based on lag 1	0.113	0.149	0.068	0.115
Quadratic estimate	0.124	0.163	0.079	0.134

Notes: Panel A reports the estimated autocovariance, the standard error of that covariance estimate clustered at the teacher level (in parentheses), and the autocorrelation (in brackets) of average test score residuals between classrooms taught by the same teacher. We measure these statistics at time lags ranging from one (i.e., two classrooms taught one year apart) to ten years (i.e., two classrooms taught ten years apart), weighting by the sum of the relevant pair of class sizes. Each covariance is estimated separately for English and math and for elementary and middle school classrooms. Panel B reports the raw standard deviation of test score residuals and decomposes this variation into components driven by idiosyncratic student-level variation, classroom shocks, and teacher-level variation. The variances in rows 2 and 3 of panel B sum to that in row 1; the variances in rows 4 and 5 sum to that in row 3. In middle school, we estimate the standard deviation of teacher effects as the square root of the covariance of mean score residuals across a random pair of classrooms within the same year. In elementary schools, we cannot separately identify class-level and teacher-level standard deviations because we observe only one classroom per year. We use the square root of the autocovariance across classrooms at a one year lag to estimate a lower bound on the within-year standard deviation for elementary schools. We also report an estimate of the standard deviation by regressing the log of first seven autocovariances in panel A on the time lag and time lag squared and extrapolating to zero to estimate the within-year covariance.

number of students per class and the number of classes per year (for middle school) as described in online Appendix A.[22]

The empirical distributions of our teacher VA estimates are plotted in online Appendix Figure 1. The standard deviation of $\hat{\mu}_{jt}$ is 0.116 in math and 0.080 in English in elementary school; the corresponding SDs are 0.092 and 0.042 in middle school. The standard deviations are smaller than the true SD of teacher effects because the best linear predictor shrinks VA estimates toward the sample mean to minimize mean-squared-error.

Out-of-Sample Forecasts.—Under the stationarity assumption in (3), an OLS regression of A_{it} on $\hat{\mu}_{jt}$—the best linear predictor of A_{it}—should yield a coefficient of 1 by construction. We confirm that this is the case in column 1 of Table 3, which reports estimates from a univariate OLS regression of test score residuals A_{it} on $\hat{\mu}_{jt}$ in the sample used to estimate the VA model. We include fixed effects for subject (math versus English) by school-level (elementary versus middle) in this and all

---

[22] Among the classrooms with the requisite controls to estimate value-added (e.g., lagged test scores), we are unable to predict teacher VA for 9 percent of student-subject-year observations because their teachers are observed in the data for only one year.

TABLE 3—ESTIMATES OF FORECAST BIAS USING PARENT CHARACTERISTICS AND LAGGED SCORES

	Score in year t (1)	Pred. score using parent chars. (2)	Score in year t (3)	Pred. score using year $t - 2$ score (4)
Teacher VA	0.998 (0.0057)	0.002 (0.0003)	0.996 (0.0057)	0.022 (0.0019)
Parent chars. controls			X	
Observations	6,942,979	6,942,979	6,942,979	5,096,518

Notes: Each column reports coefficients from an OLS regression, with standard errors clustered by school-cohort in parentheses. The regressions are run on the sample used to estimate the baseline VA model, restricted to observations with a non-missing leave-out teacher VA estimate. There is one observation for each student-subject-school year in all regressions. Teacher VA is scaled in units of student test score standard deviations and is estimated using data from classes taught by the same teacher in other years, following the procedure in Sections IB and III. Teacher VA is estimated using the baseline control vector, which includes: a cubic in lagged own- and cross-subject scores, interacted with the student's grade level; student-level characteristics including ethnicity, gender, age, lagged suspensions and absences, and indicators for grade repetition, special education, and limited English; class size and class-type indicators; cubics in class and school-grade means of lagged own- and cross-subject scores, interacted with grade level; class and school-year means of all the student-level characteristics; and grade and year dummies. When prior test scores in the other subject are missing, we set the other subject prior score to zero and include an indicator for missing data in the other subject interacted with the controls for prior own-subject test scores. In columns 1 and 3, the dependent variable is the student's test score in a given year and subject. In column 2, the dependent variable is the predicted value generated from a regression of test score on mother's age at child's birth, indicators for parent's 401(k) contributions and home ownership, and an indicator for the parent's marital status interacted with a quartic in parent household income, after residualizing all variables with respect to the baseline control vector. In column 4, the dependent variable is the predicted value generated in the same way from twice-lagged test scores. See Section IVB for details of the estimating equation for predicted scores.

subsequent regressions to obtain a single regression coefficient which is identified purely from variation within the subject-by-school-level cells. We cluster standard errors at the school-by-cohort level (where cohort is defined as the year in which a child entered kindergarten) to adjust for correlated errors across students within classrooms and the multiple observations for each student in different grades and subjects.[23] The point estimate of the coefficient on $\hat{\mu}_{jt}$ is 0.998 and the 95 percent confidence interval is (0.986, 1.010).

Panel A of Figure 2 plots the relationship between A_{it} and $\hat{\mu}_{jt}$ nonparametrically, dividing the VA estimates $\hat{\mu}_{jt}$ into 20 equal-size groups (vingtiles) and plotting the mean value of A_{it} in each bin.[24] This binned scatter plot provides a nonparametric representation of the conditional expectation function but does not show the underlying variance in the individual-level data. The regression coefficient and standard error reported in this and all subsequent figures are estimated on the microdata (not the binned averages), with standard errors clustered by school-cohort. The conditional expectation function in panel A of Figure 2 is almost perfectly linear.

---

[23] In our original working paper, we evaluated the robustness of our results to alternative forms of clustering (Chetty, Friedman, and Rockoff 2011b, Appendix Table 7). We found that school-cohort clustering yields more conservative confidence intervals than more computationally intensive techniques such as two-way clustering by student and classroom (Cameron, Gelbach, and Miller 2011).

[24] In this and all subsequent scatter plots, we first demean the x and y variables within subject by school level groups to isolate variation within these cells, as in the regressions.

Panel A. Actual score

Panel B. Predicted score using parent characteristics

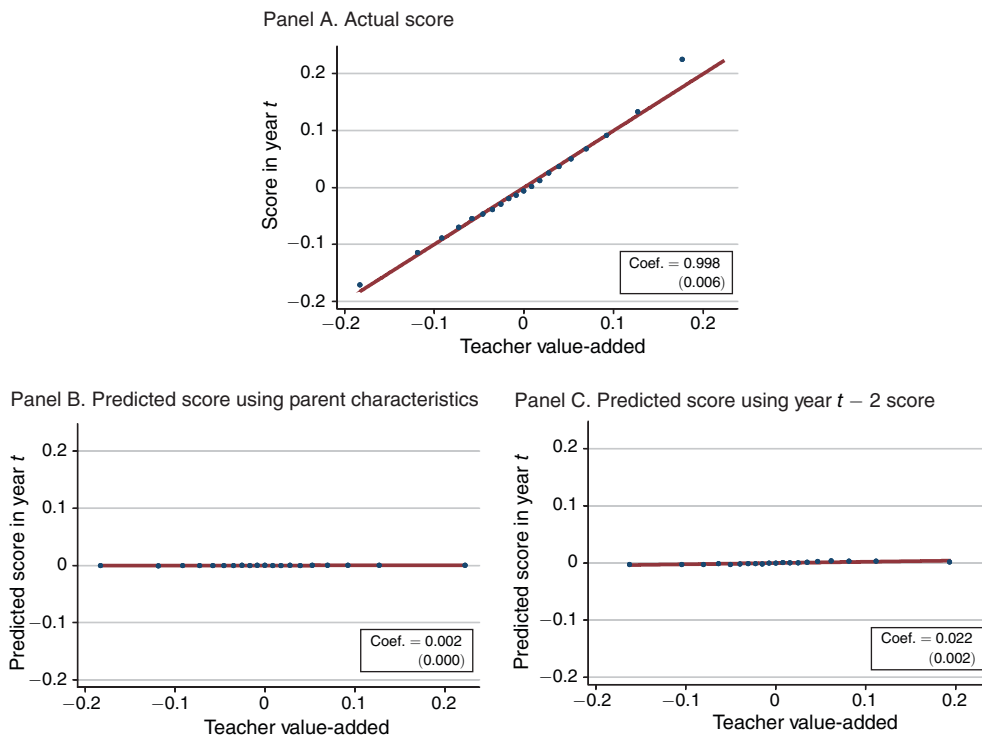Panel C. Predicted score using year $t - 2$ score

FIGURE 2. EFFECTS OF TEACHER VALUE-ADDED ON ACTUAL AND PREDICTED SCORES

Notes: These figures pool all grades and subjects and are constructed using the sample used to estimate the VA model, which has one observation per student-subject-school year. The three panels are binned scatter plots of actual scores, predicted scores based on parent characteristics, and predicted scores based on twice-lagged test scores versus teacher VA. These plots correspond to the regressions in columns 1, 2, and 4 of Table 3 and use the same sample restrictions and variable definitions. To construct these binned scatter plots, we first residualize the *y*-axis variable with respect to the baseline control vector (defined in the notes to Table 3) separately within each subject by school-level cell, using within-teacher variation to estimate the coefficients on the controls as described in Section IB. We then divide the VA estimates $\hat{\mu}_{jt}$ into 20 equal-sized groups (vingtiles) and plot the means of the *y*-variable residuals within each bin against the mean value of $\hat{\mu}_{jt}$ within each bin. The solid line shows the best linear fit estimated on the underlying micro data using OLS. The coefficients show the estimated slope of the best-fit line, with standard errors clustered at the school-cohort level reported in parentheses.

Teacher VA has a 1–1 relationship with test score residuals throughout the distribution, showing that the linear prediction model fits the data well.

The relationship between $\hat{\mu}_{jt}$ and students' test scores in panel A of Figure 2 could be driven by the causal impact of teachers on achievement (μ_{jt}) or persistent differences in student characteristics across teachers (ε_{it}). For instance, $\hat{\mu}_{jt}$ may forecast students' test scores in other years simply because some teachers are always assigned students with higher- or lower-income parents. In the next two sections, we estimate the degree to which the relationship in panel A of Figure 2 reflects teachers' causal effects versus bias due to student sorting.

IV. Estimating Bias Using Observable Characteristics

We defined forecast bias in Section IC under the assumption that students were randomly assigned to teachers in the forecast year *t*. In this section, we develop a

method of estimating forecast bias using observational data: i.e., when the school follows the same (non-random) assignment rule for teachers and students in the forecast period t as in previous periods. We first derive an estimator for forecast bias based on selection on observable characteristics and then implement it using data on parent characteristics and lagged test score gains.

A. *Methodology*

In this subsection, we show that forecast bias can be estimated by regressing predicted test scores based on observable characteristics excluded from the VA model (such as parent income) on VA estimates. To begin, observe that regressing test score residuals A_{it} on $\hat{\mu}_{jt}$ in observational data yields a coefficient of 1 because $\hat{\mu}_{jt}$ is the best linear predictor of A_{it}:

$$\frac{\text{cov}(A_{it}, \hat{\mu}_{jt})}{\text{var}(\hat{\mu}_{jt})} = \frac{\text{cov}(\mu_{jt}, \hat{\mu}_{jt}) + \text{cov}(\varepsilon_{it}, \hat{\mu}_{jt})}{\text{var}(\hat{\mu}_{jt})} = 1.$$

It follows from the definition of forecast bias that in observational data,

$$(12) \qquad\qquad B(\hat{\mu}_{jt}) = \frac{\text{cov}(\varepsilon_{it}, \hat{\mu}_{jt})}{\text{var}(\hat{\mu}_{jt})}.$$

Intuitively, the degree of forecast bias can be quantified by the extent to which students are sorted to teachers based on unobserved determinants of achievement ε_{it}.

Although we cannot observe ε_{it}, we can obtain information on components of ε_{it} using variables that predict test score residuals A_{it} but were omitted from the VA model, such as parent income. Let $\mathbf{P}_{it}^*$ denote a vector of such characteristics and $\mathbf{P}_{it}$ denote the residuals obtained after regressing the elements of $\mathbf{P}_{it}^*$ on the baseline controls $\mathbf{X}_{it}$ in a specification with teacher fixed effects, as in (5). Decompose the error in score residuals $\varepsilon_{it} = \boldsymbol{\rho}\mathbf{P}_{it} + \varepsilon_{it}'$ into the component which projects onto $\mathbf{P}_{it}$ and the remaining (unobservable) error ε_{it}'. To estimate forecast bias using $\mathbf{P}$, we make the following assumption.

ASSUMPTION 2 (Selection on Excluded Observables): *Students are sorted to teachers purely on excluded observables* $\mathbf{P}$:

$$E[\varepsilon_{it}'|j] = E[\varepsilon_{it}'].$$

Under this assumption, $B = \frac{\text{cov}(\boldsymbol{\rho}\mathbf{P}_{it}, \hat{\mu}_{jt})}{\text{var}(\hat{\mu}_{jt})}$. As in (5), we estimate the coefficient vector $\boldsymbol{\rho}$ using an OLS regression of A_{it} on $\mathbf{P}_{it}$ with teacher fixed effects:

$$(13) \qquad\qquad A_{it} = \alpha_j + \boldsymbol{\rho}\mathbf{P}_{it}.$$

This leads to the feasible estimator

$$B_p = \frac{\text{cov}\left(A_{it}^p,\ \hat{\mu}_{jt}\right)}{\text{var}\left(\hat{\mu}_{jt}\right)}, \tag{14}$$

where $A_{it}^p = \hat{\rho}\, \mathbf{P}_{it}$ is estimated using (13). Equation (14) shows that forecast bias B_p can be estimated from an OLS regression of predicted scores A_{it}^p on VA estimates under Assumption 2.

B. *Estimates of Forecast Bias*

We estimate forecast bias using two variables which are excluded from standard VA models: parent characteristics and lagged test score gains.

Parent Characteristics.—We define a vector of parent characteristics $\mathbf{P}_{it}^*$ which consists of the following variables: mother's age at child's birth, indicators for parent's 401(k) contributions and home ownership, and an indicator for the parent marital status interacted with a quartic in parent household income.[25] We construct residual parent characteristics $\mathbf{P}_{it}$ by regressing each element of $\mathbf{P}_{it}^*$ on the baseline control vector $\mathbf{X}_{it}$ and teacher fixed effects, as in (5). We then regress A_{it} on $\mathbf{P}_{it}$, again including teacher fixed effects, and calculate predicted values $A_{it}^p = \hat{\rho}\, \mathbf{P}_{it}$. We fit separate models for each subject and school level (elementary and middle) as above when constructing the residuals $\mathbf{P}_{it}$ and predicted test scores A_{it}^p.

In column 2 of Table 3, we regress A_{it}^p on $\hat{\mu}_{jt}$, including subject-by-school-level fixed effects as above. The coefficient in this regression is $B_p = 0.002$: i.e., the degree of forecast bias due to selection on parent characteristics is 0.2 percent. The upper bound on the 95 percent confidence interval for B_p is 0.25 percent. Panel B of Figure 2 presents a nonparametric analog of this linear regression. It plots A_{it}^p versus teacher VA $\hat{\mu}_{jt}$ using a binned scatter on the same scale as in panel A. The relationship between predicted scores and teacher VA is nearly flat throughout the distribution.

Another intuitive way to assess the degree of selection on parent characteristics—which corresponds to familiar methods of assessing omitted variable bias—is to control for $\mathbf{P}_{it}^*$ when estimating the impact of $\hat{\mu}_{jt}$ on test scores, as in Kane and Staiger (2008, Table 6). To implement this approach using within-teacher variation to construct residuals, we first regress raw test scores A_{it}^* on the baseline control vector used in the VA model $\mathbf{X}_{it}$, parent characteristics $\mathbf{P}_{it}^*$, and teacher fixed effects, as in (5). Again, we fit a separate model for each subject and school level (elementary and middle). We then regress the residuals from this regression (adding back the teacher fixed effects) on $\hat{\mu}_{jt}$, including subject-by-school-level fixed effects as in column 1. Column 3 of Table 3 shows that the coefficient on VA falls to 0.996 after controlling for parent characteristics. The difference between the point estimates in columns 1 and 3 is 0.002. This difference coincides exactly with our estimate of B_p

---

[25] We code the parent characteristics as zero for the 12.3 percent of students whom we are unable to match to a parent either because we could not match the student to the tax data (10.1 percent) or because we could not find a parent for a matched student (2.2 percent). We include indicators for missing parent data in both cases. We also code mother's age at child's birth as zero for observations where we match parents but do not have valid data on the parent's age, and include an indicator for such cases.

in column 2 because A_{it}^p is simply the difference between test score residuals with and without controlling for parent characteristics.

The magnitude of forecast bias due to selection on parent characteristics is very small for two reasons. First, a large fraction of the variation in test scores that project onto parent characteristics is captured by lagged test scores and other controls in the school district data. The standard deviation of class-average predicted score residuals based on parent characteristics is 0.014. Intuitively, students from high-income families have higher test scores not just in the current grade but in previous grades as well, and thus previous scores capture a large portion of the variation in family income. However, because lagged test scores are noisy measures of latent ability, parent characteristics still have significant predictive power for test scores even conditional on $\mathbf{X}_{it}$. The *F*-statistic on the parent characteristics in the regression of test score residuals A_{it} on $\mathbf{P}_{it}$ is 84, when run at the classroom level (which is the variation relevant for detecting class-level sorting to teachers). This leads to the second reason that forecast bias is so small: the remaining variation in parent characteristics after conditioning on $\mathbf{X}_{it}$ is essentially unrelated to teacher VA. The correlation between A_{it}^p and $\hat{\mu}_{jt}$ is 0.014. If this correlation were 1 instead, the bias due to sorting on parent characteristics would be $B_p = 0.149$. This shows that there is substantial scope to detect sorting of students to teachers based on parent characteristics even conditional on the school district controls. In practice, such sorting turns out to be quite small in magnitude, leading to minimal forecast bias.

Prior Test Scores.—Another natural set of variables to evaluate bias is prior test scores (Rothstein 2010). Value-added models typically control for $A_{i,t-1}$, but one can evaluate sorting on $A_{i,t-2}$ (or, equivalently, on lagged gains, $A_{i,t-1} - A_{i,t-2}$). The question here is effectively whether controlling for additional lags affects substantially VA estimates once one controls for $A_{i,t-1}$. For this analysis, we restrict attention to the subsample of students with data on both lagged and twice-lagged scores, essentially dropping fourth grade from our sample. We reestimate VA $\hat{\mu}_{jt}$ on this sample to obtain VA estimates on exactly the sample used to evaluate bias.

We assess forecast bias due to sorting on lagged score gains using the same approach as with parent characteristics. Column 4 of Table 3 replicates column 2, using predicted score residuals based on $A_{i,t-2}$, which we denote by A_{it}^l, as the dependent variable. The coefficient on $\hat{\mu}_{jt}$ is 0.022, with a standard error of 0.002. The upper bound on the 95 percent confidence interval for forecast bias due to twice-lagged scores is 0.026. Panel C of Figure 2 plots predicted scores based on twice-lagged scores (A_{it}^l) against teacher VA, following the same methodology used to construct panel B. Consistent with the regression estimates, there is a slight upward-sloping relationship between VA and A_{it}^l that is small relative to the relationship between VA and test score residuals in year t.

Forecast bias due to omitting $A_{i,t-2}$ is small for the same two reasons as with parent characteristics. The baseline control vector $\mathbf{X}_{it}$ captures much of the variation in $A_{i,t-2}$: the variation in class-average predicted score residuals based on $A_{i,t-2}$ is 0.048. The remaining variation in $A_{i,t-2}$ is not strongly related to teacher VA; the correlation between $A_{i,t-2}$ and $\hat{\mu}_{jt}$ is 0.037. If this correlation were 1, forecast bias would be 0.601, again implying that sorting on lagged gains is quite minimal relative to what it could be.

We conclude that selection on two important predictors of test scores excluded from standard VA models—parent characteristics and lagged test score gains—generates negligible forecast bias in our baseline VA estimates.

C. *Unconditional Sorting*

To obtain further insight into the sorting process, we analyze unconditional correlations (without controlling for school-district observables $\mathbf{X}_{it}$) between teacher VA and excluded observables such as parent income in online Appendix D. We find that sorting of students to teachers based on parent characteristics does not generate significant forecast bias in VA estimates for two reasons.

First, although higher socioeconomic status children have higher-VA teachers on average, the magnitude of such sorting is quite small. For example, a $10,000 (0.3 SD) increase in parent income raises teacher VA by 0.00084 (online Appendix Table 2). One explanation for why even unconditional sorting of students to teacher VA based on family background is small is that 85 percent of the variation in teacher VA is within rather than between schools. The fact that parents sort primarily by choosing schools rather than teachers within schools limits the scope for sorting on unobservables.

Second, and more importantly, controlling for a student's lagged test score entirely eliminates the correlation between teacher VA and parent income. This shows that the unconditional relationship between parent income and teacher VA runs through lagged test scores and explains why VA estimates which control for lagged scores are unbiased despite sorting on parent income.

An auxiliary implication of these results is that differences in teacher quality are responsible for only a small share of the gap in achievement by family income. In the cross section, a $10,000 increase in parental income is associated with a 0.065 SD increase in eighth grade test scores (on average across math and English). Based on estimates of the persistence of teachers' impacts and the correlation between teacher VA and parent income, we estimate that only 4 percent of the cross-sectional correlation between test scores and parent income in eighth grade can be attributed to differences in teacher VA for grades K–8 (see online Appendix D). This finding is consistent with evidence on the emergence of achievement gaps at very young ages (Fryer and Levitt 2004; Heckman, Stixrud, and Urzua 2006), which also suggests that existing achievement gaps are largely driven by factors other than teacher quality. However, it is important to note that good teachers in primary school can close a significant portion of the achievement gap created by these other factors. If teacher quality for low-income students were improved by 0.1 SD in all grades from K–8, eighth grade scores would rise by 0.34 SD, enough to offset more than a $50,000 difference in family income.

V. Estimating Bias Using Teacher Switching Quasi-Experiments

The evidence in Section IV does not rule out the possibility that students are sorted to teachers based on unobservable characteristics orthogonal to parent characteristics and lagged score gains. The ideal method to assess bias due to unobservables would be to estimate the relationship between test scores and VA estimates (from preexisting observational data) in an experiment where students are randomly

assigned to teachers. In this section, we develop a quasi-experimental analog to this experiment which exploits naturally occurring teacher turnover to estimate forecast bias. Our approach yields more precise estimates of the degree of bias than experimental studies (Kane and Staiger 2008; Kane et al. 2013) at the cost of stronger identification assumptions, which we describe and evaluate in detail below.

A. *Methodology*

Adjacent cohorts of students within a school are frequently exposed to different teachers. In our core sample, 30.1 percent of teachers switch to a different grade within the same school the following year, 6.1 percent of teachers switch to a different school within the same district, and another 5.8 percent switch out of the district entirely.[26]

To understand how we use such turnover to estimate forecast bias, consider a school with three fourth-grade classrooms. Suppose one of the teachers leaves the school in 1995 and is replaced by a teacher whose VA estimate in math is 0.3 higher. Assume that the distribution of unobserved determinants of scores ε_{it} does not change between 1994 and 1995. If forecast bias $B = 0$, this change in teaching staff should raise average fourth grade math scores in the school by $0.3/3 = 0.1$. More generally, we can estimate B by comparing the change in mean scores across cohorts to the change in mean VA driven by teacher turnover provided that student quality is stable over time.

To formalize this idea, let $\hat{\mu}_{jt}^{-\{t,t-1\}}$ denote the VA estimate for teacher j in year t constructed as in Section III using data from all years except $t-1$ and t. Similarly, let $\hat{\mu}_{j,t-1}^{-\{t,t-1\}}$ denote the VA estimate for teacher j in year $t-1$ based on data from all years except $t-1$ and t.[27] Let Q_{sgt} denote the (student-weighted) mean of $\hat{\mu}_{jt}^{-\{t,t-1\}}$ across teachers in school s in grade g. We define the change in mean teacher value-added from year $t-1$ to year t in grade g in school s as $\Delta Q_{sgt} = Q_{sgt} - Q_{sg,t-1}$. By leaving out both years t and $t-1$ when estimating VA, we ensure that the variation in ΔQ_{sgt} is driven by changes in the teaching staff rather than changes in VA estimates.[28] Leaving out two years eliminates the correlation between changes in mean test scores across cohorts t and $t-1$ and estimation error in ΔQ_{sgt}.[29]

Let A_{sgt} denote the mean value of A_{it} for students in school s in grade g in year t and define the change in mean residual scores as $\Delta A_{sgt} = A_{sgt} - A_{sg,t-1}$. We estimate the degree of forecast bias B by regressing changes in mean test scores across cohorts on changes in mean teacher VA:

$$(15) \qquad \Delta A_{sgt} = a + b \Delta Q_{sgt} + \Delta \chi_{sgt}.$$

---

[26] Although teachers switch grades frequently, only 4.5 percent of students have the same teacher for two consecutive years.

[27] As above, we estimate VA separately for each subject (math and English) and school-level (elementary and middle), and hence only identify forecast bias from teacher switches within subject by school-level cells.

[28] Part of the variation in $\Delta \bar{\mu}_{sgt}$ comes from drift. Even for a given teacher, predicted VA will change because our forecast of VA varies across years. Because the degree of drift is small across a single year, drift accounts for 5.5 percent of the variance in $\Delta \bar{\mu}_{sgt}$. As a result, isolating the variation due purely to teacher switching using an instrumental variables specification yields very similar results (not reported).

[29] Formally, not using a two-year leave out would immediately violate Assumption 3 below, because unobserved determinants of scores (ε_{sgt} and $\varepsilon_{sg,t-1}$) would appear directly in ΔQ_{sgt}.

The coefficient b in (15) identifies the degree of forecast bias as defined in (10) under the following identification assumption.

ASSUMPTION 3 (Teacher Switching as a Quasi-Experiment): *Changes in teacher VA across cohorts within a school grade are orthogonal to changes in other determinants of student scores:*

$$(16) \qquad\qquad \text{cov}\left(\Delta Q_{sgt}, \, \Delta \chi_{sgt}\right) \, = \, 0.$$

Under Assumption 3, the regression coefficient in (15) measures the degree of forecast bias: $b = \lambda = 1 - B\left(\hat{\mu}_{jt}^{-\{t,t-1\}}\right)$. We present a simple derivation of this result in online Appendix E. Intuitively, (15) is a quasi-experimental analog of the regression in (10) used to define forecast bias, differencing, and averaging across cohorts. Importantly, because we analyze student outcomes at the school-grade-cohort level in (15), we do not exploit information on classroom assignment, thus overcoming the nonrandom assignment of students to classrooms within each school-grade-cohort.

Assumption 3 could potentially be violated by endogenous student or teacher sorting to schools over time. Student sorting at an annual frequency is minimal because of the costs of changing schools. During the period we study, most students would have to move to a different neighborhood to switch schools, which families would be unlikely to do simply because a single teacher leaves or enters a given grade. While endogenous teacher sorting is plausible over long horizons, the high-frequency changes we analyze are driven likely by idiosyncratic shocks such as changes in staffing needs, maternity leaves, or the relocation of spouses. Hence, we believe that (16) is a plausible assumption and we present evidence supporting its validity below.

The estimate of λ obtained from (15) is analogous to a local average treatment effect (LATE) because it applies to the teacher switches which occur in our data rather than all potential teacher switches in the population. For example, if teachers of honors math classes never teach regular math classes, we would be unable to determine if VA estimates for honors teachers are biased relative to those of regular teachers. This is unlikely to be a serious limitation in practice because we observe a wide variety of staff changes in the district we study (including switches from honors to regular classes), as is typical in large school districts (Cook and Mansfield 2013). Moreover, as long as the policies under consideration induce changes similar to those which occur already—for example, policies do not switch teachers from one type of course to another if such changes have never occurred in the past—this limitation does not affect the policy relevance of our estimates.

If observable characteristics $\mathbf{X}_{it}$ are also orthogonal to changes in teacher quality across cohorts (i.e., satisfy Assumption 3), we can implement (15) simply by regressing the change in raw test scores ΔA_{sgt}^* on ΔQ_{sgt}. We therefore begin by analyzing changes in raw test scores ΔA_{sgt}^* across cohorts and then confirm that changes in control variables across cohorts are uncorrelated with ΔQ_{sgt}. Because we do not need controls, in this section we use the core sample of 10.7 million observations described in Section IIC rather than the subset of 7.6 million observations which have lagged scores and the other controls needed to estimate the VA model.

Our research design is related to recent work analyzing the impacts of teacher turnover on student achievement, but is the first to use turnover to validate VA

models. Rivkin, Hanushek, and Kain (2005) identify the variance of teacher effects from differences in variances of test score gains across schools with low versus high teacher turnover. We directly identify the causal impacts of teachers from first moments—the relationship between changes in mean scores across cohorts and mean teacher value-added—rather than second moments. Jackson and Bruegmann (2009) analyze whether the VA of teachers who enter or exit affects the test scores of *other* teachers' students in their school-grade cell, but do not compare changes in mean test scores by cohort to the predictions of VA models.[30]

B. *Event Studies*

To illustrate our research design, we begin with event studies of test scores around the entry and exit of high- and low-VA teachers (Figure 3). Let year 0 denote the school year when a teacher enters or exits a school-grade-subject cell and define all other school years relative to that year (e.g., if the teacher enters in 1995, year 1992 is -3 and year 1997 is $+2$). We define an entry event as the arrival of a teacher who did not teach in that school-grade-subject cell for the three preceding years; analogously, we define an exit event as the departure of a teacher who does not return to the same school-grade-subject cell for at least three years. To obtain a balanced sample, we analyze events for which we have data on average test scores at the school-grade-subject level for at least three years before and after the event.[31]

We define a teacher as *high-VA* if her estimated VA in her year of entry or exit is in the top 5 percent of the distribution of all entrants or leavers in her subject. We define *low-VA* teachers analogously as those who have estimated VA in the bottom 5 percent of the distribution of those who enter or exit in their subjects.[32] We estimate VA for each teacher in the year of entry or exit using data *outside* the six-year window used for the event studies.[33] Because VA is measured with error, some teachers who are classified as high-VA are those who had very good students by chance. Thus, if we were to define a high-VA teacher as one whose students scored highly within the event window, spuriously we would find a relationship between test scores and the entry of such a teacher even if she had no causal impact on student performance.

Panel A of Figure 3 plots the impact of the entry of a high-VA teacher on mean test scores. The solid series plots school-grade-subject-year means of test scores in the three years before and after a high-VA teacher enters the school-grade-subject cell. The dashed line in panel A plots test scores in the previous school year (i.e., the previous grade) for the same cohorts of students. To eliminate any secular trends, we remove year effects by regressing mean test scores on year dummies and plotting the

---

[30] The peer effects documented by Jackson and Bruegmann (2009) could affect our estimates of forecast bias using the switcher design. However, peer learning effects are likely to be smaller with teacher exits than entry, provided that knowledge does not deteriorate very rapidly. We find that teacher entry and exit yield broadly similar results, suggesting that spillovers across teachers are not a first-order source of bias for our technique.

[31] In school-grade-subject cells with multiple events (e.g., entries in both 1995 and 1999), we include all such events by stacking the data and using the three years before and after each event.

[32] In cases where multiple teachers enter or exit at the same time, we use the teachers' mean VA to decide whether the event falls in the top or bottom 5 percent of the relevant VA distribution.

[33] More precisely, we estimate VA for each teacher in each year excluding a five-year window (two years prior, the current year, and two years post). Coupled with our definitions of entry and exit—which require that the teacher not be present in the school-grade-subject cell for three years before or after the event—this ensures that we do not use any data from the relevant cell between event years -3 and $+2$ to compute teacher VA.

Panel A. High value-added teacher entry

Panel B. High value-added teacher exit

Panel C. Low value-added teacher entry

Panel D. Low value-added teacher exit

FIGURE 3. IMPACTS OF TEACHER ENTRY AND EXIT ON TEST SCORES

Notes: These figures plot event studies of current scores (solid line) and scores in the previous school year (dashed line) by cohort as teachers enter or leave a school-grade-subject cell in year $t = 0$. Panels A and B analyze the entry and exit of a high-VA teacher (teachers with VA in the top 5 percent of the distribution); panels C and D analyze the entry and exit of a low-VA (bottom 5 percent) teacher. All panels are plotted using the core sample collapsed to school-grade-subject-year means, as described in Section VB. To construct each panel, we first identify the set of teachers who entered or exited a school-grade-subject cell and define event time as the school year relative to the year of entry or exit. We then estimate each teacher's value-added in event year $t = 0$ using data from classes taught excluding event years $t \in [-3, 2]$. In panel A, we identify the subset of teachers with VA estimates in the top 5 percent of the distribution among entering teachers. We then plot mean current and lagged scores in the relevant school-grade-subject cell for the event years before and after the entry of such a teacher. Panels B–D are constructed analogously. We demean test scores by school year to eliminate secular time trends and normalize the residual scores to zero at event year $t = -3$ for teacher entry and at event year $t = 2$ for teacher exit. Each panel reports the change in mean score gains (current minus lag scores) from $t = -1$ to $t = 0$ and the change in mean estimated VA. We report p-values from F-tests of the hypotheses that the change in score gains from $t = -1$ to $t = 0$ equals the change in mean VA and that the change in score gains equals 0. Mean teacher VA is calculated using a student-weighted average, imputing the sample mean VA (0) for teachers who do not have data outside the $t \in [-3, 2]$ window necessary to compute a leave-out VA estimate.

residuals. We normalize both current and previous scores to 0 in the first year of the event study to facilitate interpretation of the scale. We do not condition on any other covariates in this figure: each point simply shows average test scores for different cohorts of students within a school-grade-subject cell adjusted for year effects.

When a high-VA teacher arrives, end-of-year test scores in the subject and grade taught by that teacher rise immediately. But test scores in the prior grade remain stable, as one would expect: the entry of a high-VA teacher in grade five should have no impact on the same cohort's fourth grade test scores. The stability of prior test scores supports the identification assumption in (16) that school quality and student attributes are not changing sharply around the entry of a high-VA teacher. We also find that class size does not change significantly around the entry and exit events we study.

The mean test score in the grade in which the high-VA teacher enters rises by 0.042 SD from year -1 to 0, while the mean lagged test score changes by 0.008. Hence, the entry of a high-VA teacher increases mean test score gains by 0.035. The null hypothesis that this change is 0 is rejected with $p < 0.001$, with standard errors clustered by school-cohort as above. More importantly, the magnitude of the increase in mean test score gains is very similar to the change in mean teacher VA in the school-grade-subject cell, which is 0.042.[34] The hypothesis that the observed impact on mean score gains equals the increase in mean VA is not rejected ($p = 0.34$), consistent with the hypothesis that the VA estimates are forecast-unbiased.

The remaining panels of Figure 3 repeat the event study in panel A for other types of arrivals and departures. Panel B of Figure 3 examines current and lagged test scores around the departure of a high-VA teacher. In this figure, we normalize both current and lagged mean scores to 0 in the final year of the event study to facilitate interpretation. There is a smooth negative trend in both current and lagged scores, suggesting that high-VA teachers leave schools with declining scores. However, scores in the grade taught by the teacher who leaves drop sharply relative to prior scores in the event year, showing that the departure of the high-VA teacher lowers the achievement of subsequent cohorts of students. Panels C and D analyze the arrival and departure of low-VA teachers. Test scores in the grade taught by the teacher fall relative to prior scores when low-VA teachers enter a school-grade cell and rise when low-VA teachers leave.

In every case, the change in test score gains is significantly different from 0 with $p < 0.01$ but is not significantly different from what one would forecast based on the change in mean teacher VA. Together, these event studies provide direct evidence that deselecting low-VA teachers and retaining high-VA teachers improves the academic achievement of students.

C. *Estimates of Forecast Bias*

The event studies focus on the tails of the teacher VA distribution and thus exploit only a small fraction of the variation arising from teacher turnover in the data. We now exploit all the variation in teacher VA across cohorts due to teaching staff changes to obtain a broader and more precise estimate of forecast bias. To do so, we first construct a leave-two-year-out estimate of VA $\hat{\mu}_{jt}^{-\{t,t-1\}}$ for each teacher using data for all years except $t - 1$ and t. We then calculate the change in mean teacher VA for each school-grade-subject-year cell ΔQ_{sgt} as described in Section VA.

Panel A of Figure 4 presents a binned scatterplot of the changes in mean raw test scores across cohorts ΔA_{sgt}^{*} against changes in mean teacher value-added ΔQ_{sgt}, weighting by the number of students in each cell. We include year fixed effects so that the estimate is identified purely from differential changes in teacher VA across school-grade-subject cells over time and restrict the sample to classrooms with non-missing teacher VA estimates. Changes in the quality of the teaching staff predict strongly changes in test scores across cohorts in a school-grade-subject

---

[34] When computing this change in mean VA, we weight teachers by the number of students they teach. For teachers who do not have any VA measures from classrooms outside the leave-out window, we impute VA as 0 (the sample mean). We discuss the robustness of our results to this imputation below.

Panel A. Changes in actual scores
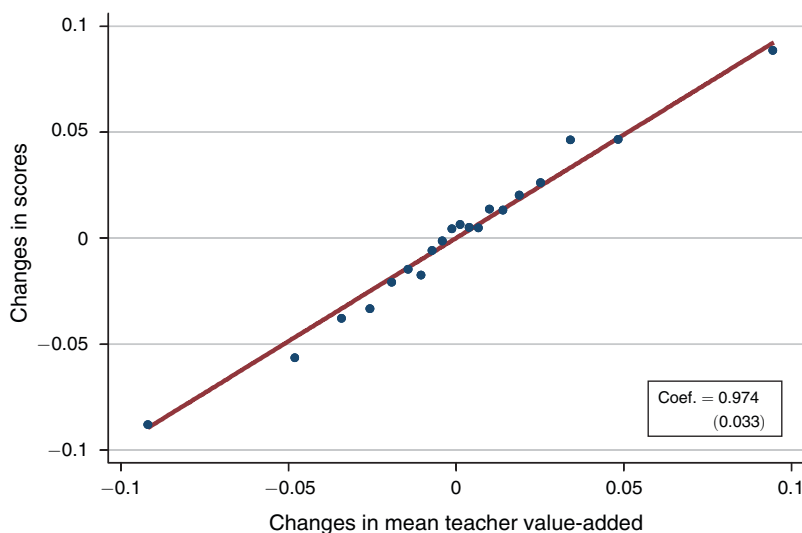
Panel B. Changes in predicted scores based on parent characteristics
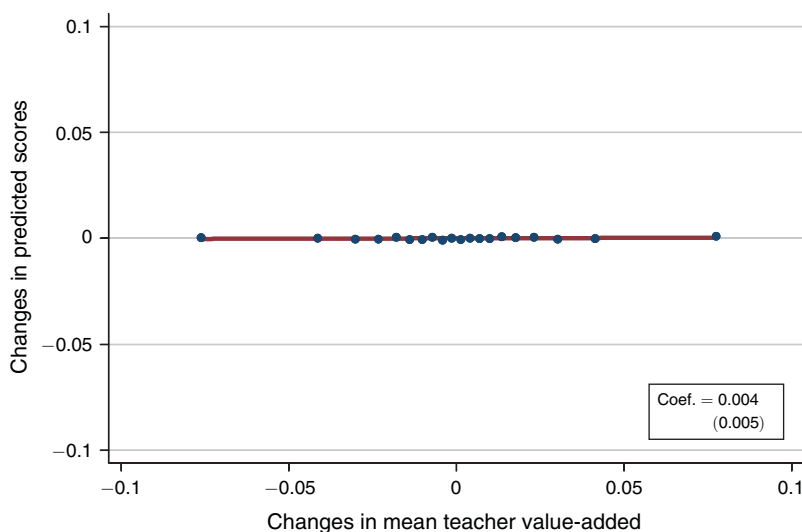

FIGURE 4. EFFECTS OF CHANGES IN TEACHING STAFF ON SCORES ACROSS COHORTS

Notes: This figure plots changes in average test scores across cohorts versus changes in average teacher VA across cohorts, generalizing the event study in Figure 3 to include all changes in teaching staff. Panel A is a binned scatterplot of changes in actual scores versus changes in mean VA, corresponding to the regression in column 1 of Table 4. Panel B is a binned scatterplot of changes in predicted scores based on parent characteristics versus changes in mean VA, corresponding to the regression in column 4 of Table 4. See notes to Table 4 for details on variable definitions and sample restrictions. Both panels are plotted using the core sample collapsed to school-grade-subject-year means, as described in Section VC. To construct these binned scatterplots, we first demean both the *x*- and *y*-axis variables by school year to eliminate any secular time trends. We then divide the observations into 20 equal-size groups (vingtiles) based on their change in mean VA and plot the means of the *y* variable within each bin against the mean change in VA within each bin, weighting by the number of students in each school-grade-subject-year cell. The solid line shows the best linear fit estimated on the underlying microdata using a weighted OLS regression as in Table 4. The coefficients show the estimated slope of the best-fit line, with standard errors clustered at the school-cohort level reported in parentheses.

TABLE 4—QUASI-EXPERIMENTAL ESTIMATES OF FORECAST BIAS

	Δ score (1)	Δ score (2)	Δ score (3)	Δ predicted score (4)	Δ other subject score (5)	Δ other subject score (6)
Changes in mean teacher	0.974	0.957	0.950	0.004	0.038	0.237
VA across cohorts	(0.033)	(0.034)	(0.023)	(0.005)	(0.083)	(0.028)
Year fixed effects	X				X	X
School $\times$ year fixed effects		X	X	X		
Lagged score controls			X			
Lead and lag changes in teacher VA			X			
Other-subject change in mean teacher VA					X	X
Grades	4 to 8	4 to 8	4 to 8	4 to 8	Middle school	Elementary school
Number of school $\times$ grade $\times$ subject $\times$ year cells	59,770	59,770	46,577	59,323	13,087	45,646

Notes: Each column reports coefficients from an OLS regression, with standard errors clustered by school-cohort in parentheses. The regressions are estimated on a dataset containing school-grade-subject-year means from the core sample described in Section IIC, excluding classrooms in which we cannot construct the leave-two-year-out VA estimate described below. All regressions are weighted by the number of students in the school-grade-subject-year cell. We calculate changes in mean teacher VA across consecutive cohorts within a school-grade-subject cell as follows. First, we calculate teacher VA for each teacher in a school-grade-subject cell in each adjacent pair of school years using information excluding those two years. We then calculate mean VA across all teachers, weighting by the number of students they teach. Finally, we compute the difference in mean teacher VA (year t minus year $t - 1$) to obtain the independent variable. The dependent variables are defined by calculating the change in the mean of the dependent variable (year t minus year $t - 1$) within a school-grade-subject cell. In columns 1–3, the dependent variable is the change in mean test scores within subject (English or math). In column 4, it is the change in the predicted score, constructed based on parental characteristics, as described in Section VD. In columns 5 and 6, the dependent variable is the change in the score in the other subject (e.g., math scores for English teachers). Column 5 restricts the sample to middle schools, where different teachers teach math and English; column 6 restricts the sample to elementary schools, where the same teacher teaches the two subjects. Column 1 includes only year fixed effects and no other controls. Columns 2 and 4 include school-by-year fixed effects. In column 3, we add a cubic in the change in mean lagged scores to the specification in column 2, as well as controls for the lead and lag change in mean teacher value-added. Columns 5 and 6 control for the change in mean teacher VA in the other subject as well as year fixed effects.

cell. The estimated coefficient on ΔQ_{sgt} is 0.974, with a standard error of 0.033 (column 1 of Table 4). The implied forecast bias of 2.6 percent is not statistically distinguishable from 0 and the upper bound of the 95 percent confidence interval is 9.1 percent.

The conclusion that VA measures exhibit little or no forecast bias rests on the validity of the identification assumption in (16). One natural concern is that improvements in teacher quality may be correlated with other improvements in a school that also increase test scores and thus lead us to underestimate forecast bias. To address this concern, column 2 of Table 4 replicates the baseline specification in column 1 including school by year fixed effects instead of just year effects. In this specification, the only source of identifying variation comes from differential changes in teacher quality across subjects and grades *within* a school in a given year. The coefficient on ΔQ_{sgt} remains virtually unchanged.

Column 3 further accounts for secular trends in subject- or grade-specific quality by controlling for the change in mean teacher VA in the prior and subsequent year as well as cubics in the change in prior-year mean own-subject and other-subject scores across cohorts. Including these additional controls for trends has little impact on our estimate of forecast bias.

D. *Placebo Tests*

We further evaluate (16) using a series of placebo tests. Column 4 of Table 4 replicates the specification in column 2, replacing the change in actual scores with the change in predicted scores based on parent characteristics. We predict scores using an OLS regression of A_{it}^* on the five parent characteristics $\mathbf{P}_{it}^*$ used in Section IV, with no other control variables. The estimated effect on predicted test scores is not significantly different from 0, and the upper bound on the 95 percent confidence interval is 0.013. Panel B of Figure 4 presents a binned scatterplot corresponding to this regression. There is little relationship between changes in mean teacher VA and mean predicted scores throughout the distribution, supporting the assumption that changes in the quality of the teaching staff are uncorrelated with changes in student characteristics.

Next, we use contemporaneous test scores in the *other* subject (math or English) to evaluate changes in student quality. In this placebo test, it is important to distinguish between elementary and middle schools. In middle school, students have different teachers for math and English. Therefore, if (16) holds, we would expect changes in mean math teacher VA to have little impact on English test scores, holding fixed the quality of English teachers.[35] We test this hypothesis in column 5 of Table 4 by regressing changes in mean scores on ΔQ_{sgt} in both the same subject and the other subject. The estimated effect on test scores in the other subject is 0.038 and is not statistically distinguishable from 0 ($p = 0.64$).

Panel A of Figure 5 presents a nonparametric analog of this regression by plotting changes in mean test scores across cohorts versus changes in mean teacher VA in the other subject. To partial out the effect of changes in mean teacher VA in the same subject, we regress both the *x* and *y* variables on changes in mean teacher VA in the own subject and compute residuals. We then bin the *x*-axis residuals into 20 equal-sized groups (vingtiles) and plot the mean of the *x* and *y* residuals within each bin. Consistent with the regression estimate, there is no relationship between changes in teacher VA and scores in the other subject throughout the distribution. We view this result as strong evidence supporting (16), as it is unlikely that changes in student unobservables ε_{sgt} would have no impact on scores in the other subject.

In elementary school, students have one teacher for both math and English. Because elementary school teachers' math and English VA are highly correlated ($\rho = 0.6$) and because VA is measured with error, changes in teacher VA in one subject have signal content for teaching quality in the other subject. Therefore, in elementary school, we should expect mean teacher VA to have nonzero effects on scores in the other subject. Panel B of Figure 5 replicates panel A for elementary school and shows that this is indeed the case. A 1 unit improvement in mean teacher VA raises scores in the other subject by 0.237 (column 6 of Table 4).

Together, these tests imply that any violation of (16) would have to be driven by unobserved determinants of test scores ε_{sgt} that (i) are uncorrelated with parent characteristics; (ii) are unrelated to prior test scores and contemporaneous test scores in the other subject; and (iii) change differentially across grades within schools at an annual frequency. We believe that such sorting on unobservables is implausible

---

[35] If math teacher quality directly affects English scores (or vice versa), one will obtain a positive coefficient in this regression even if (16) holds. Hence, this test is a sufficient but not necessary condition for (16).

Panel A. Changes in other-subject scores: Middle schools
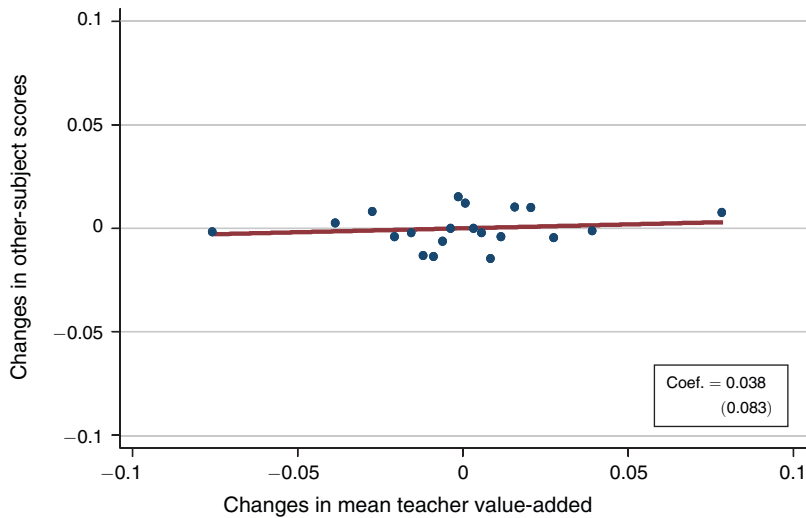
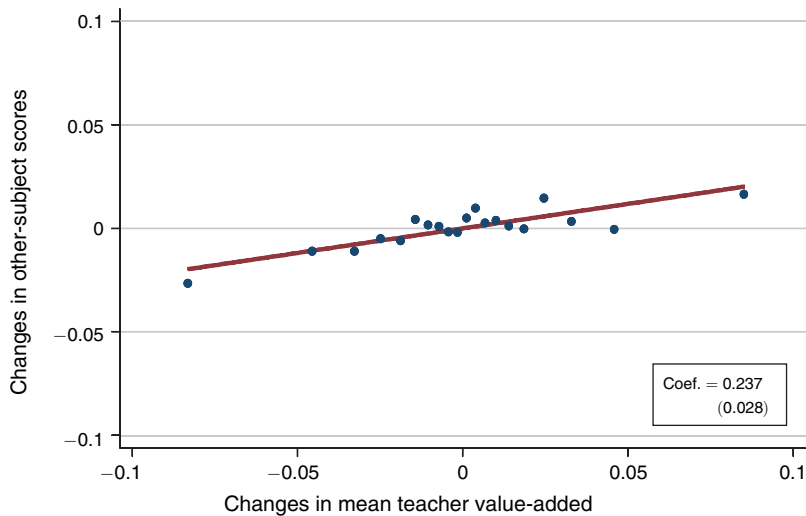Panel B. Changes in other-subject scores: Elementary schools

Figure 5. Effects of Changes in Teaching Staff on Scores in Other Subject

Notes: This figure plots changes in average test scores in the other subject across cohorts versus changes in average teacher VA, controlling for changes in other-subject VA. Panel A restricts the sample to middle schools, corresponding to the regression in column 5 of Table 4. Panel B restricts the sample to elementary schools, corresponding to the regression in column 6 of Table 4. See notes to Table 4 for details on variable definitions and sample restrictions. Both panels are plotted using the core sample collapsed to school-grade-subject-year means, as described in Section VC. To construct these binned scatterplots, we first regress both the *x*- and *y*-axis variables on changes in mean teacher VA in the other subject as well as year fixed effects and compute residuals, weighting by the number of students in each school-grade-subject-year cell. We then divide the *x* residuals into 20 equal-size groups (vingtiles) and plot the means of the *y* residuals within each bin against the mean of the *x* residuals within each bin, again weighting by the number of students in each school-grade-subject-year cell. The solid line shows the best linear fit estimated on the underlying microdata using a weighted OLS regression as in Table 4. The coefficients show the estimated slope of the best-fit line, with standard errors clustered at the school-cohort level reported in parentheses.

TABLE 5—QUASI-EXPERIMENTAL ESTIMATES OF FORECAST BIAS: ROBUSTNESS CHECKS

Specification: Dependent variable:	Teacher exit only Δ score (1)	Full sample Δ score (2)	<25 percent imputed VA Δ score (3)	0 percent imputed VA Δ score (4)
Changes in mean teacher VA across cohorts	1.045 (0.107)	0.877 (0.026)	0.952 (0.032)	0.990 (0.045)
Year fixed effects	X	X	X	X
Number of school × grade × subject × year cells	59,770	62,209	38,958	17,859
Percent of observations with non-imputed VA	100.0	83.6	93.8	100.0

Notes: Each column reports coefficients from a regression with standard errors clustered by school-cohort in parentheses. The regressions are estimated on a dataset containing school-grade-subject-year means from the core sample described in Section IIC. The dependent variable in all specifications is the change in mean test scores (year t minus year $t − 1$) within a school-grade-subject cell. The independent variable is the change in mean teacher VA across consecutive cohorts within a school-grade-subject cell; see notes to Table 4 for details on the construction of this variable. All regressions are weighted by the number of students in the school-grade-subject-year cell and include year fixed effects. In column 1, we report 2SLS (two-stage least-squares) estimates, instrumenting for changes in mean teacher VA with the fraction of students in the prior cohort taught by teachers who leave the school multiplied by the mean-VA among these school-leavers. Columns 2–4 replicate the specification in column 1 of Table 4, varying the way in which we handle classrooms with missing teacher VA. Column 2 includes all classrooms, imputing the sample mean VA (0) to classrooms with missing teacher VA. Column 3 replicates column 2, excluding entire school-grade-subject-year cells in which more than 25 percent of student observations have missing teacher VA. Column 4 restricts to entire school-grade-subject-year cells with no missing teacher VA. The last row reports the percentage of students for whom teacher VA is not imputed in each estimation sample.

given the information available to teachers and students and the constraints they face in switching across schools at high frequencies.

E. *Additional Robustness Checks*

In Table 5, we further assess the robustness of our estimates to alternative specifications and sample selection criteria. Our preceding specifications pool all sources of variation in teacher quality, including switches across grades within a school as well as entry into and exit from schools. In column 1 of Table 5, we isolate the variation in teacher quality due to departures from schools, which are perhaps least likely to be correlated with high-frequency fluctuations in student quality across cohorts. We instrument for the change in mean teacher VA ΔQ_{sgt} with the fraction of students in the prior cohort taught by teachers who leave the school, multiplied by the mean VA among school-leavers (with the instrument defined as 0 for cells with no school-leavers).[36] We include year fixed effects in this specification, as in column 1 of Table 4. This specification yields a coefficient on ΔQ_{sgt} of 1.045, similar to our baseline estimate of 0.974.

The preceding estimates all use the subsample of students for whom we have estimates of teacher VA. While this corresponds directly to the sample used to estimate VA, restricting the sample to classrooms with non-missing VA estimates could lead to violations of the identification assumption in (16) because we do not use the entire

---

[36] For example, if the previous cohort had four equally sized classrooms, with one taught by a school-leaver with $\hat{\mu}_{jt}^{-\{t,t-1\}} = 0.2$, the value of our instrument would be $0.25 \times 0.2 = 0.05$. This is the expected loss in mean VA across cohorts due to school-leavers. The first stage coefficient on this variable is $−0.97$ (t-statistic $= −53$).

school-grade-subject cohort for identification.[37] To evaluate how missing data on teacher VA affects our results, in column 2 of Table 5, we replicate the baseline specification (column 1 of Table 4) using all students in the core sample. We impute teacher VA as the sample mean (0) for the 16 percent of observations for whom we have no leave-two-year-out VA estimate, either because we have no teacher information or because the teacher taught for only those two years in the district. The coefficient on ΔQ_{sgt} with this imputation procedure in the full sample is 0.877. The estimate falls because the imputation of teacher VA generates measurement error in average teaching quality, leading to attenuation bias. For example, if a highly effective teacher enters the district for only a single year, so that we are not able to calculate VA from other years of data, our imputation procedure will treat this teacher as being average, leading to measurement error in mean VA in the school-grade-subject cell.[38]

To assess whether measurement error due to imputation is responsible for the smaller coefficient, we restrict the sample to school-grade-subject-year cells in which the fraction of observations with imputed teacher VA estimates is less than 25 percent in both the current and preceding year. Unlike the sample restrictions imposed in Table 4, restrictions at the school-grade-subject level cannot generate violations of (16) due to selection bias because we exclude entire cohorts rather than individual classrooms. In this subsample, the fraction of observations with non-missing teacher VA information is 94 percent. Accordingly, the coefficient on ΔQ_{sgt} rises to 0.952, as shown in column 3 of Table 5. In column 4, we further restrict the sample to school-grade-subject-year cells with no missing teacher VA observations in the current and preceding year. In this sample, we obtain an estimate on ΔQ_{sgt} of 0.990, implying forecast bias of 1.0 percent. Hence, missing data on teacher VA does not appear to affect our conclusions significantly.

VI. Comparing Bias across VA Models

Our analysis thus far has evaluated bias in a value-added specification which includes a very rich set of controls. Which of these controls are most important in obtaining unbiased estimates of VA? To answer this question, we report estimates of bias for various VA models in Table 6. Each row of the table considers a different VA specification. Column 1 reports correlations between the VA estimates obtained from each model and the baseline estimates. Column 2 reports estimates of forecast bias using our preferred quasi-experimental specification (column 1 of Table 4).

The first row of the table replicates our baseline VA model as a reference. For comparability, we estimate all the remaining models on the sample used in this row.

In row 2, we follow the traditional approach used in prior work of constructing student test score residuals by regressing raw scores A_{it}^* on controls $\mathbf{X}_{it}$ in a

---

[37] To take an extreme example, suppose teacher information is reported for only one of five classrooms in a given school-grade-subject cell. In this case, comparisons of mean scores across students in two cohorts with non-missing teacher information is equivalent to comparing mean scores across a single classroom in two different years. This could violate the identifying assumption of our quasi-experimental design if assignment to classrooms is non-random.

[38] We measure ΔQ_{sgt} with error because we code VA as 0 for teachers with missing VA before we define ΔQ_{sgt}. If we were to instead set ΔQ_{sgt} to 0 in all cells where VA is missing for any teacher, we would mechanically obtain the same coefficient as in column 1 of Table 4.

Table 6—Comparisons of Forecast Bias across Value-Added Models

	Correlation with baseline VA estimates	Quasi-experimental estimate of bias (%)
	(1)	(2)
1. Baseline	1.000	2.58
		(3.34)
2. Baseline, no teacher FE	0.979	2.23
		(3.50)
3. Baseline, with teacher experience	0.989	6.66
		(3.28)
4. Prior test scores	0.962	3.82
		(3.30)
5. Student's lagged scores in both subjects	0.868	4.83
		(3.29)
6. Student's lagged score in same subject only	0.787	10.25
		(3.17)
7. Non-score controls	0.662	45.39
		(2.26)
8. No controls	0.409	65.58
		(3.73)

Notes: In this table, we estimate seven alternative VA models and report correlations of the resulting VA estimates with the baseline VA estimates in column 1. In column 2, we report quasi-experimental estimates of forecast bias for each model, defined as 1 minus the coefficient in a regression of the cross-cohort change in scores on the cross-cohort change in mean teacher VA. These coefficients are estimated using exactly the specification in column 1 of Table 4. All the VA models are estimated on a constant sample of students for whom all the variables in the baseline control vector are non-missing. All models are estimated separately by school level and subject; the correlations and estimates of forecast bias pool VA estimates across all groups. Each model only varies the control vector used to estimate student test score residuals in equation (4); the remaining steps of the procedure used to construct VA estimates (described in online Appendix A) are the same for all the models. Model 1 replicates the baseline model as a reference; see notes to Table 3 for definition of the baseline control vector. The estimated forecast bias for this model coincides with that implied by column 1 of Table 4. Model 2 uses all of the baseline controls but omits teacher fixed effects when estimating equation (5), so that the coefficients on the controls are identified from both within- and between-teacher variation as in traditional VA specifications. Model 3 includes all the controls in the baseline specification as well as indicators for years of teacher experience; teachers with six or more years of experience are pooled into a single group and are the omitted category. Model 4 includes only student-, class-, and school-level test score controls from the baseline control vector along with grade and year fixed effects. Model 5 includes only cubic polynomials in prior-year scores in math and English along with grade and year fixed effects. Model 6 replicates model 5, dropping the cubic polynomial for prior scores in the other subject. Model 7 removes all controls related to test scores from the baseline specification, leaving only non-score controls at the student, class, and school level (e.g., demographics, free lunch participation, etc.). Model 8 drops all controls except grade and year fixed effects.

specification without teacher fixed effects, unlike in (5). As discussed in Section IB, this method exploits variation both within and across teachers to identify the coefficients on the control vector and thus can understate teacher effects by over-attributing test score growth to covariates if there is sorting. In practice, this turns out to be a minor concern: VA estimates constructed from test score residuals which are identified using the traditional approach have a correlation of 0.979 with our baseline VA estimates. Correspondingly, they exhibit forecast bias of 2.2 percent, very similar to our baseline model. This is because unconditional sorting is relatively minimal in practice, as shown in Section IVC, and thus most of the variation in the controls $\mathbf{X}_{it}$ is within rather than between teachers.

In row 3 of Table 6, we replicate our baseline specification adding indicators for teacher experience to the control vector $\mathbf{X}_{it}$. We include indicators for years of experience from zero to five, with the omitted group being teachers with six or more years of experience. The resulting VA estimates have a correlation of 0.989 with our baseline estimates because teacher experience has small effects on student achievement (Kane, Rockoff, and Staiger 2008).

The remaining rows of the table strip out elements of the baseline control vector to determine which controls matter most for forecast bias. In row 4, we include only year fixed effects and the controls based on prior-year test scores: cubic polynomials in student, classroom, and school-grade math and English scores interacted with grade level. These VA estimates remain highly correlated with the baseline estimates and continue to exhibit small and statistically insignificant forecast bias of 3.8 percent.

In row 5 of Table 6, we further reduce the controls to include only the cubics in student-level prior math and English test scores (without grade interactions) along with grade and year fixed effects. Forecast bias is 4.8 percent, somewhat higher than with our much richer baseline control vector. Nevertheless, even with this very parsimonious control vector, we do not reject the hypothesis that VA estimates are forecast unbiased. In row 6, we replicate row 5, dropping the controls for test scores in the other subject. Forecast bias rises to 10.2 percent in this specification, showing that prior test scores in the other subject are useful in accounting for sorting.

In row 7, we control for all variables in the baseline model *except* those based on prior test scores. Here, the degree of forecast bias jumps to 45.4 percent. Finally, row 8 estimates VA without any controls except year and grade fixed effects: i.e., using raw mean test scores by teacher. These VA estimates are very poorly correlated with the other VA measures and are biased by nearly 66 percent.

The lesson of this analysis is that controlling for prior student-level test scores is the key to obtaining unbiased VA estimates. One potential explanation for this result is that classroom assignments in large schools are made primarily on the basis of prior-year test performance and its correlates.

VII. Relationship to Prior Work

Prior research on bias in value-added models has reached conflicting conclusions (e.g., Kane and Staiger 2008; Rothstein 2010; Koedel and Betts 2011; Kinsler 2012; Goldhaber and Chaplin 2012; Kane et al. 2013). Our findings help reconcile these results and show that there is actually consensus in the literature on the central empirical issues.

Rothstein (2009, 2010) initiated the recent body of research on bias in VA models by reporting two results. First, there is significant grouping of students into classrooms based on twice-lagged scores (lagged gains), even conditional on once-lagged scores (Rothstein 2010, Table 4). Second, this grouping on lagged gains generates minimal bias in VA estimates: controlling for twice-lagged scores does not have a significant effect on VA estimates (Rothstein 2010, Table 6).[39] We replicate these

---

[39] An interesting question is how Rothstein's two findings are consistent with each other. There are two explanations for this pattern. First, the degree of grouping that Rothstein finds on $A_{i,t-2}$ has small effects on residual test score gains because the correlation between $A_{i,t-2}$ and A_{it} conditional on $A_{i,t-1}$ is relatively small. Second, if

results in our data (see Table 3), as do Kane and Staiger (2008, Table 6). Therefore, the literature is in agreement that VA measures do not suffer from substantial bias due to selection on lagged score gains.

Rothstein (2010) emphasizes that his findings raise concerns about the *potential* for bias due to selection on unobservable student characteristics.[40] To address this concern, Kane and Staiger (2008) and Kane et al. (2013) report estimates of bias from experiments in which students were randomly assigned to teachers. Their point estimates suggest that selection on unobservables is small, but their 95 percent confidence intervals are consistent with a large amount of bias—up to 50 percent in some cases—because of constraints on sample size and compliance. Our quasi-experimental estimates show that the degree of bias due to selection on unobservables turns out to be negligible with much greater precision in a more representative sample. The recent replication of our quasi-experimental methodology by Kane, Staiger, and Bacher-Hicks (2014) in the Los Angeles Unified School District also reaches the same conclusion. Hence, studies that have directly tested for selection on unobservables are in agreement that VA estimates are not significantly biased by such selection.

In future research, it may be interesting to explore why the grouping of students on lagged test score gains documented originally by Rothstein (2010) does not ultimately lead to significant forecast bias in VA estimates. However, the findings in this paper and the related literature are sufficient to conclude that standard estimates of teacher VA provide unbiased forecasts of teachers' impacts on students' test scores.

VIII. Conclusion

The main lesson of this study is that value-added models which control for a student's prior-year test scores provide unbiased forecasts of teachers' causal impacts on student achievement. Because the dispersion in teacher effects is substantial, this result implies that improvements in teacher quality can raise students' test scores significantly.

Although our analysis shows that test score-based VA measures are useful predictors of teacher quality, one can potentially improve such predictions in at least two dimensions. First, incorporating other measures of teacher quality—such as principal evaluations or information on teacher characteristics—may yield more precise forecasts of teachers' impacts on test scores. One could use the quasi-experimental methodology developed here to validate such measures of teacher quality. For example, when a teacher with good evaluations switches into a school, do outcomes improve? Second, it would be valuable to develop richer measures of teacher quality which go beyond the mean test score impacts that we analyzed here. One could develop VA measures at various percentiles of the test score distribution and for specific demographic subgroups. For instance, are some teachers better with boys rather than girls or high achievers rather than low achievers?

---

the component of $A_{i,t-2}$ on which there is grouping is not the same as the component that is correlated with $A_{i,t}$, VA estimates may be completely unaffected by grouping on $A_{i,t-2}$. For both reasons, one cannot infer from grouping on $A_{i,t-2}$ that VA estimates are significantly biased by selection on $A_{i,t-2}$. See Goldhaber and Chaplin (2012) for further discussion of these and related issues.

[40] In personal correspondence, Rothstein notes that his findings are "neither necessary nor sufficient for there to be bias in a VA estimate" and that "if the selection is just on observables, the bias is too small to matter. The worrying scenario is selection on unobservables."

In this paper, we established that value-added measures can help us identify which teachers have the greatest ability to raise students' test scores. One cannot conclude from this result that teacher VA is a good measure of teacher *quality*, however, as test scores are not the ultimate outcome of interest. Do high-VA teachers also improve students' long-term outcomes, or are they simply better at teaching to the test? We turn to this question in the next paper in this volume.

REFERENCES

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

Baker, Eva L., Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd, Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard. 2010. "Problems with the Use of Student Test Scores to Evaluate Teachers." Economic Policy Institute Briefing Paper 278.

Barlevy, Gadi, and Derek Neal. 2012. "Pay for Percentile." *American Economic Review* 102 (5): 1805–31.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas Miller. 2011. "Robust Inference with Multi-way Clustering." *Journal of Business and Economic Statistics* 29 (2): 238–49.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2011a. "New Evidence on the Long-Term Impacts of Tax Credits." *Statistics of Income White Paper.* The Internal Revenue Service. (accessed June 25, 2013).

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2011b. "The Long Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood." National Bureau of Economic Research Working Paper 17699.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates: Dataset." *American Economic Review.* http://dx.doi.org/10.1257/aer.104.9.2593.

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* 104 (9): 2633–79.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." National Bureau of Economic Research Working Paper 19843.

Cook, Jason, and Richard Mansfield. 2013. "Task-Specific Experience and Task-Specific Talent: Decomposing the Productivity of High School Teachers." Unpublished.

Corcoran, Sean P. 2010. "Can Teachers be Evaluated by Their Students' Test Scores? Should They Be? The Use of Value-Added Measures of Teacher Effectiveness in Policy and Practice." Report for the Annenberg Institute for School Reform at Brown University, Education Policy for Action Series.

Fryer, Roland G., Jr., and Steven D. Levitt. 2004. "Understanding the Black-White Test Score Gap in the First Two Years of School." *The Review of Economics and Statistics* 86 (2): 447–64.

Goldhaber, Dan, and Duncan Chaplin. 2012. "Assessing the 'Rothstein Falsification Test': Does It Really Show Teacher Value-Added Models Are Biased?" Center for Education Data and Research (CEDR) Working Paper 2011–5.

Goldhaber, Dan, and Michael Hansen. 2010. "Using Performance on the Job to Inform Teacher Tenure Decisions." *American Economic Review* 100 (2): 250–55.

Gordon, Robert, Thomas J. Kane, and Douglas O. Staiger. 2006. "Identifying Effective Teachers Using Performance on the Job." The Brookings Institution, Hamilton Project Discussion Paper 2006–01.

Hanushek, Eric. 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." *American Economic Review* 61 (2): 280–88.

Hanushek, Eric A. 2009. "Teacher Deselection." In *Creating a New Teaching Profession*, edited by Dan Goldhaber and Jane Hannaway, 165–80. Washington, DC: Urban Institute Press.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24 (3): 411–82.

Jackson, C. Kirabo. 2009. "Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation." *Journal of Labor Economics* 27 (2): 213–56.

Jackson, C. Kirabo. 2010. "Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence From Teachers." National Bureau of Economic Research Working Paper 15990.

Jackson, C. Kirabo, and Elias Bruegmann. 2009. "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers." *American Economic Journal: Applied Economics* 1 (4): 85–108.

Jacob, Brian A., Lars Lefgren, and David Sims. 2010. "The Persistence of Teacher-Induced Learning Gains." *Journal of Human Resources* 45 (4): 915–43.

Kane, Thomas J., Daniel F. McCaffrey, Trey Miller, and Douglas O. Staiger. 2013. *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment.* Seattle: Bill & Melinda Gates Foundation.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. "What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27 (6): 615–31.

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.

Kane, Thomas J., Douglas O. Staiger, and Andrew Bacher-Hicks. 2014. "Validating Teacher Effect Estimates Using Between School Movers: A Replication and Extension of Chetty et al." Unpublished.

Kinsler, Josh. 2012. "Assessing Rothstein's Critique of Teacher Value-Added Models." *Quantitative Economics* 3 (2): 333–62.

Koedel, Cory, and Julian R. Betts. 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.

Mansfield, Richard K. 2013. "Teacher Quality and Student Inequality." Unpublished.

McCaffrey, Daniel F., Tim R. Sass, J. R. Lockwood, and Kata Mihaly. 2009. "The Intertemporal Variability of Teacher Effect Estimates." *Education Finance and Policy* 4 (4): 572–606.

Murnane, Richard J. 1975. *The Impact of School Resources on Learning of Inner City Children.* Cambridge, MA: Ballinger Publishing.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73 (2): 417–58.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94 (2): 247–52.

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

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125 (1): 175–214.

This article has been cited by:

1. Jason A. Grissom, Brendan Bartanen. 2022. Potential Race and Gender Biases in High-Stakes Teacher Observations. *Journal of Policy Analysis and Management* **41**:1, 131-161. [Crossref]

2. Maximiliaan W.P. Thijssen, Mari Rege, Oddny J. Solheim. 2022. Teacher relationship skills and student learning. *Economics of Education Review* **89**, 102251. [Crossref]

3. James Cowan, Dan Goldhaber, Roddy Theobald. 2022. Performance Evaluations as a Measure of Teacher Effectiveness When Implementation Differs: Accounting for Variation across Classrooms, Schools, and Districts. *Journal of Research on Educational Effectiveness* **15**:3, 510-531. [Crossref]

4. Roddy J. Theobald, Dan D. Goldhaber, Kristian L. Holden, Marcy L. Stein. 2022. Special Education Teacher Preparation, Literacy Instructional Alignment, and Reading Achievement for Students With High-incidence Disabilities. *Exceptional Children* **88**:4, 381-400. [Crossref]

5. My Nguyen, Kien Le. 2022. Racial/ethnic match and student–teacher relationships. *Bulletin of Economic Research* **51**. . [Crossref]

6. Diether W Beuermann, C Kirabo Jackson, Laia Navarro-Sola, Francisco Pardo. 2022. What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output. *The Review of Economic Studies* **82**. . [Crossref]

7. Aisha Salim Al-Harthi, Waheed Hammad, Fawzia Al-Seyabi, Noor Al-Najjar, Sulaiman Al-Balushi, Mahmoud Emam. 2022. Evaluating the effectiveness of teacher education in Oman: a multiple case study of multiple stakeholders. *Quality Assurance in Education* **43**. . [Crossref]

8. Eric Isenberg, Jeffrey Max, Philip Gleason, Jonah Deutsch. 2022. Do Low-Income Students Have Equal Access to Effective Teachers?. *Educational Evaluation and Policy Analysis* **44**:2, 234-256. [Crossref]

9. Asma Benhenda. 2022. Absence, substitutability and productivity: Evidence from teachers. *Labour Economics* **76**, 102167. [Crossref]

10. Rachael Meager. 2022. Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature. *American Economic Review* **112**:6, 1818-1847. [Abstract] [View PDF article] [PDF with links]

11. Adriana Lleras-Muney, Joseph Price, Dahai Yue. 2022. The association between educational attainment and longevity using individual-level data from the 1940 census. *Journal of Health Economics* 102649. [Crossref]

12. Jeffrey Smith. 2022. Treatment Effect Heterogeneity. *Evaluation Review* **5**, 0193841X2210907. [Crossref]

13. Dan Goldhaber, John Krieg, Roddy Theobald, Marcelle Goggins. 2022. Front End to Back End: Teacher Preparation, Workforce Entry, and Attrition. *Journal of Teacher Education* **73**:3, 253-270. [Crossref]

14. Vasiliki Polymeropoulou, Angeliki Lazaridou. 2022. Quality Teaching: Finding the Factors That Foster Student Performance in Junior High School Classrooms. *Education Sciences* **12**:5, 327. [Crossref]

15. Matthew P. Steinberg, Haisheng Yang. 2022. Does Principal Professional Development Improve Schooling Outcomes? Evidence from Pennsylvania's Inspired Leadership Induction Program. *Journal of Research on Educational Effectiveness* **1**, 1-49. [Crossref]

16. Enrico Cantoni, Vincent Pons. 2022. Does Context Outweigh Individual Characteristics in Driving Voting Behavior? Evidence from Relocations within the United States. *American Economic Review* **112**:4, 1226-1272. [Abstract] [View PDF article] [PDF with links]

17. Yan Zhu, Gabriele Kaiser. 2022. Impacts of classroom teaching practices on students' mathematics learning interest, mathematics self-efficacy and mathematics test achievements: a secondary analysis of Shanghai data from the international video study Global Teaching InSights. *ZDM – Mathematics Education* **3**. . [Crossref]

18. Ray Miller. 2022. The Effect of Private Schools on Measures of Socio-Emotional Development in Adolescence: Evidence from India. *Journal of Human Capital* . [Crossref]

19. Anthony Bald, Eric Chyn, Justine Hastings, Margarita Machelett. 2022. The Causal Impact of Removing Children from Abusive and Neglectful Homes. *Journal of Political Economy* . [Crossref]

20. Felipe Arteaga, Adam J Kapor, Christopher A Neilson, Seth D Zimmerman. 2022. Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice. *The Quarterly Journal of Economics* **126**. . [Crossref]

21. Mark Hoekstra, CarlyWill Sloan. 2022. Does Race Matter for Police Use of Force? Evidence from 911 Calls. *American Economic Review* **112**:3, 827-860. [Abstract] [View PDF article] [PDF with links]

22. Xuan Zhang. 2022. The effects of physician retirement on patient outcomes: Anticipation and disruption. *Journal of Public Economics* **207**, 104603. [Crossref]

23. Thurston Domina, Patrick Akos, Kevin C. Bastian, Jeremy Godwin. 2022. The Impact of School Counselor Resources in Elementary and Middle Grades. *Professional School Counseling* **26**:1a, 2156759X2210867. [Crossref]

24. Pedro Carneiro, Yyannu Cruz-Aguayo, Ruthy Intriago, Juan Ponce, Norbert Schady, Sarah Schodt. When Promising Interventions Fail: Personalized Coaching for Teachers in a Middle-Income Country **5**, . [Crossref]

25. Julien Combe, Olivier Tercieux, Camille Terrier. 2022. The Design of Teacher Assignment: Theory and Evidence. *The Review of Economic Studies* **88**. . [Crossref]

26. Xumei Fan. 2022. Teachers' perspectives on the evaluation of teacher effectiveness: A focus on student learning objectives. *Teaching and Teacher Education* **110**, 103604. [Crossref]

27. Roberto V. Penaloza, Mark Berends. 2022. The Mechanics of Treatment-effect Estimate Bias for Nonexperimental Data. *Sociological Methods & Research* **51**:1, 165-202. [Crossref]

28. Hua-Yu Sebastian Cherng, Peter F. Halpin, Luis A. Rodriguez. 2022. Teaching Bias? Relations between Teaching Quality and Classroom Demographic Composition. *American Journal of Education* **128**:2, 171-201. [Crossref]

29. Colin Harris, Andrew Myers, Christienne Briol, Sam Carlen. The Binding Force of Economics 69-103. [Crossref]

30. Allyson L. Snyder, Xin Tong, Angeline S. Lillard. 2022. Standardized Test Proficiency in Public Montessori Schools. *Journal of School Choice* **16**:1, 105-135. [Crossref]

31. Darci L. Gueta, Alexandra Babino. Teaching Across the Demographic Divide With Latinx Texts on Immigration 120-140. [Crossref]

32. Todd D. Reeves, Valerie Hamilton, Yasemin Onder. 2022. Which teacher induction practices work? Linking forms of induction to teacher practices, self-efficacy, and job satisfaction. *Teaching and Teacher Education* **109**, 103546. [Crossref]

33. Martin Braun, Valentin Verdier. 2022. Estimation of spillover effects with matched data or longitudinal network data. *Journal of Econometrics* **67**. . [Crossref]

34. David S. Yeager, Jamie M. Carroll, Jenny Buontempo, Andrei Cimpian, Spencer Woody, Robert Crosnoe, Chandra Muller, Jared Murray, Pratik Mhatre, Nicole Kersting, Christopher Hulleman, Molly Kudym, Mary Murphy, Angela Lee Duckworth, Gregory M. Walton, Carol S. Dweck. 2022. Teacher Mindsets Help Explain Where a Growth-Mindset Intervention Does and Doesn't Work. *Psychological Science* **33**:1, 18-32. [Crossref]

35. Minseok Yang, Ho Jun Lee. 2022. Do school resources reduce socioeconomic achievement gap? Evidence from PISA 2015. *International Journal of Educational Development* **88**, 102528. [Crossref]

36. Pedro Carneiro, Yyannu Cruz-Aguayo, Ruthy Intriago, Juan Ponce, Norbert Schady, Sarah Schodt. 2022. When promising interventions fail: Personalized coaching for teachers in a middle-income country. *Journal of Public Economics Plus* **3**, 100012. [Crossref]

37. Drew Atchison, Michael S. Garet, Toni M. Smith, Mengli Song. 2022. The Validity of Measures of Instructional Alignment With State Standards Based on Surveys of Enacted Curriculum. *AERA Open* **8**, 233285842210987. [Crossref]

38. Louise Ashton, Jed Friedman, Diana Goldemberg, Mustafa Zakir Hussain, Thomas Kenyon, Akib Khan, Mo Zhou. A Puzzle with Missing Pieces: Explaining the Effectiveness of World Bank Development Projects **61**, . [Crossref]

39. Tim T. Morris, Danny Dorling, Neil M. Davies, George Davey Smith. 2021. Associations between school enjoyment at age 6 and later educational achievement: evidence from a UK cohort study. *npj Science of Learning* **6**:1. . [Crossref]

40. Stephen Gibbons, Vincenzo Scrutinio, Shqiponja Telhaj. 2021. Teacher turnover: Effects, mechanisms and organisational responses. *Labour Economics* **73**, 102079. [Crossref]

41. Katharina Wedel. 2021. Instruction time and student achievement: The moderating role of teacher qualifications. *Economics of Education Review* **85**, 102183. [Crossref]

42. Simon Burgess, Robert Metcalfe, Sally Sadoff. 2021. Understanding the response to financial and non-financial incentives in education: Field experimental evidence using high-stakes assessments. *Economics of Education Review* **85**, 102195. [Crossref]

43. Deon Filmer, Vatsal Nahata, Shwetlena Sabarwal. Preparation, Practice, and Beliefs: A Machine Learning Approach to Understanding Teacher Effectiveness **1**, . [Crossref]

44. Raoul van Maarseveen. 2021. The urban–rural education gap: do cities indeed make us smarter?. *Journal of Economic Geography* **21**:5, 683-714. [Crossref]

45. Frédéric Aubery, David E. Sahn. 2021. Cognitive achievement production in Madagascar: a value-added model approach. *Education Economics* **29**:6, 670-699. [Crossref]

46. Moses Oketch, Caine Rolleston, Jack Rossiter. 2021. Diagnosing the learning crisis: What can value-added analysis contribute?. *International Journal of Educational Development* **87**, 102507. [Crossref]

47. Stéphane Bonhomme, Martin Weidner. 2021. Posterior Average Effects. *Journal of Business & Economic Statistics* 1-38. [Crossref]

48. Nicola Limodio. 2021. Bureaucrat Allocation in the Public Sector: Evidence from the World Bank. *The Economic Journal* **131**:639, 3012-3040. [Crossref]

49. Marianne Bitler, Sean P. Corcoran, Thurston Domina, Emily K. Penner. 2021. Teacher Effects on Student Achievement and Height: A Cautionary Tale. *Journal of Research on Educational Effectiveness* **14**:4, 900-924. [Crossref]

50. Simon Burgess, Shenila Rawal, Eric S. Taylor. 2021. Teacher Peer Observation and Student Test Scores: Evidence from a Field Experiment in English Secondary Schools. *Journal of Labor Economics* **39**:4, 1155-1186. [Crossref]

51. Audrey Amrein-Beardsley, Kevin Close. 2021. Teacher-Level Value-Added Models on Trial: Empirical and Pragmatic Issues of Concern Across Five Court Cases. *Educational Policy* **35**:6, 866-907. [Crossref]

52. José Víctor Orón Semper, Inmaculada Lizasoain, Jaione Abaurrea, Concepción González-García, Esperanza Ayuga-Téllez. 2021. What Kind of School Organizational Decisions Serve to Enhance Sustainable Personal and Social Growth?. *Sustainability* **13**:17, 9995. [Crossref]

53. Andrew E. Clark, Huifu Nong, Hongjia Zhu, Rong Zhu. 2021. Compensating for academic loss: Online learning and student performance during the COVID-19 pandemic. *China Economic Review* **68**, 101629. [Crossref]

54. Barbara Biasi. 2021. The Labor Market for Teachers under Different Pay Schemes. *American Economic Journal: Economic Policy* **13**:3, 63-102. [Abstract] [View PDF article] [PDF with links]

55. David D. Liebowitz. 2021. Teacher evaluation for accountability and growth: Should policy treat them as complements or substitutes?. *Labour Economics* **71**, 102024. [Crossref]

56. Johannes Hörner, Nicolas S Lambert. 2021. Motivational Ratings. *The Review of Economic Studies* **88**:4, 1892-1935. [Crossref]

57. Amalavoyal Chari, Elaine M Liu, Shing-Yi Wang, Yongxiang Wang. 2021. Property Rights, Land Misallocation, and Agricultural Efficiency in China. *The Review of Economic Studies* **88**:4, 1831-1862. [Crossref]

58. Matthew Baird, John Engberg. 2021. Trade-offs in selecting the number of years for estimation of teacher effectiveness. *Education Economics* **29**:4, 341-358. [Crossref]

59. Niu Gao, Anastasia Semykina. 2021. Estimation of Teacher Value Added in the Presence of Missing Data. *Journal of Research on Educational Effectiveness* **14**:3, 676-695. [Crossref]

60. Jing Liu, Monica Lee, Seth Gershenson. 2021. The short- and long-run impacts of secondary school absences. *Journal of Public Economics* **199**, 104441. [Crossref]

61. Clare Leaver, Owen Ozier, Pieter Serneels, Andrew Zeitlin. 2021. Recruitment, Effort, and Retention Effects of Performance Contracts for Civil Servants: Experimental Evidence from Rwandan Primary Schools. *American Economic Review* **111**:7, 2213-2246. [Abstract] [View PDF article] [PDF with links]

62. Bernt Bratsberg, Jeremy Ferwerda, Henning Finseraas, Andreas Kotsadam. 2021. How Settlement Locations and Local Networks Influence Immigrant Political Integration. *American Journal of Political Science* **65**:3, 551-565. [Crossref]

63. Natalie Bau, Jishnu Das, Andres Yi Chang. 2021. New evidence on learning trajectories in a low-income setting. *International Journal of Educational Development* **84**, 102430. [Crossref]

64. Jason Abaluck, Mauricio Caceres Bravo, Peter Hull, Amanda Starc. 2021. Mortality Effects and Choice Across Private Health Insurance Plans. *The Quarterly Journal of Economics* **136**:3, 1557-1610. [Crossref]

65. Lucy C. Sorensen, Stephen B. Holt. 2021. Sorting it Out: The Effects of Charter Expansion on Teacher and Student Composition at Traditional Public Schools. *Economics of Education Review* **82**, 102095. [Crossref]

66. Aaron L. Schwartz. 2021. Accuracy versus Incentives. *American Journal of Health Economics* **7**:3, 333-360. [Crossref]

67. Peter Hinrichs. 2021. What kind of teachers are schools looking for? Evidence from a randomized field experiment. *Journal of Economic Behavior & Organization* **186**, 395-411. [Crossref]

68. David Silver. 2021. Haste or Waste? Peer Pressure and Productivity in the Emergency Department. *The Review of Economic Studies* **88**:3, 1385-1417. [Crossref]

69. Marta De Philippis. 2021. Multi-Task Agents and Incentives: The Case of Teaching and Research for University Professors. *The Economic Journal* **131**:636, 1643-1681. [Crossref]

70. Joshua S. Gans, Avi Goldfarb, Mara Lederman. 2021. Exit, Tweets, and Loyalty. *American Economic Journal: Microeconomics* **13**:2, 68-112. [Abstract] [View PDF article] [PDF with links]

71. Benjamin Feigenberg, Conrad Miller. 2021. Racial Divisions and Criminal Justice: Evidence from Southern State Courts. *American Economic Journal: Economic Policy* **13**:2, 207-240. [Abstract] [View PDF article] [PDF with links]

72. Sabrin Beg, Anne Fitzpatrick, Adrienne M. Lucas. 2021. Gender Bias in Assessments of Teacher Performance. *AEA Papers and Proceedings* **111**, 190-195. [Abstract] [View PDF article] [PDF with links]

73. Gregory Elacqua, Isabela Munevar, Fabio Sanchez, Humberto Santos. 2021. The impact of decentralized decision-making on student outcomes and teacher quality: Evidence from Colombia. *World Development* **141**, 105378. [Crossref]

74. Wesley A. Sims, Kathleen R. King, Wendy M. Reinke, Keith Herman, T. Chris Riley-Tillman. 2021. Development and Preliminary Validity Evidence for the Direct Behavior Rating-Classroom Management (DBR-CM). *Journal of Educational and Psychological Consultation* **31**:2, 215-245. [Crossref]

75. Mengli Song, Andrew J. Wayne, Michael S. Garet, Seth Brown, Jordan Rickles. 2021. Impact of Providing Teachers and Principals with Performance Feedback on Their Practice and Student Achievement: Evidence from a Large-Scale Randomized Experiment. *Journal of Research on Educational Effectiveness* **14**:2, 353-378. [Crossref]

76. Martin J. Williams. 2021. Beyond state capacity: bureaucratic performance, policy implementation and reform. *Journal of Institutional Economics* **17**:2, 339-357. [Crossref]

77. Desire Kedagni, Kala Krishna, Rigissa Megalokonomou, Yingyan Zhao. 2021. Does class size matter? How, and at what cost?. *European Economic Review* **133**, 103664. [Crossref]

78. Carmen Aina, Massimiliano Bratti, Enrico Lippo. 2021. Ranking high schools using university student performance in Italy. *Economia Politica* **38**:1, 293-321. [Crossref]

79. Karthik Muralidharan, Paul Niehaus, Sandip Sukhtankar, Jeffrey Weaver. 2021. Improving Last-Mile Service Delivery Using Phone-Based Monitoring. *American Economic Journal: Applied Economics* **13**:2, 52-82. [Abstract] [View PDF article] [PDF with links]

80. Kristin F. Butcher, Kyung H. Park, Anne Morrison Piehl. 2021. Judge Effects, Case Characteristics, and Plea Bargaining. *Journal of Labor Economics* **39**:S2, S543-S574. [Crossref]

81. Zachary Griffen, Aaron Panofsky. 2021. Ambivalent economizations: the case of value added modeling in teacher evaluation. *Theory and Society* **50**:3, 515-539. [Crossref]

82. Natalie Bau, Jishnu Das, Andres Yi Chang. New Evidence on Learning Trajectories in a Low-Income Setting **2**, . [Crossref]

83. Roddy J. Theobald, Dan D. Goldhaber, Trevor M. Gratz, Kristian L. Holden. 2021. High School English Language Arts Teachers and Postsecondary Outcomes for Students With and Without Disabilities. *Journal of Disability Policy Studies* **31**:4, 217-229. [Crossref]

84. Catherine Haeck, Pierre Lefebvre. 2021. Trends in Cognitive Skill Inequalities by Socio-Economic Status across Canada. *Canadian Public Policy* **47**:1, 88-116. [Crossref]

85. Anikó Fehérvári. 2021. An education programme for the elimination of inequalities in Hungary. *International Journal of Inclusive Education* **2**, 1-15. [Crossref]

86. Sarah R. Cohodes, Elizabeth M. Setren, Christopher R. Walters. 2021. Can Successful Schools Replicate? Scaling Up Boston's Charter School Sector. *American Economic Journal: Economic Policy* **13**:1, 138-167. [Abstract] [View PDF article] [PDF with links]

87. Mathias Reynaert, James M. Sallee. 2021. Who Benefits When Firms Game Corrective Policies?. *American Economic Journal: Economic Policy* **13**:1, 372-412. [Abstract] [View PDF article] [PDF with links]

88. Rafael Camelo, Vladimir Ponczek. 2021. Teacher Turnover and Financial Incentives in Underprivileged Schools: Evidence from a Compensation Policy in a Developing Country. *Economics of Education Review* **80**, 102067. [Crossref]

89. Christopher R. Berry, Anthony Fowler. 2021. Leadership or luck? Randomization inference for leader effects in politics, business, and sports. *Science Advances* **7**:4. . [Crossref]

90. Rodney Hughes, Lauren Dahlin, Tara Tucci. 2021. An Investigation of a Multiple-Measures Teaching Evaluation System and Its Relationship With Students' College-Going Outcomes. *Educational Policy* **35**:1, 131-158. [Crossref]

91. Julie Berry Cullen, Cory Koedel, Eric Parsons. 2021. The Compositional Effect of Rigorous Teacher Evaluation on Workforce Quality. *Education Finance and Policy* **16**:1, 7-41. [Crossref]

92. Joshua Angrist, Peter Hull, Parag A. Pathak, Christopher Walters. 2021. Simple and Credible Value-Added Estimation Using Centralized School Assignment. *SSRN Electronic Journal* . [Crossref]

93. Eric Chyn, Samantha Gold, Justine Hastings. 2021. The returns to early-life interventions for very low birth weight children. *Journal of Health Economics* **75**, 102400. [Crossref]

94. Anne-Françoise de Chambrier, Ariane Baye, Mélanie Tinnes-Vigne, Youssef Tazouti, Joëlle Vlassis, Débora Poncelet, Nadine Giauque, Annick Fagnant, Christophe Luxembourger, Amélie Auquière, Sylvie Kerger, Christophe Dierendonck. 2021. Enhancing children's numerical skills through a play-based intervention at kindergarten and at home: a quasi-experimental study. *Early Childhood Research Quarterly* **54**, 164-178. [Crossref]

95. José Felipe Martínez, María Paz Fernández. Teacher Evaluation with Multiple Indicators: Conceptual and Methodological Considerations Regarding Validity 373-394. [Crossref]

96. Robert M. Klassen, Lisa E. Kim. What Does 'Teacher Effectiveness' Look like? 15-32. [Crossref]

97. Robert M. Klassen, Lisa E. Kim. The Importance of Selecting the Most Effective Teachers 1-11. [Crossref]

98. Andrew Zeitlin. 2020. Teacher Turnover in Rwanda. *Journal of African Economies* **34**. . [Crossref]

99. Charalambos Y. Charalambous, Heather C. Hill, Mark J. Chin, Daniel McGinn. 2020. Mathematical content knowledge and knowledge for teaching: exploring their distinguishability and contribution to student learning. *Journal of Mathematics Teacher Education* **23**:6, 579-613. [Crossref]

100. C. Kirabo Jackson, Shanette C. Porter, John Q. Easton, Alyssa Blanchard, Sebastián Kiguel. 2020. School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment. *American Economic Review: Insights* **2**:4, 491-508. [Abstract] [View PDF article] [PDF with links]

101. Ismail Aslantas. 2020. Impact of Contextual Predictors on Value-Added Teacher Effectiveness Estimates. *Education Sciences* **10**:12, 390. [Crossref]

102. Jenny L. Apriesnig, Dale T. Manning, Jordan F. Suter, Sheryl Magzamen, Jennifer E. Cross. 2020. Academic stars and Energy Stars, an assessment of student academic achievement and school building energy efficiency. *Energy Policy* **147**, 111859. [Crossref]

103. Michael Gottfried, J. Jacob Kirksey, Ethan Hutt. 2020. Can teacher education programs help prepare new kindergarten and first grade teachers to address student absenteeism?. *Children and Youth Services Review* **119**, 105618. [Crossref]

104. Michael A. Gottfried, J. Jacob Kirksey, Christopher S. Ozuna. 2020. Exploring the Links Between Student and Classmate Chronic Absenteeism. *Teachers College Record: The Voice of Scholarship in Education* **122**:12, 1-28. [Crossref]

105. Allison Atteberry, Sarah E. Lacour. 2020. Putting Performance Pay to the Test: Effects of Denver ProComp on the Teacher Workforce and Student Outcomes. *Teachers College Record: The Voice of Scholarship in Education* **122**:12, 1-48. [Crossref]

106. Dara Shifrer. 2020. Contextualizing Educational Disparities and the Evaluation of Teacher Quality. *Social Problems* **49**. . [Crossref]

107. Richard Murphy, Felix Weinhardt. 2020. Top of the Class: The Importance of Ordinal Rank. *The Review of Economic Studies* **87**:6, 2777-2826. [Crossref]

108. Lee Crawfurd, Caine Rolleston. 2020. Long-run effects of teachers in developing countries. *Review of Development Economics* **24**:4, 1279-1299. [Crossref]

109. Geoffrey T. Wodtke. 2020. Regression-based Adjustment for Time-varying Confounders. *Sociological Methods & Research* **49**:4, 906-946. [Crossref]

110. Marina Bassi, Costas Meghir, Ana Reynoso. 2020. Education Quality and Teaching Practices. *The Economic Journal* **130**:631, 1937-1965. [Crossref]

111. Victor Lavy. 2020. Teachers' Pay for Performance in the Long-Run: The Dynamic Pattern of Treatment Effects on Students' Educational and Labour Market Outcomes in Adulthood. *The Review of Economic Studies* **87**:5, 2322-2355. [Crossref]

112. Amelie Schiprowski. 2020. The Role of Caseworkers in Unemployment Insurance: Evidence from Unplanned Absences. *Journal of Labor Economics* **38**:4, 1189-1225. [Crossref]

113. Shao-Hsun Keng. 2020. Gender bias and statistical discrimination against female instructors in student evaluations of teaching. *Labour Economics* **66**, 101889. [Crossref]

114. Li Han, Jiaxin Xie. 2020. Can conditional grants attract better students? Evidence from Chinese teachers' colleges. *Economics of Education Review* **78**, 102034. [Crossref]

115. J. R. Lockwood, D. McCaffrey. 2020. Using hidden information and performance level boundaries to study student–teacher assignments: implications for estimating teacher causal effects. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **183**:4, 1333-1362. [Crossref]

116. Nicole Black, Sonja C. de New. 2020. Short, Heavy and Underrated? Teacher Assessment Biases by Children's Body Size. *Oxford Bulletin of Economics and Statistics* **82**:5, 961-987. [Crossref]

117. Hege Marie Gjefsen, Trude Gunnes. 2020. How school accountability affects teacher mobility and teacher sorting. *Education Economics* **28**:5, 455-473. [Crossref]

118. Mitchell Hoffman, Steven Tadelis. 2020. People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis. *Journal of Political Economy* . [Crossref]

119. Camille Terrier. 2020. Boys lag behind: How teachers' gender biases affect student achievement. *Economics of Education Review* **77**, 101981. [Crossref]

120. Abhijeet Singh. 2020. Learning More with Every Year: School Year Productivity and International Learning Divergence. *Journal of the European Economic Association* **18**:4, 1770-1813. [Crossref]

121. Mitchell Hoffman, Elizabeth Lyons. 2020. A time to make laws and a time to fundraise? On the relation between salaries and time use for state politicians. *Canadian Journal of Economics/Revue canadienne d'économique* **53**:3, 1318-1358. [Crossref]

122. Alejandra Mizala, Ben Schneider. 2020. Promoting quality education in Chile: the politics of reforming teacher careers. *Journal of Education Policy* **35**:4, 529-555. [Crossref]

123. Rune Johan Krumsvik. 2020. Home schooling, remote teaching and digital Bildungin societal crisis. *Nordic Journal of Digital Literacy* **15**:2, 71-85. [Crossref]

124. Saurabh A. Lall, Li-Wei Chen, Peter W. Roberts. 2020. Are we accelerating equity investment into impact-oriented ventures?. *World Development* **131**, 104952. [Crossref]

125. Cecilia Speroni, Alison Wellington, Paul Burkander, Hanley Chiang, Mariesa Herrmann, Kristin Hallgren. 2020. Do Educator Performance Incentives Help Students? Evidence from the Teacher Incentive Fund National Evaluation. *Journal of Labor Economics* **38**:3, 843-872. [Crossref]

126. Youngjin Stephanie Hong, Julia R. Henly. 2020. Supplemental nutrition assistance program and school readiness skills. *Children and Youth Services Review* **114**, 105034. [Crossref]

127. Matthew Ryan Lavery, Audrey Amrein-Beardsley, Tray Geiger, Margarita Pivovarova. 2020. Value-Added Model (VAM) Scholars on Using VAMs for Teacher Evaluation After the Passage of the Every

Student Succeeds Act. *Teachers College Record: The Voice of Scholarship in Education* **122**:7, 1-34. [Crossref]

128. Philip J. Cook, Songman Kang. 2020. Girls to the front: How redshirting and test-score gaps are affected by a change in the school-entry cut date. *Economics of Education Review* **76**, 101968. [Crossref]

129. Mark A. Paige, Audrey Amrein-Beardsley. 2020. "Houston, We Have a Lawsuit": A Cautionary Tale for the Implementation of Value-Added Models for High-Stakes Employment Decisions. *Educational Researcher* **49**:5, 350-359. [Crossref]

130. Sam Sims, Harry Fletcher-Wood. 2020. Identifying the characteristics of effective teacher professional development: a critical review. *School Effectiveness and School Improvement* **30**, 1-17. [Crossref]

131. Atila Abdulkadiroğlu, Parag A. Pathak, Jonathan Schellenberg, Christopher R. Walters. 2020. Do Parents Value School Effectiveness?. *American Economic Review* **110**:5, 1502-1539. [Abstract] [View PDF article] [PDF with links]

132. Anna Godøy, Ingrid Huitfeldt. 2020. Regional variation in health care utilization and mortality. *Journal of Health Economics* **71**, 102254. [Crossref]

133. Nicholas W. Papageorge, Seth Gershenson, Kyung Min Kang. 2020. Teacher Expectations Matter. *The Review of Economics and Statistics* **102**:2, 234-251. [Crossref]

134. Alhasan Allmnakrah, Colin Evers. 2020. The need for a fundamental shift in the Saudi education system: Implementing the Saudi Arabian economic vision 2030. *Research in Education* **106**:1, 22-40. [Crossref]

135. Miraç Fatih İLGÜN. 2020. Industry 4.0 and Transformation in Public Finance: An Assessment by Government Expenditures. *Sosyoekonomi* 237-260. [Crossref]

136. Katherine Landau Wright, Tracey S. Hodges, Sherry Dismuke, Peter Boedeker. 2020. Writing Motivation and Middle School: An Examination of Changes in Students' Motivation for Writing. *Literacy Research and Instruction* **59**:2, 148-168. [Crossref]

137. Markus Nagler, Marc Piopiunik, Martin R. West. 2020. Weak Markets, Strong Teachers: Recession at Career Start and Teacher Effectiveness. *Journal of Labor Economics* **38**:2, 453-500. [Crossref]

138. Dan Goldhaber, John Krieg, Roddy Theobald. 2020. Effective like me? Does having a more productive mentor improve the productivity of mentees?. *Labour Economics* **63**, 101792. [Crossref]

139. Tara Béteille, Namrata Tognatta, Michelle Riboud, Shinsaku Nomura, Yashodhan Ghorpade. South Asia Has Too Few Effective Teachers 115-130. [Crossref]

140. Valentin Verdier. 2020. Estimation and Inference for Linear Models with Two-Way Fixed Effects and Sparsely Matched Data. *The Review of Economics and Statistics* **102**:1, 1-16. [Crossref]

141. Godstime Osekhebhen Eigbiremolen, Jonathan Emenike Ogbuabor, Chioma Sylvia Nwambe. 2020. Estimating Private School Effects for School Children in Peru: Evidence from Individual-level Panel Data. *Journal of International Development* **32**:2, 131-149. [Crossref]

142. Kevin C. Bastian, C. Kevin Fortner. 2020. Is Less More? Subject-Area Specialization and Outcomes in Elementary Schools. *Education Finance and Policy* **15**:2, 357-382. [Crossref]

143. Carolyn J. Heinrich, Jennifer Darling-Aduana, Caroline Martin. 2020. The potential and prerequisites of effective tablet integration in rural Kenya. *British Journal of Educational Technology* **51**:2, 498-514. [Crossref]

144. Larissa da Silva Marioni, Ricardo Da Silva Freguglia, Naercio A Menezes-Filho. 2020. The impacts of teacher working conditions and human capital on student achievement: evidence from brazilian longitudinal data. *Applied Economics* **52**:6, 568-582. [Crossref]

145. Sarah R. Cohodes. 2020. The Long-Run Impacts of Specialized Programming for High-Achieving Students. *American Economic Journal: Economic Policy* **12**:1, 127-166. [Abstract] [View PDF article] [PDF with links]

146. Natalie Bau, Jishnu Das. 2020. Teacher Value Added in a Low-Income Country. *American Economic Journal: Economic Policy* **12**:1, 62-96. [Abstract] [View PDF article] [PDF with links]

147. Christopher Day. 2020. How teachers' individual autonomy may hinder students' academic progress and attainment: Professionalism in practice. *British Educational Research Journal* **46**:1, 247-264. [Crossref]

148. Shu Han, S. Abraham Ravid. 2020. Star Turnover and the Value of Human Capital—Evidence from Broadway Shows. *Management Science* **66**:2, 958-978. [Crossref]

149. Evan Totty. 2020. High school value-added and college outcomes. *Education Economics* **28**:1, 67-95. [Crossref]

150. Robert M. Klassen, Lisa E. Kim, Jade V. Rushby, Lisa Bardach. 2020. Can we improve how we screen applicants for initial teacher education?. *Teaching and Teacher Education* **87**, 102949. [Crossref]

151. Patrick L. Baude, Marcus Casey, Eric A. Hanushek, Gregory R. Phelan, Steven G. Rivkin. 2020. The Evolution of Charter School Quality. *Economica* **87**:345, 158-189. [Crossref]

152. Josh Feng, Xavier Jaravel. 2020. Crafting Intellectual Property Rights: Implications for Patent Assertion Entities, Litigation, and Innovation. *American Economic Journal: Applied Economics* **12**:1, 140-181. [Abstract] [View PDF article] [PDF with links]

153. Bjarne Strøm, Torberg Falch. The role of teacher quality in education production 307-319. [Crossref]

154. Jessalynn James, James Wyckoff. Teacher labor markets: An overview 355-370. [Crossref]

155. Godstime Osekhebhen Eigbiremolen. 2020. Estimating Private School Premium for Primary School Children in Ethiopia: Evidence from Individual-level Panel Data. *Progress in Development Studies* **20**:1, 26-44. [Crossref]

156. Andrew C. Johnston. 2020. Teacher Preferences, Working Conditions, and Compensation Structure. *SSRN Electronic Journal* . [Crossref]

157. Juliane Begenau, Emil Siriwardane. 2020. How Do Private Equity Fees Vary Across Public Pensions?. *SSRN Electronic Journal* . [Crossref]

158. Marit Hinnosaar, Elaine Liu. 2020. Persistence in Alcohol Consumption: Evidence from Migrants. *SSRN Electronic Journal* . [Crossref]

159. Jack Mountjoy, Brent Hickman. 2020. The Returns to College(s): Estimating Value-Added and Match Effects in Higher Education. *SSRN Electronic Journal* . [Crossref]

160. Paul Bruno, Katharine O. Strunk. 2019. Making the Cut: The Effectiveness of Teacher Screening and Hiring in the Los Angeles Unified School District. *Educational Evaluation and Policy Analysis* **41**:4, 426-460. [Crossref]

161. Jan Feld, Nicolás Salamanca, Ulf Zölitz. 2019. Students are almost as effective as professors in university teaching. *Economics of Education Review* **73**, 101912. [Crossref]

162. Andrew Bacher-Hicks, Mark J. Chin, Thomas J. Kane, Douglas O. Staiger. 2019. An experimental evaluation of three teacher quality measures: Value-added, classroom observations, and student surveys. *Economics of Education Review* **73**, 101919. [Crossref]

163. Maria Liria Dacanay, Mark Lawrence Gale, William Bill Turnbull. Development and Validation of the New Teachers' Behavior Inventory 632-639. [Crossref]

164. Joseph Doyle, John Graves, Jonathan Gruber. 2019. Evaluating Measures of Hospital Quality: Evidence from Ambulance Referral Patterns. *The Review of Economics and Statistics* **101**:5, 841-852. [Crossref]

165. Monazza Aslam, Rabea Malik, Shenila Rawal, Pauline Rose, Anna Vignoles. 2019. Do government schools improve learning for poor students? Evidence from rural Pakistan. *Oxford Review of Education* **45**:6, 802-824. [Crossref]

166. Ishuan Li. 2019. Book Review: Education and the commercial mindset , by Abrams, S. E. *The American Economist* **64**:2, 327-329. [Crossref]

167. Dan Goldhaber, Umut Özek. 2019. How Much Should We Rely on Student Test Achievement as a Measure of Success?. *Educational Researcher* **48**:7, 479-483. [Crossref]

168. Kevin C. Bastian. 2019. A Degree Above? The Value-Added Estimates and Evaluation Ratings of Teachers with a Graduate Degree. *Education Finance and Policy* **14**:4, 652-678. [Crossref]

169. Shannon W. Anderson, Amanda Kimball. 2019. Evidence for the Feedback Role of Performance Measurement Systems. *Management Science* **65**:9, 4385-4406. [Crossref]

170. W. Bentley MacLeod, Miguel Urquiola. 2019. Is Education Consumption or Investment? Implications for School Competition. *Annual Review of Economics* **11**:1, 563-589. [Crossref]

171. David C Chan, Michael J Dickstein. 2019. Industry Input in Policy Making: Evidence from Medicare*. *The Quarterly Journal of Economics* **134**:3, 1299-1342. [Crossref]

172. Michela Carlana. 2019. Implicit Stereotypes: Evidence from Teachers' Gender Bias*. *The Quarterly Journal of Economics* **134**:3, 1163-1224. [Crossref]

173. Derek C. Briggs, Jessica L. Alzen. 2019. Making Inferences About Teacher Observation Scores Over Time. *Educational and Psychological Measurement* **79**:4, 636-664. [Crossref]

174. Helen Johnson, Sandra McNally, Heather Rolfe, Jenifer Ruiz-Valenzuela, Robert Savage, Janet Vousden, Clare Wood. 2019. Reprint of: Teaching assistants, computers and classroom management. *Labour Economics* **59**, 17-32. [Crossref]

175. Paul Hanselman. 2019. Access to Effective Teachers and Economic and Racial Disparities in Opportunities to Learn. *The Sociological Quarterly* **60**:3, 498-534. [Crossref]

176. John D. Singleton. 2019. Incentives and the Supply of Effective Charter Schools. *American Economic Review* **109**:7, 2568-2612. [Abstract] [View PDF article] [PDF with links]

177. Thomas S. Dee, Will Dobbie, Brian A. Jacob, Jonah Rockoff. 2019. The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations. *American Economic Journal: Applied Economics* **11**:3, 382-423. [Abstract] [View PDF article] [PDF with links]

178. Javaeria Qureshi, Ben Ost. 2019. DOES TEACHER-FAMILY EXPERIENCE AFFECT TEST SCORES?. *Contemporary Economic Policy* **37**:3, 509-523. [Crossref]

179. Michael Geruso, Timothy Layton. 2019. Upcoding: Evidence from Medicare on Squishy Risk Adjustment. *Journal of Political Economy* . [Crossref]

180. Benjamin Williams. 2019. Controlling for ability using test scores. *Journal of Applied Econometrics* **34**:4, 547-565. [Crossref]

181. Helen Johnson, Sandra McNally, Heather Rolfe, Jenifer Ruiz-Valenzuela, Robert Savage, Janet Vousden, Clare Wood. 2019. Teaching assistants, computers and classroom management. *Labour Economics* **58**, 21-36. [Crossref]

182. William Gort, Rebecca King, Alan Weiss. 2019. Measuring School Efficiency and Effectiveness to Inform the Design of a School Funding Formula. *Economic Record* **95**:S1, 65-78. [Crossref]

183. Jack Britton, Neil Shephard, Anna Vignoles. 2019. A comparison of sample survey measures of earnings of English graduates with administrative data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **182**:3, 719-754. [Crossref]

184. Javier Tourón, Enrique Navarro-Asencio, Luis Lizasoain, Emelina López-González, María José García-San Pedro. 2019. How teachers' practices and students' attitudes towards technology affect mathematics achievement: results and insights from PISA 2012. *Research Papers in Education* **34**:3, 263-275. [Crossref]

185. Kyle Rozema, Max Schanzenbach. 2019. Good Cop, Bad Cop: Using Civilian Allegations to Predict Police Misconduct. *American Economic Journal: Economic Policy* **11**:2, 225-268. [Abstract] [View PDF article] [PDF with links]

186. Alex Bell, Raj Chetty, Xavier Jaravel, Neviana Petkova, John Van Reenen. 2019. Who Becomes an Inventor in America? The Importance of Exposure to Innovation*. *The Quarterly Journal of Economics* **134**:2, 647-713. [Crossref]

187. Mark C. White. 2019. Analysis of Nominally Equivalent Outcome Studies: A Comment on Rickles, Heppen, Allensworth, Sorensen, and Walters (2018). *Educational Researcher* **48**:4, 239-240. [Crossref]

188. Lenora M. Crabtree, Sonyia C. Richardson, Chance W. Lewis. 2019. The Gifted Gap, STEM Education, and Economic Immobility. *Journal of Advanced Academics* **30**:2, 203-231. [Crossref]

189. Karthik Muralidharan, Abhijeet Singh, Alejandro J. Ganimian. 2019. Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. *American Economic Review* **109**:4, 1426-1460. [Abstract] [View PDF article] [PDF with links]

190. Nolan G. Pope. 2019. The effect of teacher ratings on teacher performance. *Journal of Public Economics* **172**, 84-110. [Crossref]

191. Eric Parsons, Cory Koedel, Li Tan. 2019. Accounting for Student Disadvantage in Value-Added Models. *Journal of Educational and Behavioral Statistics* **44**:2, 144-179. [Crossref]

192. Isaac M. Opper. 2019. Does Helping John Help Sue? Evidence of Spillovers in Education. *American Economic Review* **109**:3, 1080-1115. [Abstract] [View PDF article] [PDF with links]

193. John Gathergood, Neale Mahoney, Neil Stewart, Jörg Weber. 2019. How Do Individuals Repay Their Debt? The Balance-Matching Heuristic. *American Economic Review* **109**:3, 844-875. [Abstract] [View PDF article] [PDF with links]

194. Jade Marcus Jenkins, Sudhanshu Handa. 2019. Parenting skills and early childhood development: production function estimates from longitudinal data. *Review of Economics of the Household* **17**:1, 121-147. [Crossref]

195. Marie Hull, Katherine Duch. 2019. One-to-One Technology and Student Outcomes: Evidence From Mooresville's Digital Conversion Initiative. *Educational Evaluation and Policy Analysis* **41**:1, 79-97. [Crossref]

196. Robert Vagi, Margarita Pivovarova, Wendy Miedel Barnard. 2019. Keeping Our Best? A Survival Analysis Examining a Measure of Preservice Teacher Quality and Teacher Attrition. *Journal of Teacher Education* **70**:2, 115-127. [Crossref]

197. Zhilei Tian, Yi Wei, Fang Li. 2019. Who are better teachers? The effects of tenure-track and part-time faculty on student achievement. *China Economic Review* **53**, 140-151. [Crossref]

198. Elizabeth I. Rivera Rodas. 2019. Which New Yorkers vote with their wallets? The impact of teacher quality data on household sorting, and residential and school demographics. *Economics of Education Review* **68**, 104-121. [Crossref]

199. Walker A. Swain, Luis A. Rodriguez, Matthew G. Springer. 2019. Selective retention bonuses for highly effective teachers in high poverty schools: Evidence from Tennessee. *Economics of Education Review* **68**, 148-160. [Crossref]

200. Robert M. Klassen, Lisa E. Kim. 2019. Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review* **26**, 32-51. [Crossref]

201. Francis A. Pearman. 2019. The Effect of Neighborhood Poverty on Math Achievement: Evidence From a Value-Added Design. *Education and Urban Society* **51**:2, 289-307. [Crossref]

202. Karen Phelan Kozlowski, Douglas Lee Lauen. 2019. Understanding Teacher Pay for Performance: Flawed Assumptions and Disappointing Results. *Teachers College Record: The Voice of Scholarship in Education* **121**:2, 1-38. [Crossref]

203. Jamie N. Mikeska, Steven Holtzman, Daniel F. McCaffrey, Shuangshuang Liu, Tamara Shattuck. 2019. Using classroom observations to evaluate science teaching: Implications of lesson sampling for measuring science teaching effectiveness across lesson types. *Science Education* **103**:1, 123-144. [Crossref]

204. 2019. OUP accepted manuscript. *Cesifo Economic Studies* . [Crossref]

205. Nirav Mehta. 2019. Measuring quality for use in incentive schemes: The case of "shrinkage" estimators. *Quantitative Economics* **10**:4, 1537-1577. [Crossref]

206. Nayoung Rim, Roman Rivera, Bocar A. Ba. 2019. In-group Bias and the Police: Evidence from Award Nominations. *SSRN Electronic Journal* . [Crossref]

207. E. Jason Baron. 2019. Union Reform, Performance Pay, and New Teacher Supply: Evidence from Wisconsin's Act 10. *SSRN Electronic Journal* . [Crossref]

208. Umut Özek, Zeyu Xu. 2019. Misattribution of Teacher Value Added. *Education Finance and Policy* **14**:1, 115-148. [Crossref]

209. Pablo Munoz, Mounu Prem. 2019. Productivity and Sectoral Allocation: The Labor Market of School Principals. *SSRN Electronic Journal* . [Crossref]

210. Achyuta Adhvaryu, Vittorio Bassi, Anant Nyshadham, Jorge A. Tamayo. 2019. No Line Left Behind: Assortative Matching Inside the Firm. *SSRN Electronic Journal* . [Crossref]

211. Jeremy A. Cook. 2018. Student demographics, school working conditions, and teacher mobility. *Applied Economics Letters* **25**:21, 1499-1506. [Crossref]

212. Anna M. Makles, Kerstin Schneider, Alexandra Schwarz. 2018. Potenziale schulstatistischer Individualdaten für die Bildungsforschung und Bildungspolitik – Das Beispiel Bremen. *Zeitschrift für Erziehungswissenschaft* **21**:6, 1229-1259. [Crossref]

213. Mette Gørtz, Eva Rye Johansen, Marianne Simonsen. 2018. Academic achievement and the gender composition of preschool staff. *Labour Economics* **55**, 241-258. [Crossref]

214. Thurston Domina, Nikolas Pharris-Ciurej, Andrew M. Penner, Emily K. Penner, Quentin Brummet, Sonya R. Porter, Tanya Sanabria. 2018. Is Free and Reduced-Price Lunch a Valid Measure of Educational Disadvantage?. *Educational Researcher* **47**:9, 539-555. [Crossref]

215. Kevin M. Williams, Teny Maghakian Shapiro. 2018. Academic achievement across the day: Evidence from randomized class schedules. *Economics of Education Review* **67**, 158-170. [Crossref]

216. Peng Ding, Luke Keele. 2018. Rank tests in unmatched clustered randomized trials applied to a study of teacher training. *The Annals of Applied Statistics* **12**:4. . [Crossref]

217. Scott E. Carrell, Mark Hoekstra, Elira Kuka. 2018. The Long-Run Effects of Disruptive Peers. *American Economic Review* **108**:11, 3377-3415. [Abstract] [View PDF article] [PDF with links]

218. Joseph G. Altonji, Richard K. Mansfield. 2018. Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects. *American Economic Review* **108**:10, 2902-2946. [Abstract] [View PDF article] [PDF with links]

219. Tim T. Morris, Neil M. Davies, Danny Dorling, Rebecca C. Richmond, George Davey Smith. 2018. Testing the validity of value-added measures of educational progress with genetic data. *British Educational Research Journal* **44**:5, 725-747. [Crossref]

220. Kelly N. Vosters, Cassandra M. Guarino, Jeffrey M. Wooldridge. 2018. Understanding and evaluating the SAS® EVAAS® Univariate Response Model (URM) for measuring teacher effectiveness. *Economics of Education Review* **66**, 191-205. [Crossref]

221. Mark Hoekstra, Pierre Mouganie, Yaojing Wang. 2018. Peer Quality and the Academic Benefits to Attending Better Schools. *Journal of Labor Economics* **36**:4, 841-884. [Crossref]

222. Lisa Fitton, Autumn L. McIlraith, Carla L. Wood. 2018. Shared Book Reading Interventions With English Learners: A Meta-Analysis. *Review of Educational Research* **88**:5, 712-751. [Crossref]

223. Brian A. Jacob, Jonah E. Rockoff, Eric S. Taylor, Benjamin Lindy, Rachel Rosen. 2018. Teacher applicant hiring and teacher performance: Evidence from DC public schools. *Journal of Public Economics* **166**, 81-97. [Crossref]

224. Fred C. Lunenburg. National Policy/Standards 379-405. [Crossref]

225. Jun Wang, Bo Li. 2018. Supplementary education, student development and education equity: evidence from primary schools in Beijing, China. *Education Economics* **26**:5, 459-487. [Crossref]

226. Alan Sánchez, Abhijeet Singh. 2018. Accessing higher education in developing countries: Panel data analysis from India, Peru, and Vietnam. *World Development* **109**, 261-278. [Crossref]

227. Jason A. Grissom, Richard S. L. Blissett, Hajime Mitani. 2018. Evaluating School Principals: Supervisor Ratings of Principal Practice and Principal Job Performance. *Educational Evaluation and Policy Analysis* **40**:3, 446-472. [Crossref]

228. Isaac Sorkin. 2018. Ranking Firms Using Revealed Preference*. *The Quarterly Journal of Economics* **133**:3, 1331-1393. [Crossref]

229. Raj Chetty, Nathaniel Hendren. 2018. The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates*. *The Quarterly Journal of Economics* **133**:3, 1163-1228. [Crossref]

230. Arlen Guarín, Carlos Medina, Christian Posso. 2018. Calidad, cobertura y costos ocultos de la educación secundaria pública y privada en Colombia. *Revista Desarrollo y Sociedad* :81, 61-114. [Crossref]

231. Hugh Macartney, John D. Singleton. 2018. School boards and student segregation. *Journal of Public Economics* **164**, 165-182. [Crossref]

232. Nicolas Grau, Daniel Hojman, Alejandra Mizala. 2018. School closure and educational attainment: Evidence from a market-based system. *Economics of Education Review* **65**, 1-17. [Crossref]

233. Stéphane Lavertu, Travis St. Clair. 2018. Beyond spending levels: Revenue uncertainty and the performance of local governments. *Journal of Urban Economics* **106**, 59-80. [Crossref]

234. Efrat Kass, Erez C. Miller. 2018. Career choice among academically excellent students: Choosing teaching career as a corrective experience. *Teaching and Teacher Education* **73**, 90-98. [Crossref]

235. Eric S. Taylor. 2018. Skills, Job Tasks, and Productivity in Teaching: Evidence from a Randomized Trial of Instruction Practices. *Journal of Labor Economics* **36**:3, 711-742. [Crossref]

236. Johan Coenen, Ilja Cornelisz, Wim Groot, Henriette Maassen van den Brink, Chris Van Klaveren. 2018. TEACHER CHARACTERISTICS AND THEIR EFFECTS ON STUDENT TEST SCORES: A SYSTEMATIC REVIEW. *Journal of Economic Surveys* **32**:3, 848-877. [Crossref]

237. David Blazar. 2018. Validating Teacher Effects on Students' Attitudes and Behaviors: Evidence from Random Assignment of Teachers to Students. *Education Finance and Policy* **13**:3, 281-309. [Crossref]

238. Cheng Cheng. 2018. Greek membership and academic performance: evidence from student-level data. *Applied Economics* **50**:29, 3185-3195. [Crossref]

239. Albert Cheng, Gema Zamarro. 2018. Measuring teacher non-cognitive skills and its impact on students: Insight from the Measures of Effective Teaching Longitudinal Database. *Economics of Education Review* **64**, 251-260. [Crossref]

240. Barbara Bruns, Leandro Costa, Nina Cunha. 2018. Through the looking glass: Can classroom observation and coaching improve teacher performance in Brazil?. *Economics of Education Review* **64**, 214-250. [Crossref]

241. Andrew J. Hill, Daniel B. Jones. 2018. A teacher who knows me: The academic benefits of repeat student-teacher matches. *Economics of Education Review* **64**, 1-12. [Crossref]

242. Brian Stacy, Cassandra Guarino, Jeffrey Wooldridge. 2018. Does the precision and stability of value-added estimates of teacher performance depend on the types of students they serve?. *Economics of Education Review* **64**, 50-74. [Crossref]

243. Ben Backes, Dan Goldhaber, Whitney Cade, Kate Sullivan, Melissa Dodson. 2018. Can UTeach? Assessing the relative effectiveness of STEM teachers. *Economics of Education Review* **64**, 184-198. [Crossref]

244. Stephen Machin, Sandra McNally, Martina Viarengo. 2018. Changing How Literacy Is Taught: Evidence on Synthetic Phonics. *American Economic Journal: Economic Policy* **10**:2, 217-241. [Abstract] [View PDF article] [PDF with links]

245. Andrea Canales, Luis Maldonado. 2018. Teacher quality and student achievement in Chile: Linking teachers' contribution and observable characteristics. *International Journal of Educational Development* **60**, 33-50. [Crossref]

246. Todd Pugatch, Elizabeth Schroeder. 2018. Teacher pay and student performance: evidence from the Gambian hardship allowance. *Journal of Development Effectiveness* **10**:2, 249-276. [Crossref]

247. Paul Hanselman. 2018. Do School Learning Opportunities Compound or Compensate for Background Inequalities? Evidence from the Case of Assignment to Effective Teachers. *Sociology of Education* **91**:2, 132-158. [Crossref]

248. Benny Geys, Rune J. Sørensen. 2018. Never Change a Winning Policy? Public Sector Performance and Politicians' Preferences for Reforms. *Public Administration Review* **78**:2, 206-216. [Crossref]

249. Ricardo Perez-Truglia. 2018. Political Conformity: Event-Study Evidence from the United States. *The Review of Economics and Statistics* **100**:1, 14-28. [Crossref]

250. Ben Backes, Michael Hansen. 2018. The Impact of Teach for America on Non-Test Academic Outcomes. *Education Finance and Policy* **13**:2, 168-193. [Crossref]

251. Will Dobbie, Jacob Goldin, Crystal S. Yang. 2018. The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges. *American Economic Review* **108**:2, 201-240. [Abstract] [View PDF article] [PDF with links]

252. Ben Backes, James Cowan, Dan Goldhaber, Cory Koedel, Luke C. Miller, Zeyu Xu. 2018. The common core conundrum: To what extent should we worry that changes to assessments will affect test-based measures of teacher performance?. *Economics of Education Review* **62**, 48-65. [Crossref]

253. Dan Goldhaber, Vanessa Quince, Roddy Theobald. 2018. Has It Always Been This Way? Tracing the Evolution of Teacher Quality Gaps in U.S. Public Schools. *American Educational Research Journal* **55**:1, 171-201. [Crossref]

254. Atila Abdulkadiroğlu, Parag A. Pathak, Christopher R. Walters. 2018. Free to Choose: Can School Choice Reduce Student Achievement?. *American Economic Journal: Applied Economics* **10**:1, 175-206. [Abstract] [View PDF article] [PDF with links]

255. Stephen L. Morgan, Daniel T. Shackelford. School and Teacher Effects 513-534. [Crossref]

256. Cassandra M. Guarino. Value-Added and Growth Models in Education Research 213-230. [Crossref]

257. Peter Hull. 2018. Estimating Hospital Quality with Quasi-Experimental Data. *SSRN Electronic Journal* . [Crossref]

258. Jens Prufer, Patricia Prufer. 2018. Data Science for Institutional and Organizational Economics. *SSRN Electronic Journal* . [Crossref]

259. Shannon W. Anderson, Amanda Kimball. 2018. Evidence for the Feedback Role of Performance Measurement Systems. *SSRN Electronic Journal* . [Crossref]

260. Mika Kortelainen, Kalle Manninen. 2018. Effectiveness of Private and Public High Schools: Evidence from Finland. *SSRN Electronic Journal* . [Crossref]

261. Stephanie Lee, Ju-Ho Lee, Youngsik Jeong. 2018. The Effects of Digital Textbooks on Students' Academic Performance, Learning Skills, and Digital Addiction. *SSRN Electronic Journal* . [Crossref]

262. My Nguyen. 2018. The Relationship between Race-Congruent Students and Teachers: Does Racial Discrimination Exist?. *SSRN Electronic Journal* . [Crossref]

263. Seth Gershenson, Michael S. Hayes. 2018. The Implications of Summer Learning Loss for Value-Added Estimates of Teacher Effectiveness. *Educational Policy* **32**:1, 55-85. [Crossref]

264. Mark Egan, Gregor Matvos, Amit Seru. 2018. Arbitration with Uninformed Consumers. *SSRN Electronic Journal* . [Crossref]

265. Adam Isen, Maya Rossin-Slater, Reed Walker. 2017. Relationship between season of birth, temperature exposure, and later life wellbeing. *Proceedings of the National Academy of Sciences* **114**:51, 13447-13452. [Crossref]

266. Jamie N. Mikeska, Tamara Shattuck, Steven Holtzman, Daniel F. McCaffrey, Nancy Duchesneau, Yi Qi, Leslie Stickler. 2017. Understanding science teaching effectiveness: examining how science-specific and generic instructional practices relate to student achievement in secondary science classrooms. *International Journal of Science Education* **39**:18, 2594-2623. [Crossref]

267. Dario Sansone. 2017. Why does teacher gender matter?. *Economics of Education Review* **61**, 9-18. [Crossref]

268. Jeffrey Penney. 2017. A self-reference problem in test score normalization. *Economics of Education Review* **61**, 79-84. [Crossref]

269. Dara Shifrer, Ruth López Turley, Holly Heard. 2017. Do Teacher Financial Awards Improve Teacher Retention and Student Achievement in an Urban Disadvantaged School District?. *American Educational Research Journal* **54**:6, 1117-1153. [Crossref]

270. Julie A. Marsh, Susan Bush-Mecenas, Katharine O. Strunk, Jane Arnold Lincove, Alice Huguet. 2017. Evaluating Teachers in the Big Easy: How Organizational Context Shapes Policy Responses in New Orleans. *Educational Evaluation and Policy Analysis* **39**:4, 539-570. [Crossref]

271. Dan Goldhaber, Trevor Gratz, Roddy Theobald. 2017. What's in a teacher test? Assessing the relationship between teacher licensure test scores and student STEM achievement and course-taking. *Economics of Education Review* **61**, 112-129. [Crossref]

272. Marie C. Hull. 2017. The time-varying role of the family in student time use and achievement. *IZA Journal of Labor Economics* **6**:1. . [Crossref]

273. Michael Jones, Michael T. Hartney. 2017. Show Who the Money? Teacher Sorting Patterns and Performance Pay across U.S. School Districts. *Public Administration Review* **77**:6, 919-931. [Crossref]

274. Marine de Talancé. 2017. Better Teachers, Better Results? Evidence from Rural Pakistan. *The Journal of Development Studies* **53**:10, 1697-1713. [Crossref]

275. Yi Long, Christopher Nyland, Russell Smyth. 2017. Fiscal Decentralisation, the Knowledge Economy and School Teachers' Wages in Urban China. *The Journal of Development Studies* **53**:10, 1731-1747. [Crossref]

276. Yyannú Cruz-Aguayo, Pablo Ibarrarán, Norbert Schady. 2017. Do tests applied to teachers predict their effectiveness?. *Economics Letters* **159**, 108-111. [Crossref]

277. Edward Sloat, Audrey Amrein-Beardsley, Kent E. Sabo. 2017. Examining the Factor Structure Underlying the TAP System for Teacher and Student Advancement. *AERA Open* **3**:4, 233285841773552. [Crossref]

278. Helen F. Ladd, Charles T. Clotfelter, John B. Holbein. 2017. The Growing Segmentation of the Charter School Sector in North Carolina. *Education Finance and Policy* **12**:4, 536-563. [Crossref]

279. Jeffrey Penney. 2017. Racial Interaction Effects and Student Achievement. *Education Finance and Policy* **12**:4, 447-467. [Crossref]

280. Andrew McEachin, Allison Atteberry. 2017. The Impact of Summer Learning Loss on Measures of School Performance. *Education Finance and Policy* **12**:4, 468-491. [Crossref]

281. Matthew P. Steinberg, Matthew A. Kraft. 2017. The Sensitivity of Teacher Performance Ratings to the Design of Teacher Evaluation Systems. *Educational Researcher* **46**:7, 378-396. [Crossref]

282. Geoffrey T. Wodtke, Matthew Parbst. 2017. Neighborhoods, Schools, and Academic Achievement: A Formal Mediation Analysis of Contextual Effects on Reading and Mathematics Abilities. *Demography* **54**:5, 1653-1676. [Crossref]

283. Dongwoo Kim, Cory Koedel, Shawn Ni, Michael Podgursky. 2017. Labor market frictions and production efficiency in public schools. *Economics of Education Review* **60**, 54-67. [Crossref]

284. Dan Goldhaber, James Cowan, Roddy Theobald. 2017. Evaluating Prospective Teachers: Testing the Predictive Validity of the edTPA. *Journal of Teacher Education* **68**:4, 377-393. [Crossref]

285. David Lazer, Jason Radford. 2017. Data ex Machina: Introduction to Big Data. *Annual Review of Sociology* **43**:1, 19-39. [Crossref]

286. Vladimir Kogan, Stéphane Lavertu, Zachary Peskowitz. 2017. Direct Democracy and Administrative Disruption. *Journal of Public Administration Research and Theory* **27**:3, 381-399. [Crossref]

287. Jason A. Grissom, Susanna Loeb. 2017. Assessing Principals' Assessments: Subjective Evaluations of Teacher Effectiveness in Low- and High-Stakes Environments. *Education Finance and Policy* **12**:3, 369-395. [Crossref]

288. Halil Dundar, Benoît Millot, Michelle Riboud, Mari Shojo, Harsha Aturupane, Sangeeta Goyal, Dhushyanth Raju. Primary and Secondary Education: The Quality Challenge 81-121. [Crossref]

289. Jesse Rothstein. 2017. Measuring the Impacts of Teachers: Comment. *American Economic Review* **107**:6, 1656-1684. [Abstract] [View PDF article] [PDF with links]

290. Raj Chetty, John N. Friedman, Jonah E. Rockoff. 2017. Measuring the Impacts of Teachers: Reply. *American Economic Review* **107**:6, 1685-1717. [Abstract] [View PDF article] [PDF with links]

291. Dennis Epple, Richard E. Romano, Miguel Urquiola. 2017. School Vouchers: A Survey of the Economics Literature. *Journal of Economic Literature* **55**:2, 441-492. [Abstract] [View PDF article] [PDF with links]

292. Maarten Goos, Anna Salomons. 2017. Measuring teaching quality in higher education: assessing selection bias in course evaluations. *Research in Higher Education* **58**:4, 341-364. [Crossref]

293. Vince Geiger, Judy Anderson, Derek Hurrell. 2017. A case study of effective practice in mathematics teaching and learning informed by Valsiner's zone theory. *Mathematics Education Research Journal* **29**:2, 143-161. [Crossref]

294. Derek C. Briggs, Nathan Dadey. 2017. Principal holistic judgments and high-stakes evaluations of teachers. *Educational Assessment, Evaluation and Accountability* **29**:2, 155-178. [Crossref]

295. Joshua D. Angrist, Peter D. Hull, Parag A. Pathak, Christopher R. Walters. 2017. Leveraging Lotteries for School Value-Added: Testing and Estimation*. *The Quarterly Journal of Economics* **132**:2, 871-919. [Crossref]

296. Jessica L. Alzen, Erin M. Fahle, Benjamin W. Domingue. 2017. The Implications of Reduced Testing for Teacher Accountability. *AERA Open* **3**:2, 233285841770441. [Crossref]

297. Dan Goldhaber, Cyrus Grout, Nick Huntington-Klein. 2017. Screen Twice, Cut Once: Assessing the Predictive Validity of Applicant Selection Tools. *Education Finance and Policy* **12**:2, 197-223. [Crossref]

298. Seth Gershenson, Alison Jacknowitz, Andrew Brannegan. 2017. Are Student Absences Worth the Worry in U.S. Primary Schools?. *Education Finance and Policy* **12**:2, 137-165. [Crossref]

299. Jens Dietrichson, Martin Bøg, Trine Filges, Anne-Marie Klint Jørgensen. 2017. Academic Interventions for Elementary and Middle School Students With Low Socioeconomic Status: A Systematic Review and Meta-Analysis. *Review of Educational Research* **87**:2, 243-282. [Crossref]

300. Lisa Grazzini. 2017. The Importance of the Quality of Education: Some Determinants and its Effects on Earning Returns and Economic Growth. *ECONOMIA PUBBLICA* :2, 43-82. [Crossref]

301. Dhushyanth Raju. 2017. Public School Teacher Management in Sri Lanka. *South Asia Economic Journal* **18**:1, 39-63. [Crossref]

302. David Blazar, Matthew A. Kraft. 2017. Teacher and Teaching Effects on Students' Attitudes and Behaviors. *Educational Evaluation and Policy Analysis* **39**:1, 146-170. [Crossref]

303. Min Sun, Susanna Loeb, Jason A. Grissom. 2017. Building Teacher Teams. *Educational Evaluation and Policy Analysis* **39**:1, 104-125. [Crossref]

304. Allison Atteberry, Susanna Loeb, James Wyckoff. 2017. Teacher Churning. *Educational Evaluation and Policy Analysis* **39**:1, 3-30. [Crossref]

305. Susan Payne Carter, Kyle Greenberg, Michael S. Walker. 2017. The impact of computer usage on academic performance: Evidence from a randomized trial at the United States Military Academy. *Economics of Education Review* **56**, 118-132. [Crossref]

306. Massimo G. Colombo, Samuele Murtinu. 2017. Venture Capital Investments in Europe and Portfolio Firms' Economic Performance: Independent Versus Corporate Investors. *Journal of Economics & Management Strategy* **26**:1, 35-66. [Crossref]

307. Kimberlee C. Everson. 2017. Value-Added Modeling and Educational Accountability. *Review of Educational Research* **87**:1, 35-70. [Crossref]

308. Hanley S. Chiang, Melissa A. Clark, Sheena McConnell. 2017. Supplying Disadvantaged Schools with Effective Teachers: Experimental Evidence on Secondary Math Teachers from Teach For America. *Journal of Policy Analysis and Management* **36**:1, 97-125. [Crossref]

309. J. Rothstein, T. von Wachter. Social Experiments in the Labor Market 555-637. [Crossref]

310. Craig M. Burnett, Vladimir Kogan. 2017. The Politics of Potholes: Service Quality and Retrospective Voting in Local Elections. *The Journal of Politics* **79**:1, 302-314. [Crossref]

311. Mark Egan, Stefan Lewellen. 2017. The Cross Section of Bank Value. *SSRN Electronic Journal* . [Crossref]

312. Teevrat Garg, Maulik Jagnani, Vis Taraz. 2017. Human Capital Costs of Climate Change: Evidence from Test Scores in India. *SSRN Electronic Journal* . [Crossref]

313. Barbara Biasi. 2017. Unions, Salaries, and the Market for Teachers: Evidence from Wisconsin. *SSRN Electronic Journal* . [Crossref]

314. Dionissi Aliprantis. 2017. Evidence on the Production of Cognitive Achievement from Moving to Opportunity. *SSRN Electronic Journal* . [Crossref]

315. Gonzalo Gaete. 2017. Follow the Leader: Student Strikes, School Absenteeism and Long Term Implications for Education Outcomes. *SSRN Electronic Journal* . [Crossref]

316. Jan Feld, Nicolas Salamanca, Ulf ZZlitz. 2017. Students Are Almost as Effective as Professors in University Teaching. *SSRN Electronic Journal* . [Crossref]

317. Barbara Biasi. 2017. Retirement Age Effects of Pension and Salary Reforms: Evidence from Wisconsin Teachers. *SSRN Electronic Journal* . [Crossref]

318. Liran Einav, Amy Finkelstein, Neale Mahoney. 2017. Provider Incentives and Healthcare Costs: Evidence from Long-Term Care Hospitals. *SSRN Electronic Journal* . [Crossref]

319. Evan Totty. 2017. High School Value-Added and Labor Market Outcomes. *SSRN Electronic Journal* . [Crossref]

320. Kumar Aniket. 2017. Man Versus Machine: Automation, Market Structure and Skills. *SSRN Electronic Journal* . [Crossref]

321. Heinrich Hock, Eric Isenberg. 2017. Methods for Accounting for Co-Teaching in Value-Added Models. *Statistics and Public Policy* **4**:1, 1-11. [Crossref]

322. Dan Goldhaber, Richard Startz. 2017. On the Distribution of Worker Productivity: The Case of Teacher Effectiveness and Student Achievement. *Statistics and Public Policy* **4**:1, 1-12. [Crossref]

323. Anthony Milanowski. 2017. Lower Performance Evaluation Practice Ratings for Teachers of Disadvantaged Students. *AERA Open* **3**:1, 233285841668555. [Crossref]

324. Cory Koedel, Diyi Li, Morgan S. Polikoff, Tenice Hardaway, Stephani L. Wrabel. 2017. Mathematics Curriculum Effects on Student Achievement in California. *AERA Open* **3**:1, 233285841769051. [Crossref]

325. Aenneli Houkes-Hommes, Bas ter Weel, Karen van der Wiel. 2016. Measuring the Contribution of Primary-School Teachers to Education Outcomes in The Netherlands. *De Economist* **164**:4, 357-364. [Crossref]

326. Han Feng, Jiayao Li. 2016. Head teachers, peer effects, and student achievement. *China Economic Review* **41**, 268-283. [Crossref]

327. Eric A. Hanushek, Steven G. Rivkin, Jeffrey C. Schiman. 2016. Dynamic effects of teacher turnover on the quality of instruction. *Economics of Education Review* **55**, 132-148. [Crossref]

328. Esteban M. Aucejo, Teresa Foy Romano. 2016. Assessing the effect of school days and absences on test score performance. *Economics of Education Review* **55**, 70-87. [Crossref]

329. Andrew J. McEachin, Richard Osbourne Welsh, Dominic James Brewer. 2016. The Variation in Student Achievement and Behavior Within a Portfolio Management Model. *Educational Evaluation and Policy Analysis* **38**:4, 669-691. [Crossref]

330. José Felipe Martínez, Jonathan Schweig, Pete Goldschmidt. 2016. Approaches for Combining Multiple Measures of Teacher Performance. *Educational Evaluation and Policy Analysis* **38**:4, 738-756. [Crossref]

331. John Garen. 2016. Assessing the literature on school reform from an entrepreneurship perspective. *Journal of Entrepreneurship and Public Policy* **5**:3, 383-403. [Crossref]

332. Stéphane Lavertu. 2016. We All Need Help: "Big Data" and the Mismeasure of Public Administration. *Public Administration Review* **76**:6, 864-872. [Crossref]

333. Patrick Kline, Christopher R. Walters. 2016. Evaluating Public Programs with Close Substitutes: The Case of Head Start*. *The Quarterly Journal of Economics* **131**:4, 1795-1848. [Crossref]

334. William H. Schmidt, Nathan A. Burroughs. 2016. Influencing public school policy in the United States: the role of large-scale assessments. *Research Papers in Education* **31**:5, 567-577. [Crossref]

335. . Reductions in Inequality: A Policy Perspective 129-170. [Crossref]

336. David Card, Laura Giuliano. 2016. Can Tracking Raise the Test Scores of High-Ability Minority Students?. *American Economic Review* **106**:10, 2783-2816. [Abstract] [View PDF article] [PDF with links]

337. Rosa Rodríguez, Gonzalo Rubio. 2016. Teaching quality and academic research. *International Review of Economics Education* **23**, 10-27. [Crossref]

338. Jessica Van Parys. 2016. Variation in Physician Practice Styles within and across Emergency Departments. *PLOS ONE* **11**:8, e0159882. [Crossref]

339. Brian Jacob, Jesse Rothstein. 2016. The Measurement of Student Ability in Modern Assessment Systems. *Journal of Economic Perspectives* **30**:3, 85-108. [Citation] [View PDF article] [PDF with links]

340. Jason B. Cook, Richard K. Mansfield. 2016. Task-specific experience and task-specific talent: Decomposing the productivity of high school teachers. *Journal of Public Economics* **140**, 51-72. [Crossref]

341. M. Caridad Araujo, Pedro Carneiro, Yyannú Cruz-Aguayo, Norbert Schady. 2016. Teacher Quality and Learning Outcomes in Kindergarten *. *The Quarterly Journal of Economics* **131**:3, 1415-1453. [Crossref]

342. Joniada Milla, Ernesto San Martín, Sébastien Van Bellegem. 2016. Higher Education Value Added Using Multiple Outcomes. *Journal of Educational Measurement* **53**:3, 368-400. [Crossref]

343. Martin Carnoy, Tatiana Khavenson, Prashant Loyalka, William H. Schmidt, Andrey Zakharov. 2016. Revisiting the Relationship Between International Assessment Outcomes and Educational Production. *American Educational Research Journal* **53**:4, 1054-1085. [Crossref]

344. Eric Parsons. 2016. Does Attending a Low-Achieving School Affect High-Performing Student Outcomes?. *Teachers College Record: The Voice of Scholarship in Education* **118**:8, 1-36. [Crossref]

345. James Cowan, Dan Goldhaber. 2016. National Board Certification and Teacher Effectiveness: Evidence From Washington State. *Journal of Research on Educational Effectiveness* **9**:3, 233-258. [Crossref]

346. Michela Braga, Marco Paccagnella, Michele Pellizzari. 2016. The Impact of College Teaching on Students' Academic and Labor Market Outcomes. *Journal of Labor Economics* **34**:3, 781-822. [Crossref]

347. Matthew P. Steinberg, Morgaen L. Donaldson. 2016. The New Educational Accountability: Understanding the Landscape of Teacher Evaluation in the Post-NCLB Era. *Education Finance and Policy* **11**:3, 340-359. [Crossref]

348. Hanley Chiang, Stephen Lipscomb, Brian Gill. 2016. Is School Value Added Indicative of Principal Quality?. *Education Finance and Policy* **11**:3, 283-309. [Crossref]

349. Amitabh Chandra, Douglas Staiger. 2016. Sources of Inefficiency in Healthcare and Education. *American Economic Review* **106**:5, 383-387. [Abstract] [View PDF article] [PDF with links]

350. Joshua Angrist, Peter Hull, Parag Pathak, Christopher Walters. 2016. Interpreting Tests of School VAM Validity. *American Economic Review* **106**:5, 388-392. [Abstract] [View PDF article] [PDF with links]

351. Raj Chetty, John N. Friedman, Jonah Rockoff. 2016. Using Lagged Outcomes to Evaluate Bias in Value-Added Models. *American Economic Review* **106**:5, 393-399. [Abstract] [View PDF article] [PDF with links]

352. Mark Ehlert, Cory Koedel, Eric Parsons, Michael Podgursky. 2016. Selecting Growth Measures for Use in School Evaluation Systems. *Educational Policy* **30**:3, 465-500. [Crossref]

353. Pedro Bernal, Nikolas Mittag, Javaeria A. Qureshi. 2016. Estimating effects of school quality using multiple proxies. *Labour Economics* **39**, 1-10. [Crossref]

354. Seth Gershenson. 2016. Linking Teacher Quality, Student Attendance, and Student Achievement. *Education Finance and Policy* **11**:2, 125-149. [Crossref]

355. David Blazar, Erica Litke, Johanna Barmore. 2016. What Does It Mean to Be Ranked a "High" or "Low" Value-Added Teacher? Observing Differences in Instructional Quality Across Districts. *American Educational Research Journal* **53**:2, 324-359. [Crossref]

356. Jane Waldfogel. 2016. Presidential Address:The Next War on Poverty. *Journal of Policy Analysis and Management* **35**:2, 267-278. [Crossref]

357. Eric A. Hanushek. 2016. School human capital and teacher salary policies. *Journal of Professional Capital and Community* **1**:1, 23-40. [Crossref]

358. Ludger Woessmann. 2016. The economic case for education. *Education Economics* **24**:1, 3-32. [Crossref]

359. Ryan D. Shaw. 2016. Arts teacher evaluation: How did we get here?. *Arts Education Policy Review* **117**:1, 1-12. [Crossref]

360. Emmanuel Thanassoulis, Kristof De Witte, Jill Johnes, Geraint Johnes, Giannis Karagiannis, Conceição S. Portela. Applications of Data Envelopment Analysis in Education 367-438. [Crossref]

361. D. Figlio, K. Karbownik, K.G. Salvanes. Education Research and Administrative Data 75-138. [Crossref]

362. Kimberly Kappler Hewitt, Audrey Amrein-Beardsley. Might the Tidal Wave Recede? Considering the Future of Student Growth Measures in Teacher Accountability 261-283. [Crossref]

363. Margarita Pivovarova, Audrey Amrein-Beardsley, Jennifer Broatch. 2016. Value-Added Models (VAMs): Caveat Emptor. *Statistics and Public Policy* **3**:1, 1-9. [Crossref]

364. Mariesa Herrmann, Elias Walsh, Eric Isenberg. 2016. Shrinkage of Value-Added Estimates and Characteristics of Students with Hard-to-Predict Achievement Levels. *Statistics and Public Policy* **3**:1, 1-10. [Crossref]

365. Hugh Macartney. 2016. The Dynamic Effects of Educational Accountability. *Journal of Labor Economics* **34**:1, 1-28. [Crossref]

366. Cory Koedel, Jiaxi Li. 2016. THE EFFICIENCY IMPLICATIONS OF USING PROPORTIONAL EVALUATIONS TO SHAPE THE TEACHING WORKFORCE. *Contemporary Economic Policy* **34**:1, 47-62. [Crossref]

367. Matthew M. Chingos. 2016. Instructional Quality and Student Learning in Higher Education: Evidence from Developmental Algebra Courses. *The Journal of Higher Education* **87**:1, 84-114. [Crossref]

368. Jonathan T. Rothwell. 2016. Classroom Inequality and the Cognitive Race Gap: Evidence from 4-Year Olds in Public PreK. *SSRN Electronic Journal* . [Crossref]

369. Johannes Horner, Nicolas S. Lambert. 2016. Motivational Ratings. *SSRN Electronic Journal* . [Crossref]

370. Loretti Dobrescu, Alberto Motta, Chun Yee Wong. 2016. Can Videogames Fix Online Learning? The Effect of Educational Videogames on Student Learning and Engagement. *SSRN Electronic Journal* . [Crossref]

371. Brian Gill, Jennifer Lerner, Paul Meosky. 2016. Reimagining Accountability in K-12 Education: A Behavioral Science Perspective. *SSRN Electronic Journal* . [Crossref]

372. Mike Helal. 2016. How Principals Affect Schools. *SSRN Electronic Journal* . [Crossref]

373. Andrew McEachin, Allison Atteberry. 2016. The Impact of Summer Learning Loss on Measures of School Performance. *SSRN Electronic Journal* . [Crossref]

374. Matthew D. Hendricks. 2016. Teacher Characteristics and Productivity: Quasi-Experimental Evidence from Teacher Mobility. *SSRN Electronic Journal* . [Crossref]

375. Roman Sittl, Arne Jonas Warnke. 2016. Competitive Balance and Assortative Matching in the German Bundesliga. *SSRN Electronic Journal* . [Crossref]

376. Shu Han. 2016. Actor Replacement as a Pseudo-Natural Experiment in the Value of Human Capital-Evidence from Broadway Shows. *SSRN Electronic Journal* . [Crossref]

377. Joshua S. Gans. 2016. Exit, Tweets, and Loyalty. *SSRN Electronic Journal* . [Crossref]

378. Scott A. Imberman, Michael F. Lovenheim. 2016. Does the market value value-added? Evidence from housing prices after a public release of school and teacher value-added. *Journal of Urban Economics* **91**, 104-121. [Crossref]

379. Robert J. Stonebraker, Gary S. Stone. 2015. Too Old to Teach? The Effect of Age on College and University Professors. *Research in Higher Education* **56**:8, 793-812. [Crossref]

380. Irina Horoi, Ben Ost. 2015. Disruptive peers and the estimation of teacher value added. *Economics of Education Review* **49**, 180-192. [Crossref]

381. Alain de Crombrugghe, Marc Romainville. 2015. Cours interactif et performance académique d'étudiants de première année universitaire en économie. *Revue internationale de pédagogie de l'enseignement supérieur* **31**:3. . [Crossref]

382. Christopher R. Walters. 2015. Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start. *American Economic Journal: Applied Economics* **7**:4, 76-102. [Abstract] [View PDF article] [PDF with links]

383. Matthew A. Kraft. 2015. Teacher Layoffs, Teacher Quality, and Student Achievement: Evidence from a Discretionary Layoff Policy. *Education Finance and Policy* **10**:4, 467-507. [Crossref]

384. John P. Papay, Matthew A. Kraft. 2015. Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics* **130**, 105-119. [Crossref]

385. Cory Koedel, Kata Mihaly, Jonah E. Rockoff. 2015. Value-added modeling: A review. *Economics of Education Review* **47**, 180-195. [Crossref]

386. David R. Agrawal, William F. Fox, Joel Slemrod. 2015. Competition and Subnational Governments: Tax Competition, Competition in Urban Areas, and Education Competition. *National Tax Journal* **68**:3S, 701-734. [Crossref]

387. David Blazar. 2015. Grade Assignments and the Teacher Pipeline. *Educational Researcher* **44**:4, 213-227. [Crossref]

388. Derek C. Briggs, Frederick A. Peck. 2015. Using Learning Progressions to Design Vertical Scales that Support Coherent Inferences about Student Growth. *Measurement: Interdisciplinary Research and Perspectives* **13**:2, 75-99. [Crossref]

389. Anna J. Egalite, Brian Kisida, Marcus A. Winters. 2015. Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review* **45**, 44-52. [Crossref]

390. Cassandra M. Guarino, Michelle Maxfield, Mark D. Reckase, Paul N. Thompson, Jeffrey M. Wooldridge. 2015. An Evaluation of Empirical Bayes's Estimation of Value-Added Teacher Performance Measures. *Journal of Educational and Behavioral Statistics* **40**:2, 190-222. [Crossref]

391. Abhijeet Singh. 2015. Private school effects in urban and rural India: Panel estimates at primary and secondary school ages. *Journal of Development Economics* **113**, 16-32. [Crossref]

392. Henry Braun. 2015. The Value in Value Added Depends on the Ecology. *Educational Researcher* **44**:2, 127-131. [Crossref]

393. Gema Zamarro, John Engberg, Juan Esteban Saavedra, Jennifer Steele. 2015. Disentangling Disadvantage: Can We Distinguish Good Teaching From Classroom Composition?. *Journal of Research on Educational Effectiveness* **8**:1, 84-111. [Crossref]

394. Jesse Rothstein. 2015. Teacher Quality Policy When Supply Matters. *American Economic Review* **105**:1, 100-130. [Abstract] [View PDF article] [PDF with links]

395. Elias Walsh, Eric Isenberg. 2015. How Does Value Added Compare to Student Growth Percentiles?. *Statistics and Public Policy* **2**:1, 1-13. [Crossref]

396. Seth Gershenson. 2015. Performance Standards and Employee Effort: Evidence from Teacher Absences. *SSRN Electronic Journal* . [Crossref]

397. Matthew D. Hendricks. 2015. Public Schools Are Hemorrhaging Talented Teachers: Can Higher Salaries Function as a Tourniquet?. *SSRN Electronic Journal* . [Crossref]

398. Michael Geruso, Timothy J Layton. 2015. Upcoding: Evidence from Medicare on Squishy Risk Adjustment. *SSRN Electronic Journal* . [Crossref]

399. Craig M. Burnett, Vladimir Kogan. 2015. The Politics of Potholes: Service Quality and Retrospective Voting in Local Elections. *SSRN Electronic Journal* . [Crossref]

400. Jonathan T. Rothwell. 2015. Defining Skilled Technical Work. *SSRN Electronic Journal* . [Crossref]

401. Deven Carlson, Joshua M. Cowen. 2015. Student Neighborhoods, Schools, and Test Score Growth. *Sociology of Education* **88**:1, 38-55. [Crossref]

402. Stéphane Lavertu. 2014. Performance Measurement: A Balanced Solution to Inequities in Election Administration?. *Election Law Journal: Rules, Politics, and Policy* **13**:4, 550-554. [Crossref]

403. Liran Einav, Jonathan Levin. 2014. Economics in the age of big data. *Science* **346**:6210. . [Crossref]

404. Raj Chetty, John N. Friedman, Jonah E. Rockoff. 2014. Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review* **104**:9, 2633-2679. [Abstract] [View PDF article] [PDF with links]

405. Ricardo Perez Truglia. 2014. Conformity Effects and Geographic Polarization: Evidence from an Event-Study Analysis of Residential Mobility in the U.S. *SSRN Electronic Journal* . [Crossref]

406. Paul T. von Hippel, Cynthia Osborne, Jane Arnold Lincove, Nicholas Mills, Laura Bellows. 2014. Seeking Exceptional Teacher Preparation Programs Among Many Noisy Estimates. *SSRN Electronic Journal* . [Crossref]

407. David R. Agrawal, William F. Fox, Joel B. Slemrod. 2014. Competition and Subnational Governments: Tax Competition, Competition in Urban Areas, and Education Competition. *SSRN Electronic Journal* . [Crossref]

408. Peter Hinrichs. 2014. What Kind of Teachers are Schools Looking for? Evidence from a Randomized Field Experiment. *SSRN Electronic Journal* . [Crossref]

409. Seth Gershenson. 2014. Should Value-Added Models Control for Student Absences?. *SSRN Electronic Journal* . [Crossref]

410. Massimo G. Colombo, Samuele Murtinu. 2014. Venture Capital Investments in Europe and Portfolio Firms' Economic Performance: Independent versus Corporate Investors. *SSRN Electronic Journal* . [Crossref]

411. Michela Braga, Marco Paccagnella, MIchele Pellizzari. 2014. The Academic and Labor Market Returns of University Professors. *SSRN Electronic Journal* . [Crossref]