American Economic Review 2014, 104(9): 2633–2679
http://dx.doi.org/10.1257/aer.104.9.2633

Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood†

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? This question has sparked debate partly because of a lack of evidence on whether high value-added (VA) teachers improve students' long-term outcomes. Using school district and tax records for more than one million children, we find that students assigned to high-VA teachers are more likely to attend college, earn higher salaries, and are less likely to have children as teenagers. Replacing a teacher whose VA is in the bottom 5 percent with an average teacher would increase the present value of students' lifetime income by approximately $250,000 per classroom. (*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

* Chetty: Harvard University, Littauer Center 226, Cambridge, MA 02138 (e-mail: chetty@fas.harvard.edu); Friedman: Harvard University, Taubman Center 356, Cambridge, MA 02138 (e-mail: john_friedman@harvard.edu); Rockoff: Columbia University, Uris 603, New York, NY 10027 (e-mail: jonah.rockoff@columbia.edu). We thank Joseph Altonji, Josh Angrist, David Card, Gary Chamberlain, David Deming, Caroline Hoxby, Guido Imbens, Brian Jacob, Thomas Kane, Lawrence Katz, Michal Kolesar, Adam Looney, Phil Oreopoulos, Jesse Rothstein, Douglas Staiger, Danny Yagan, anonymous referees, and numerous seminar participants for helpful discussions and comments. This paper is the second of two companion papers on teacher quality. The results in the two papers were previously combined in NBER Working Paper 17699, entitled "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood," issued in December 2011. On May 4, 2012, Raj Chetty was retained as an expert witness by Gibson, Dunn, and Crutcher LLP to testify about the importance of teacher effectiveness for student learning in *Vergara v. California* based on the findings in NBER Working Paper 17699. John Friedman is currently on leave from Harvard, working at the National Economic Council; this work does not represent the views of the NEC. All results based on tax data contained in this paper were originally reported in an IRS Statistics of Income white paper (Chetty, Friedman, and Rockoff 2011a). Sarah Abraham, Alex Bell, Peter Ganong, Sarah Griffis, Jessica Laird, Shelby Lin, Alex Olssen, Heather Sarsons, and Michael Stepner provided outstanding research assistance. Financial support from the Lab for Economic Applications and Policy at Harvard and the National Science Foundation is gratefully acknowledged. Publicly available portions of the analysis code are posted at: http://obs.rc.fas.harvard.edu/chetty/cfr_analysis_code.zip.

† Go to http://dx.doi.org/10.1257/aer.104.9.2633 to visit the article page for additional materials and author disclosure statement(s).

[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?

We addressed the first question in the previous paper in this volume (Chetty, Friedman, and Rockoff 2014) and concluded that VA measures which control for lagged test scores exhibit little or no bias. This paper addresses the second question.[2] We study the long-term impacts of teachers by linking information from an administrative dataset on students and teachers in grades 3–8 from a large urban school district spanning 1989–2009 with selected data from United States tax records spanning 1996–2011. We match approximately 90 percent of the observations in the school district data to the tax data, allowing us to track approximately one million individuals from elementary school to early adulthood, where we measure outcomes such as earnings, college attendance, and teenage birth rates.

We use two research designs to estimate the long-term impacts of teacher quality: cross-sectional comparisons across classrooms and a quasi-experimental design based on teacher turnover. The first design compares the outcomes of students who were assigned to teachers with different VA, controlling for a rich set of student characteristics such as prior test scores and demographics. We implement this approach by regressing long-term outcomes on the test-score VA estimates constructed in Chetty, Friedman, and Rockoff (2014). The identification assumption underlying this research design is selection on observables: unobserved determinants of outcomes in adulthood such as student ability must be unrelated to teacher VA conditional on the observable characteristics. Although this is a very strong assumption, the estimates from this approach closely match the quasi-experimental estimates for outcomes where we have adequate precision to implement both designs, supporting its validity.

We find that teacher VA has substantial impacts on a broad range of outcomes. A one standard deviation improvement in teacher VA in a single grade raises the probability of college attendance at age 20 by 0.82 percentage points, relative to a sample mean of 37 percent. Improvements in teacher quality also raise the quality of the colleges which students attend, as measured by the average earnings of previous graduates of that college. Students who are assigned higher VA teachers have steeper earnings trajectories in their 20s. At age 28, the oldest age at which we currently have a sufficiently large sample size to estimate earnings impacts, a one standard deviation increase in teacher quality in a single grade raises annual earnings by 1.3 percent. If the impact on earnings remains constant at 1.3 percent over the life cycle, students would gain approximately \$39,000 on average in cumulative lifetime income from a one standard deviation improvement in teacher VA in a single grade. Discounting at a 5 percent rate yields a present value gain of \$7,000 at age 12, the mean age at which the interventions we study occur. We also find that improvements in teacher quality significantly reduce the probability of having a child while being a teenager, increase the quality of the neighborhood in which the student lives (as measured by the percentage of college graduates in that zip code) in adulthood, and raise participation rates in 401(k) retirement savings plans.

[2] Recent work has shown that early childhood education has significant long-term impacts (e.g., Heckman et al. 2010a, 2010b, Chetty et al. 2011), but there is no evidence to date on the long-term impacts of teacher quality as measured by value-added.

Our second design relaxes the assumption of selection on observables by exploiting teacher turnover as a quasi-experimental source of variation in teacher quality. To understand this research design, suppose a high-VA fourth grade teacher moves from school *A* to another school in 1995. Because of this staff change, students entering grade 4 in school *A* in 1995 will have lower quality teachers on average than those in the prior cohort. If high-VA teachers improve long-term outcomes, we would expect college attendance rates and earnings for the 1995 cohort to be lower on average than the previous cohort. Building on this idea, we estimate teachers' impacts by regressing changes in mean adult outcomes across consecutive cohorts of children within a school on changes in the mean VA of the teaching staff.

Using this design, we find that a one standard deviation improvement in teacher VA raises the probability of college attendance at age 20 by 0.86 percentage points, nearly identical to the estimate obtained from our first research design. Improvements in average teacher VA also increase the quality of colleges which students attend. The impacts on college outcomes are statistically significant with $p < 0.01$. Unfortunately, we have insufficient precision to obtain informative estimates for earnings using the quasi-experimental design.

Our quasi-experimental results rest on the identification assumption that high-frequency teacher turnover within school-grade cells is uncorrelated with student and school characteristics. Several pieces of evidence support this assumption. First, predetermined student and parent characteristics are uncorrelated with changes in the quality of teaching staff. Second, students' prior test scores and contemporaneous scores in the other subject are also uncorrelated with changes in the quality of teaching staff in a given subject. Third, changes in teacher VA across cohorts have sharp effects on college attendance exactly in the year of the change but not in prior years or subsequent years.

The long-term impacts of teacher VA are slightly larger for females than males. Improvements in English teacher quality have larger impacts than improvements in math teacher quality. The impacts of teacher VA are roughly constant in percentage terms by parents' income. Hence, higher income households, whose children have higher earnings on average, should be willing to pay larger amounts for higher teacher VA. Teachers' impacts are significant and large throughout grades 4–8, showing that improvements in the quality of education can have large returns well beyond early childhood.

Our conclusion that teachers have long-lasting impacts may be surprising given evidence that teachers' impacts on test scores *fade out* very rapidly in subsequent grades (Rothstein 2010; Carrell and West 2010; Jacob, Lefgren, and Sims 2010). We confirm this rapid fade-out in our data, but find that teachers' impacts on earnings are similar to what one would predict based on the cross-sectional correlation between earnings and contemporaneous test score gains. This pattern of fade-out and re-emergence echoes the findings of recent studies of early childhood interventions (Deming 2009; Heckman et al. 2010b; Chetty et al. 2011).

To illustrate the magnitude of teachers' impacts, we evaluate Hanushek's (2009) proposal to replace teachers in the bottom 5 percent of the VA distribution with teachers of average quality. We estimate that replacing a teacher whose current VA is in the bottom 5 percent with an average teacher would increase the mean present value of students' lifetime income by $250,000 per classroom over a teacher's career, accounting for drift in teacher quality over time.[3] However, because VA is estimated with noise, the gains from deselecting teachers based on data from a limited number of classrooms are smaller. The present value gain from deselecting the bottom 5 percent of teachers using three years of test score data is \$185,000 per classroom on average.[4] This gain is still about ten times larger than recent estimates of the additional salary one would have to pay teachers to compensate them for the risk of evaluation based on VA measures (Rothstein 2013). This result suggests that VA could potentially be a useful tool for evaluating teacher performance if the signal quality of VA for long-term impacts does not fall substantially when it is used to evaluate teachers.

We also evaluate the expected gains from policies which pay bonuses to high-VA teachers to increase retention rates. The gains from such policies are only slightly larger than their costs because most bonus payments end up going to high-VA teachers who would have stayed even without the additional payment. Replacing low-VA teachers may therefore be a more cost effective strategy to increase teacher quality in the short run than paying bonuses to retain high-VA teachers. In the long run, higher salaries could attract more high-VA teachers to the teaching profession, a potentially important benefit which we do not measure here.

The paper is organized as follows. In Section I, we formalize our estimating equations and explain how the parameters we estimate should be interpreted using a simple statistical model. Section II describes the data sources and reports summary statistics as well as cross-sectional correlations between test scores and adult outcomes. Sections III and IV present results on teachers' long-term impacts using the two research designs described above. We analyze the heterogeneity of teachers' impacts in Section V. Section VI presents policy simulations and Section VII concludes.

I. Conceptual Framework

In this section, we first present a simple statistical model of students' long-term outcomes as a function of their teachers' value-added. We then describe how the reduced-form parameters of this statistical model should be interpreted. Finally, we show how we estimate the impacts of teacher VA on long-term outcomes given that each teacher's true value-added is unobserved.

Statistical Model.—Consider the outcomes of a student i who is in grade g in calendar year $t_i(g)$. Let $j = j(i,t)$ denote student i's teacher in school year t; for simplicity, assume that the student has only one teacher throughout the school year, as in elementary schools. Let μ_{jt} represent teacher j's "test-score value-added" in year t, so that student i's test score in year t is

$$A_{it}^{*} = \boldsymbol{\beta}\mathbf{X}_{it} + \mu_{jt} + \varepsilon_{it}. \tag{1}$$

Here, $\mathbf{X}_{it}$ denotes observable determinants of student achievement, such as lagged

[3] This calculation discounts the earnings gains at a rate of 5 percent to age 12. The estimated total undiscounted earnings gains from this policy are approximately \$50,000 per child and \$1.4 million for the average classroom.

[4] The gains remain substantial despite noise because very few of the teachers with low VA estimates ultimately turn out to be excellent teachers. For example, we estimate that 3.2 percent of the math teachers in elementary school whose estimated VA is in the bottom 5 percent based on three years of data actually have true VA above the median.

test scores and family characteristics and ε_{it} denotes a student-level error which may be correlated across students within a classroom and with teacher value-added μ_{jt}. We scale μ_{jt} in student test-score standard deviations so that the average teacher has $\mu_{jt} = 0$ and the effect of a 1 unit increase in teacher value-added on end-of-year test scores is 1. We allow teacher quality μ_{jt} to vary with time t to account for the stochastic drift in teacher quality documented in our companion paper (Chetty, Friedman, and Rockoff 2014).

Let Y_i^* denote student i's earnings in adulthood. Throughout our analysis, we focus on earnings residuals after removing the effect of observable characteristics:

$$(2) \qquad Y_{it} = Y_i^* - \boldsymbol{\beta}^Y \mathbf{X}_{it}.$$

The earnings residuals Y_{it} vary across school years because the control vector $\mathbf{X}_{it}$ varies across school years. We estimate the coefficient vector $\boldsymbol{\beta}^Y$ using variation across students taught by the same teacher using an OLS regression:

$$(3) \qquad Y_i^* = \alpha_j + \boldsymbol{\beta}^Y \mathbf{X}_{it},$$

where α_j is a teacher fixed effect. Importantly, we estimate $\boldsymbol{\beta}^Y$ using *within-teacher* variation to account for the potential sorting of students to teachers based on VA. If teacher VA is correlated with $\mathbf{X}_{it}$, estimates of $\boldsymbol{\beta}^Y$ in a specification without teacher fixed effects are biased because part of the teacher effect is attributed to the covariates. See Section IB of our companion paper for further discussion of this issue.

We model the relationship between earnings residuals and teacher VA in school year t using the following linear specification:

$$(4) \qquad Y_{it} = a + \kappa_g m_{jt} + \eta_{it},$$

where $m_{jt} = \mu_{jt}/\sigma_\mu$ denotes teacher j's *normalized value-added* (i.e., teacher quality scaled in standard deviation (σ_μ) units of the teacher VA distribution).

Interpretation of Reduced-Form Treatment Effects.—The parameter κ_g in (4) represents the reduced-form impact of a one standard deviation increase in teachers' test-score VA in a given school year t on earnings. There are two important issues to keep in mind when interpreting this parameter, which we formalize using a dynamic model of education production in online Appendix A.

First, the reduced-form impact combines two effects: the direct impact of having a higher-VA teacher in grade g and the indirect impact of endogenous changes in other educational inputs. For example, other determinants of earnings such as investments in learning by children and their parents might respond endogenously to changes in teacher quality. One particularly important endogenous response is that a higher achieving student may be tracked into classes taught by higher-quality teachers. Such tracking would lead us to overstate the impacts of improving teacher quality in grade g holding fixed the quality of teachers in subsequent grades. In Section VC, we estimate the degree of teacher tracking and use these estimates to identify the impact of having a higher-VA teacher in each grade holding fixed future teacher quality, which we denote by $\tilde{\kappa}_g$. The degree of tracking turns out to be relatively small in our data, and thus the reduced-form estimates of κ_g reported below are similar to the net impacts of each teacher $\tilde{\kappa}_g$.[5] Although the net impacts $\tilde{\kappa}_g$ still combine several structural parameters—such as endogenous responses by parents and children to changes in teacher quality—they are relevant for policy. For example, the ultimate earnings impact of retaining teachers on the basis of their VA depends on $\tilde{\kappa}_g$.

Second, $\tilde{\kappa}_g$ measures only the portion of teachers' earnings impacts which are correlated with their impacts on test scores. As a result, $\tilde{\kappa}_g$ is a lower bound for the standard deviation of teachers' effects on earnings. Intuitively, some teachers may be effective at raising students' earnings even if they are not effective at raising test scores, for instance by directly instilling other skills which have long-term payoffs (Jackson 2013). In principle, one could estimate teacher *j*'s *earnings value-added* $\left(\mu_{jt}^{Y}\right)$ based on the mean residual earnings of her students, exactly as we estimated test-score VA in our first paper. Unfortunately, the orthogonality condition required to obtain unbiased forecasts of teachers' earnings VA—that other unobservable determinants of students' earnings are orthogonal to earnings VA estimates—does not hold in practice, as we discuss in online Appendix A. We therefore focus on estimating the effect of being assigned to a high test-score VA teacher on earnings (κ_g). Although κ_g does not correspond directly to earnings VA, it reveals the extent to which the test-score-based VA measures currently used by school districts are informative about teachers' long-term impacts.

Empirical Implementation.—There are two challenges in estimating κ_g using (4). First, unobserved determinants of earnings η_{it} may be correlated with teacher VA m_{jt}. We return to this issue in our empirical analysis and isolate variation in teacher VA which is orthogonal to unobserved determinants of earnings. Second, teachers' true test-score VA m_{jt} is unobserved. We can solve this second problem by substituting estimates of teacher VA $\hat{m}_{jt} = \hat{\mu}_{jt}/\sigma_\mu$ for true teacher VA in (4) under the following assumption.

ASSUMPTION 1 (Forecast Unbiasedness of Test-Score VA): *Test-score value-added estimates $\hat{\mu}_{jt}$ are forecast unbiased:*

$$\frac{\mathrm{Cov}(\mu_{jt}, \hat{\mu}_{jt})}{\mathrm{Var}(\hat{\mu}_{jt})} = \frac{\mathrm{Cov}(m_{jt}, \hat{m}_{jt})}{\mathrm{Var}(\hat{m}_{jt})} = 1.$$

[5] An alternative approach to identify the direct impacts of teachers in each grade *g* would be to include teacher VA in all grades simultaneously in the model in (4). Unfortunately, this is not feasible because our primary research design requires conditioning on lagged test scores and lagged scores are endogenous to teacher quality in the previous grade. For the same reason, we also cannot identify the substitutability or complementarity of teachers' impacts across grades. See online Appendix A for further details.

In our companion paper, we demonstrate that Assumption 1 holds for the VA estimates which we use in this paper. Equation (4) implies that $\text{Cov}(Y_{it}, \hat{m}_{jt}) = \kappa_g \text{Cov}(m_{jt}, \hat{m}_{jt}) + \text{Cov}(\eta_{it}, \hat{m}_{jt})$. Hence, under Assumption 1,

$$\frac{\text{Cov}(Y_{it}, \hat{m}_{jt})}{\text{Var}(\hat{m}_{jt})} = \kappa_g + \frac{\text{Cov}(\eta_{it}, \hat{m}_{jt})}{\text{Var}(\hat{m}_{jt})}.$$

It follows that we can identify the impact of a one standard deviation increase in a teacher's *true* VA m_{jt} from an OLS regression of earnings residuals Y_{it} on teacher VA *estimates* $\hat{m}_{jt}$,

$$(5) \qquad Y_{it} = \alpha + \kappa_g \hat{m}_{jt} + \eta'_{it},$$

provided that unobserved determinants of earnings are orthogonal to teacher VA estimates $\hat{m}_{jt}$.

Intuitively, we identify κ_g using two-stage least squares (2SLS) by instrumenting for true teacher VA m_{jt} with the teacher VA estimates we constructed in our companion paper. Forecast unbiasedness of test-score VA implies that the first stage of this 2SLS regression has a coefficient of 1. Thus, the reduced form coefficient obtained from an ordinary least squares (OLS) regression of earnings on VA estimates identifies κ_g.

The remainder of the paper focuses on estimating variants of (5).[6]

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. We then provide descriptive statistics and report correlations between test scores and long-term outcomes as a benchmark to interpret the magnitude of the causal effects of teachers. Note that the dataset we use in this paper is identical to that used in our first paper, except that we restrict attention to the subset of students who are old enough for us to observe outcomes in adulthood by 2011.

A. *School District Data*

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

[6] Another way to identify κ_g is to directly estimate the covariance of teachers' effects on earnings $\left(\mu_{jt}^{Y}\right)$ and test scores (μ_{jt}) in a correlated random effects or factor model. Chamberlain (2013) develops such an approach and obtains estimates similar to those reported here.

We summarize the key features of the data relevant for our analysis of teachers' long-term impacts here; see Section II of our first paper for a comprehensive description of the school district data.

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. 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 examine is comparable to the within-grade variation nationwide, so our results can be compared to estimates from other samples.

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.

Teachers.—The dataset links students in grades 3–8 to classrooms and teachers from 1991 through 2009. 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- 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 school year 1998 and 85 percent by 2003. These missing teacher links raise two potential concerns. First, our estimates (especially for grades 6–8) apply to a subset of schools with more complete information reporting systems and thus may not be representative of the district as a whole. Reassuringly, we find that these schools do not differ significantly from the sample as a whole on test scores and other observables. Second, and more importantly, missing data could generate biased estimates. We address this concern by showing that our estimates remain similar in a subsample of school-grade-subject cells with little or no missing data (online Appendix Table 7).

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 listed as 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). Fourth, 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. Finally, because the adult outcomes we analyze are measured at age 20 or afterward, in this paper we restrict the sample to students who would have graduated high school by the 2008–2009 school year (and thus turned 20 by 2011) if they progressed through school at a normal pace.[7]

B. *Tax Data*

We obtain information on students' outcomes in adulthood from US federal income tax returns spanning 1996–2011.[8] 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 of our companion paper, after which individual identifiers were removed to protect confidentiality. In all, 87.4 percent of the students and 89.2 percent of student-subject-year observations in the sample used to analyze long-term impacts were matched to the tax data.[9] We define students' outcomes in adulthood as follows.

Earnings.—Individual wage earnings data come from W-2 forms, which are available from 1999 to 2011. Importantly, W-2 data are available for *both* tax filers and non-filers, eliminating concerns about missing data on formal sector earnings. We cap earnings in each year at \$100,000 to reduce the influence of outliers; 1.3 percent of individuals in the sample report earnings above \$100,000 at age 28. We measure all monetary variables in 2010 US\$, adjusting for inflation using the Consumer Price Index. Individuals with no W-2 are coded as having zero earnings; 33.1 percent of individuals have zero wage earnings at age 28 in our sample.

Total Income.—To obtain a more comprehensive definition of income, we define *total income* as the sum of W-2 wage earnings and household self-employment earnings (as reported on the 1040). For non-filers, we define total income as just W-2 wage earnings; those with no W-2 income are coded as having zero total income. In our sample, 29.6 percent of individuals have zero total income.[10] We show that similar results are obtained using this alternative definition of income, but use W-2

[7] A few classrooms contain students at different grade levels because of retentions or split-level classroom structures. To avoid dropping a subset of students within a classroom, we include every classroom which has at least one student who would graduate school during or before 2008–2009 if she progressed at the normal pace. That is, we include all classrooms in which $\min_i(12 + \text{school year} - \text{grade}_i) \leq 2009$.

[8] 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 $t + 1$.

[9] We find little or no correlation between match rates and teacher VA in the various subsamples we use in our analysis and obtain very similar estimates of teachers' impacts on long-term outcomes when restricting the sample to school-grade-subject cells with above-median match rates (see online Appendix Table 7).

[10] According to the Current Population Survey, 27.2 percent of the noninstitutionalized population between the ages of 25 and 29 was not employed in 2011. The nonemployment rate in our sample may differ from this figure because it includes the institutionalized population in the denominator and applies to a relatively low-income urban public school district.

wage earnings as our baseline measure because it (i) is unaffected by the endogeneity of tax filing and (ii) provides a consistent definition of individual (rather than household) income for both filers and non-filers.

College Attendance.—We define college attendance as an indicator for having one or more 1098-T forms filed on one's behalf. Title IV institutions—all colleges and universities as well as vocational schools and other postsecondary institutions eligible for federal student aid—are required to file 1098-T forms which report tuition payments or scholarships received for every student. Because the 1098-T forms are filed directly by colleges independent of whether an individual files a tax return, we have complete records on college attendance for all individuals. The 1098-T data are available from 1999 to 2011. Comparisons to other data sources indicate that 1098-T forms capture college enrollment accurately (see online Appendix B).

College Quality.—We construct an earnings-based index of college quality based on data from the universe of tax returns (not just the students from our school district). Using the population of all current US citizens born in 1979 or 1980, we group individuals by the higher education institution they attended at age 20. We pool individuals who were not enrolled in any college at age 20 together in a separate "no college" category. For each college or university (including the "no college" group), we then compute the mean W-2 earnings of the students when they are age 31 (in 2010 and 2011). Among colleges attended by students in the school district studied in this paper, the average value of our earnings index is $44,048 for four-year colleges and $30,946 for two-year colleges. For students who did not attend college, the mean earnings level is $17,920.

In online Appendix B, we analyze the robustness of the college quality index to alternative specifications, such as measuring earnings and college attendance at different ages and defining the index based on total income instead of W-2 earnings. We find that rankings of college quality are very stable across cohorts and are robust to alternative specifications provided that earnings are measured after age 28 (online Appendix Figure 1, online Appendix Table 2).

Neighborhood Quality.—We use data from 1040 forms to identify each household's zip code of residence in each year. For non-filers, we use the zip code of the address to which the W-2 form was mailed. If an individual did not file and has no W-2 in a given year, we impute current zip code as the last observed zip code. We construct a measure of a neighborhood's socioeconomic status (SES) using data on the percentage of college graduates in the individual's zip code from the 2000 census.

Retirement Savings.—We measure retirement savings using contributions to 401(k) accounts reported on W-2 forms from 1999 to 2011. We define saving for retirement as an indicator for contributing to a 401(k) at age 28.

Teenage Birth.—We define a woman as having a teenage birth if she ever claims a dependent who was born while she was between the ages of 13 and 19 (as of December 31 in the year the child was born). This measure is an imperfect proxy for having a teenage birth because it only covers children who are claimed as dependents by their mothers and includes any other dependents who are not biological children but were born while the mother was a teenager. Despite these limitations, our proxy for teenage birth is closely aligned with estimates based on the American Community Survey (ACS). In our core sample, 15.8 percent of women have teenage births, compared with 14.6 percent in the 2003 ACS. The unweighted correlation between state-level teenage birth rates in our data and the ACS is 0.80.

Parent Characteristics.—We also use the tax data to obtain information on five parent characteristics which we use as controls. Students were linked to parents based on the earliest 1040 form filed between tax years 1996 and 2011 on which the student was claimed as a dependent. We identify parents for 94.8 percent of the observations in the analysis dataset conditional on being matched to the tax data.[11]

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.[12] Parents are assigned an income of zero in years when they did not file a tax return. We define parental 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. Lastly, 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.[13] 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 which include parent characteristics.

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, and class and teacher assignment if available. Each row also includes all the students' available adult outcomes (e.g., college attendance and earnings at each age). We organize the data in this format so that each row contains information on a treatment by a single teacher conditional on predetermined characteristics, facilitating the estimation of (5). We account for the

[11] The remaining students are likely to have parents who did not file tax returns in the early years of the sample when they could have claimed their child as a dependent, making it impossible to link the children to their parents. Note that this definition of parents is based on who claims the child as a dependent, and thus may not reflect the biological parent of the child.

[12] Because the children in our sample vary in age by over 25 years whereas the tax data start only in 1996, we cannot measure parent characteristics at the same age for all children. For simplicity, we instead measure parent characteristics at a fixed time. Measuring parent income at other points in time yields very similar results (not reported).

[13] 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.

fact that each student appears multiple times in the dataset by clustering standard errors as described in Section IIIA.

After imposing the sample restrictions described above, the linked analysis sample contains 6.8 million student-subject-year observations (covering 1.1 million students) which we use to study teachers' long-term impacts.[14] Table 1 reports summary statistics for this sample. Note that the summary statistics are student-subject-year means and thus weight students who are in the district for a longer period of time more heavily, as does our empirical analysis. On average, each student has 6.25 subject-school year observations.

The mean test score in the analysis sample is positive and has a standard deviation below one because we normalize the test scores in the full population which includes students in special education classrooms and schools (who typically have lower test scores). The mean age at which students are observed in school is 11.7 years. In addition, 77.1 percent of students are eligible for free or reduced price lunches.

The availability of data on adult outcomes naturally varies across cohorts. There are more than 5.9 million observations for which we observe college attendance at age 20. We observe earnings at age 25 for 2.3 million observations and at age 28 for 1.3 million observations. Because many of these observations at later ages are from older cohorts of students who were in middle school in the early 1990s, we were not able to obtain information on teachers. As a result, there are only 1.6 million student-subject-school year observations for which we see *both* teacher VA and earnings at age 25, 750,000 at age 28, and only 220,000 at age 30. The oldest age at which the sample is large enough to obtain informative estimates of teachers' impacts on earnings turns out to be age 28. Mean individual earnings at age 28 is \$20,885, while mean total income is \$21,272 (in 2010 US\$).

For students whom we are able to link to parents, the mean parent household income is \$40,808, while the median is \$31,834. 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 \$34,515, with 10 percent of parents earning more than \$100,000.

D. *Cross-Sectional Correlations*

Online Appendix Tables 3–6 report coefficients from OLS regressions of various adult outcomes on test scores. Both math and English test scores are highly positively correlated with earnings, college attendance, and neighborhood quality, and are negatively correlated with teenage births. In the cross section, a one standard deviation increase in test score is associated with a \$7,700 (36 percent) increase in earnings at age 28. Conditional on the student- and class-level controls $\mathbf{X}_{it}$ which we define in Section IIIA below, a one standard deviation increase in the current test score is associated with a \$2,600 (12 percent) increase in earnings on average.

[14] For much of the analysis in our first paper, we restricted attention to the subset of observations in the core sample which have lagged scores and other controls needed to estimate the baseline VA model. Because we do not control for individual-level variables in most of the specifications in this paper, we do not impose that restriction here.

TABLE 1—SUMMARY STATISTICS FOR LINKED ANALYSIS DATASET

Variable	Mean (1)	SD (2)	Observations (3)
Student data:			
Class size (not student-weighted)	28.2	5.8	240,459
Number of subject-school years per student	6.25	3.18	1,083,556
Test score (SD)	0.12	0.91	6,035,726
Female	50.4%		6,762,896
Age (years)	11.7	1.6	6,762,679
Free lunch eligible (1999–2009)	77.1%		3,309,198
Minority (Black or Hispanic)	72.1%		6,756,138
English language learner	4.9%		6,734,837
Special education	3.1%		6,586,925
Repeating grade	2.7%		6,432,281
Matched to tax data	89.2%		6,770,045
Matched to parents (cond. on match to tax data)	94.8%		6,036,422
Adult outcomes:			
Annual wage earnings at age 20	5,670	7,733	5,939,022
Annual wage earnings at age 25	17,194	19,889	2,321,337
Annual wage earnings at age 28	20,885	24,297	1,312,800
Total income at age 28	21,780	24,281	1,312,800
In college at age 20	35.6%		5,939,022
In college at age 25	16.5%		2,321,337
More than four years of college, ages 18–22	22.7%		4,514,758
College quality at age 20	26,408	13,461	5,934,570
Contributed to a 401(k) at age 28	19.1%		1,312,800
Percent college graduates in zip at age 28	13.7%		929,079
Had a child while a teenager (for women)	14.3%		3,032,170
Owned a house at age 25	4.3%		2,321,337
Married at age 25	11.3%		2,321,337
Parent characteristics:			
Annual household income	40,808	34,515	5,720,657
Owned a house	34.8%		5,720,657
Contributed to a 401(k)	31.3%		5,720,657
Married	42.2%		5,720,657
Age at child birth	28.3	7.8	5,615,400

Notes: All statistics reported are for the linked analysis dataset described in Section II, which includes students from classrooms in which at least one student would graduate high school in or before 2009 if progressing at a normal pace. The sample has one observation per student-subject-school year. Student data are from the administrative records of a large urban school district in the US. Adult outcomes and parent characteristics are from 1996–2011 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 refers to standardized scale score in math or English. 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. We are unable to link 10.8 percent of observations to the tax data; the summary statistics for adult outcomes and parent characteristics exclude these observations. Wage earnings are measured from W-2 forms; we assign 0's to students with no W-2s. Total income includes both W-2 wage earnings and self-employment income reported on the 1040. College attendance is measured from 1098-T forms. College quality is the average W-2 earnings at age 31 for students who attended a given college at age 20 (see Section IIB for more details). 401(k) contributions are reported on W-2 forms. Zip code of residence is determined from the address on a 1040 (or W-2 for non-filers); percent college graduates in zip is based on the 2000 Census. We measure teen births for female students as an indicator for claiming a dependent who was born fewer than 20 years after the student herself was born. We measure home ownership from the payment of mortgage interest, reported on either the 1040 or a 1099 form. We measure marriage by the filing of a joint return. Conditional on linking to the tax data, we are unable to link 5.2 percent of observations to a parent; the summary statistics for parents exclude these observations. Parent income is average adjusted gross income during the three tax-years between 2005 and 2007. For parents who do not file, household income is defined as zero. Parent age at child birth is the difference between the age of the mother (or father if single father) and the student. All parent indicator variables are defined in the same way as the equivalent for the students and are equal to 1 if the event occurs in any year between 2005 and 2007.

Online Appendix Figure 2 presents binned scatter plots of selected outcomes versus test scores both with and without controls. The unconditional relationship between scores and outcomes is S-shaped, while the relationship conditional on prior scores and other covariates is almost perfectly linear. We return to these results below and show that the causal impacts of teacher VA on earnings and other outcomes are commensurate to what one would predict based on these correlations.

III. Research Design 1: Cross-Class Comparisons

Our first method of estimating teachers' long-term impacts builds on our finding that conditioning on prior test scores and other observables is adequate to obtain unbiased estimates of teachers' causal impacts on test scores (Chetty, Friedman, and Rockoff 2014). Given this result, one may expect that comparing the long-term outcomes of students assigned to different teachers conditional on the same control vector will yield unbiased estimates of teachers' long-term impacts. The next subsection formalizes the identification assumption and methodological details of this approach. We then present results for three sets of impacts: college attendance, earnings, and other outcomes such as teenage birth.

A. *Methodology*

We begin by constructing earnings (or other long-term outcome) residuals $Y_{it} = Y_i^* - \hat{\boldsymbol{\beta}}^Y \mathbf{X}_{it}$, estimating $\hat{\boldsymbol{\beta}}^Y$ using within-teacher variation as in (3). We then regress students' earnings residuals on their teachers' normalized VA $\hat{m}_{jt}$, pooling all grades and subjects:

$$Y_{it} = \alpha + \kappa \hat{m}_{jt} + \eta'_{it}. \tag{6}$$

Note that we do not residualize $\hat{m}_{jt}$ with respect to the controls $\mathbf{X}_{it}$ when estimating (6). In a partial regression, one residualizes $\hat{m}_{jt}$ typically with respect to $\mathbf{X}_{it}$ because the OLS regression of Y_i^* on $\mathbf{X}_{it}$ used to construct earnings residuals yields an estimate of $\boldsymbol{\beta}^Y$ which is biased by the correlation between and $\hat{m}_{jt}$ and $\mathbf{X}_{it}$. This problem does not arise here because we estimate $\boldsymbol{\beta}^Y$ from a regression with teacher fixed effects, so the variation in $\mathbf{X}_{it}$ used to identify $\boldsymbol{\beta}^Y$ is orthogonal to the variation in VA across teachers.[15] Hence, regressing Y_{it} directly on $\hat{m}_{jt}$ identifies the relationship between Y_{it} and $\hat{m}_{jt}$ conditional on $\mathbf{X}_{it}$.

Recall that each student appears in our dataset once for every subject-year with the same level of Y_{it} but different values of $\hat{m}_{jt}$. Hence, κ represents the mean impact of having a higher VA teacher for a *single* grade between grades 4–8. We present results on heterogeneity in impacts across grades and subgroups in Section V.

[15] 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, one may still obtain biased estimates of $\boldsymbol{\beta}^Y$. This problem is modest because only 20 percent of the variance in $\hat{m}_{jt}$ is within teacher. Moreover, we obtain similar results when estimating $\hat{m}_{jt}$ and $\boldsymbol{\beta}^Y$ from regressions without teacher fixed effects and implementing a standard partial regression (Chetty, Friedman, and Rockoff 2011b). We estimate $\boldsymbol{\beta}^Y$ using within-teacher variation here for consistency with our approach to estimating teacher VA in the first paper; see Section IB of that paper for further details.

Estimating (6) using OLS yields an unbiased estimate of κ under the following assumption.

ASSUMPTION 2 (Selection on Observables): *Test-score value-added estimates are orthogonal to unobserved determinants of earnings conditional on* $\mathbf{X}_{it}$*:*

$$\text{Cov}(\hat{m}_{jt}, \eta'_{it}) = 0. \tag{7}$$

While we cannot be certain that conditioning on observables fully accounts for differences in student characteristics across teachers, as required by Assumption 2, the quasi-experimental evidence reported in the next section supports this assumption. In particular, quasi-experimental estimates of the impacts of teacher quality on test scores and college attendance are very similar to the estimates obtained from (6).[16] This result supports the validity of the selection on observables assumption not only for test scores and college attendance but also for other outcomes as well, as any selection effects would presumably be manifested in all of these outcomes.

Four methodological issues arise in estimating (6): (i) estimating test-score VA $\hat{m}_{jt}$; (ii) specifying a control vector $\mathbf{X}_{it}$; (iii) calculating the standard error on κ; and (iv) accounting for outliers in $\hat{m}_{jt}$. The remainder of this subsection addresses these four issues. Note that our methodology parallels closely that in our companion paper. In particular, we use the same VA estimates and control vectors and calculate standard errors in the same way. We did not address outliers in our first paper because they only affect our analysis of long-term impacts, as we explain below.

Estimating Test-Score VA.—We define normalized VA $\hat{m}_{jt} = \hat{\mu}_{jt}/\sigma_\mu$, where $\hat{\mu}_{jt}$ is the baseline estimate of test-score VA for teacher j in year t constructed in our companion paper.[17] We define σ_μ as the standard deviation of teacher effects for the corresponding subject and school-level using the estimates in Table 2 of our companion paper: 0.163 for math and 0.124 for English in elementary school and 0.134 for math and 0.098 for English in middle school. With this scaling, a one unit increase in $\hat{m}_{jt}$ corresponds to a teacher who is rated one standard deviation higher in the distribution of true teacher quality for her subject and school-level. Note that because $\hat{\mu}_{jt}$ is shrunk toward the sample mean to account for noise in VA estimates, $SD(\hat{\mu}_{jt}) < \sigma_\mu$ and hence the standard deviation of the normalized VA measure $\hat{m}_{jt}$ is less than one. We demean $\hat{m}_{jt}$ within each of the four subject (math versus English) by school level (elementary versus middle) cells in the estimation sample in (6) so that κ is identified purely from variation within the subject-by-school-level cells.

Importantly, the VA estimates $\hat{m}_{jt}$ are predictions of teacher quality in year t based on test score data from all years excluding year t. For example, when predicting teachers' effects on the outcomes of students they taught in 1995, we estimate $\hat{m}_{j,1995}$ based on residual test scores from students in all years of the sample *except* 1995. To maximize precision, the VA estimates are based on data from all years for which

[16] The test score results are reported in our companion paper (Chetty, Friedman, and Rockoff 2014). We do not have adequate precision to implement the quasi-experimental design for earnings or within specific subgroups, which is why the cross-sectional estimates remain valuable.

[17] Unless otherwise specified, the independent variable in all the regressions and figures in this paper is normalized test-score VA $\hat{m}_{jt}$. For simplicity, we refer to this measure as *value-added* or VA below.

school district data with teacher assignments are available (1991–2009), not just the subset of older cohorts for which we observe long-term outcomes.

Using a leave-year-out estimate of VA is necessary to obtain unbiased estimates of teachers' long-term impacts because of correlated errors in students' test scores and later outcomes. Intuitively, if a teacher is randomly assigned unobservably high ability students, her estimated VA will be higher. The same unobservably high ability students are likely to have high levels of earnings η'_{it}, generating a mechanical correlation between VA and earnings even if teachers have no causal effect ($\kappa = 0$). The leave-year-out approach eliminates this correlated estimation error bias because $\hat{m}_{jt}$ is estimated using a sample which excludes the observations on the left-hand side of (6).[18]

Control Vectors.—We construct residuals Y_{it} using separate models for each of the four subject-by-school-level cells. Within each of these groups, we regress raw outcomes Y_i^* on a vector of covariates $\mathbf{X}_{it}$ with teacher fixed effects, as in (3), and compute residuals Y_{it}. We partition the control vector $\mathbf{X}_{it}$ which we used to construct our baseline VA estimates into two components: student-level controls $\mathbf{X}_{it}^I$ that vary across students within a class; and classroom-level controls $\mathbf{X}_{ct}$ that vary only at the classroom level. The student-level control vector $\mathbf{X}_{it}^I$ includes cubic polynomials in prior-year math and English scores, interacted with the student's grade level to permit flexibility in the persistence of test scores as students age. We also control for the following student level characteristics: ethnicity, gender, age, lagged suspensions and absences, and indicators for grade repetition, free or reduced-price lunch, special education, and limited English. The class-level controls $\mathbf{X}_{ct}$ consist of the following elements: (i) class size and class-type indicators (honors, remedial); (ii) 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; (iii) class and school-year means of all the individual covariates $\mathbf{X}_{it}^I$; and (iv) grade and year dummies.

In our baseline analysis, we control only for the class-level controls $\mathbf{X}_{ct}$ when estimating the residuals Y_{it}. Let $Y_{ct} = Y_c^* - \boldsymbol{\beta}_C \mathbf{X}_{ct}$ denote the residual of mean outcomes Y_c^* in class c in year t, where $\boldsymbol{\beta}_C$ is estimated at the class level using within-teacher variation across classrooms as in (3), weighting by class size. We estimate the impact of teacher VA on mean outcomes using a class-level OLS regression analogous to (6), again weighting by class size:

$$Y_{ct} = \alpha + \kappa_C \hat{m}_{jt} + \eta'_{ct}. \tag{8}$$

The point estimate $\hat{\kappa}_C$ in (8) is identical to $\hat{\kappa}$ in (6) because teacher VA varies only at the classroom level. Formally, $\hat{\kappa}_C = \hat{\kappa}$ because deviations of individual-level controls $(\mathbf{X}_{it}^I - \mathbf{X}_{ct})$ and outcomes $(Y_i^* - Y_{ct}^*)$ from class means are uncorrelated with $\hat{m}_{jt}$.[19] Omitting individual-level controls allows us to implement our analysis on a dataset collapsed to classroom means, reducing computational costs.

[18] This problem does not arise when estimating the impacts of treatments such as class size because the treatment is observed; here, the size of the treatment (teacher VA) must itself be estimated, leading to correlated estimation errors.

[19] In practice, the identity $\hat{\kappa} = \hat{\kappa}_C$ does not hold exactly because the class means $\mathbf{X}_{ct}$ are defined using all observations with non-missing data for the relevant variable. Some students are not matched to the tax data and hence are missing Y_i^*, while other students are missing some of the individual-level covariates $\mathbf{X}_{it}^I$ (e.g., prior-year test

Standard Errors.—The dependent variable in (6) has a correlated error structure because students within a classroom face common class-level shocks and because our analysis dataset contains repeat observations on students in different grades. One natural way to account for these two sources of correlated errors would be to cluster standard errors by both student and classroom (Cameron, Gelbach, and Miller 2011). Unfortunately, two-way clustering of this form requires running regressions on student-level data and thus was computationally infeasible at the Internal Revenue Service. We instead cluster standard errors at the school by cohort level when estimating (8) at the class level, which adjusts for correlated errors across classrooms and repeat student observations within a school. The more conservative approach of clustering by school increases standard errors by 30 percent (for earnings), but does not affect our hypothesis tests at conventional levels of statistical significance (online Appendix Table 7).[20]

Outliers.—In our baseline specifications, we exclude classrooms taught by teachers whose estimated VA $\hat{m}_{jt}$ falls in the top 1 percent for their subject and school level (above 2.03 in math and 1.94 in English in elementary school and 1.93 in math and 1.19 in English in middle school). We do so because these teachers' impacts on test scores appear suspiciously consistent with testing irregularities indicative of test manipulation. Jacob and Levitt (2003) develop a proxy for cheating which measures the extent to which a teacher generates very large test score gains which are followed by very large test score losses for the same students in the subsequent grade. Jacob and Levitt show that this proxy for cheating is highly correlated with unusual answer sequences which reveal test manipulation directly. Teachers in the top 1 percent of our estimated VA distribution are significantly more likely to show suspicious patterns of test score gains followed by steep losses, as defined by Jacob and Levitt's proxy (see online Appendix Figure 3).[21] We therefore trim the top 1 percent of outliers in all the specifications reported in the main text. We investigate how trimming at other cutoffs affects our results in online Appendix Table 8. The qualitative conclusion that teacher VA has long-term impacts is not sensitive to trimming, but including teachers in the top 1 percent reduces our estimates of teachers' impacts on long-term outcomes by 10–30 percent. In contrast, excluding the bottom 1 percent of the VA distribution has little impact on our estimates, consistent with the view that test manipulation is responsible for the results in the upper tail. Directly excluding teachers who have suspect classrooms based on Jacob and Levitt's proxy for cheating yields similar results

scores). As a result, $\mathbf{X}_{ct}$ does not exactly equal the mean of $\mathbf{X}_{it}^{I}$ within classrooms in the final estimation sample. To verify that the small discrepancies between $\mathbf{X}_{ct}$ and $\mathbf{X}_{it}^{I}$ do not affect our estimates of κ, we show in online Appendix Table 7 that the inclusion of individual controls $\mathbf{X}_{it}^{I}$ has little impact on the point estimates of κ by estimating (6) for a selected set of specifications on the individual data.

[20] In online Appendix Table 7 of Chetty, Friedman, and Rockoff (2011b), we evaluated the robustness of our results to other forms of clustering for selected specifications. 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.

[21] Online Appendix Figure 3 plots the fraction of classrooms which are in the top 5 percent according to Jacob and Levitt's proxy, defined in the notes to the figure, versus our leave-out-year measure of teacher value-added. On average, classrooms in the top 5 percent according to the Jacob and Levitt measure have test score gains of 0.47 standard deviations in year t followed by mean test score *losses* of 0.42 standard deviations in the subsequent year. Stated differently, teachers' impacts on future test scores fade out much more rapidly in the very upper tail of the VA distribution. Consistent with this pattern, these exceptionally high VA teachers also have very little impact on their students' long-term outcomes.

to trimming on VA itself. Because we trim outliers, our baseline estimates should be interpreted as characterizing the relationship between VA and outcomes below the ninety-ninth percentile of VA.

B. *College Attendance*

We begin by analyzing the impact of teachers' test-score VA on college attendance at age 20, the age at which college attendance rates are maximized in our sample. Panel A of Figure 1 plots residual college attendance rates for students in school year t versus $\hat{m}_{jt}$, the leave-year-out estimate of their teacher's VA in year t. To construct this binned scatter plot, we first residualize college attendance rates with respect to the class-level control vector $\mathbf{X}_{ct}$ separately within each subject by school-level cell, using within-teacher variation to estimate the coefficients on the controls as described above. We then divide the VA estimates $\hat{m}_{jt}$ into 20 equal-sized groups (vingtiles) and plot the mean of the college attendance residuals in each bin against the mean of $\hat{m}_{jt}$ in each bin. Finally, we add back the mean college attendance rate in the estimation sample to facilitate interpretation of the scale.[22] Note that 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 class-level data using (8), with standard errors clustered by school-cohort.

Panel A of Figure 1 shows that being assigned to a higher VA teacher in a single grade raises a student's probability of attending college significantly. The null hypothesis that teacher VA has no effect on college attendance is rejected with a t-statistic above 11 ($p < 0.001$). On average across subjects and grades, a one standard deviation increase in a teacher's true test score VA in a single grade increases the probability of college attendance at age 20 by $\kappa = 0.82$ percentage points, relative to a mean college attendance rate of 37.2 percent.[23]

We evaluate the robustness of this estimate to alternative control vectors in Table 2. Column 1 of Table 2 replicates the specification with the baseline control vector $\mathbf{X}_{ct}$ in panel A of Figure 1 as a reference. Column 2 replicates column 1, adding parent controls $\mathbf{P}^*_{ct}$ to the control vector. The parent characteristics $\mathbf{P}^*_{ct}$ consist of classroom means 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's marital status interacted with a quartic in parent's household income.[24] The estimate in column 2 is quite similar to that in column 1. Column 3 of Table 2 replicates column 1, adding class means of twice-lagged test scores $A^*_{c,t-2}$ to the control vector instead of

[22] In this and all subsequent scatter plots, we also demean $\hat{m}_{jt}$ within subject-by-school-level groups to isolate variation within these cells as in the regressions, and then add back the unconditional mean of $\hat{m}_{jt}$ in the estimation sample.

[23] These and all other estimates reported below reflect the value of a one standard deviation improvement in *actual* teacher VA m_{jt}, as shown in Section I. Being assigned to a teacher with higher *estimated* VA yields smaller gains because of noise in $\hat{m}_{jt}$ and drift in teacher quality, an issue we revisit in Section VI.

[24] We code the parent characteristics as zero for the 5.2 percent of students whom we matched to the tax data but were unable to link to a parent, and include an indicator for having no parent matched to the student. We also code mother's age at child's birth as zero for the small number of observations where we match parents but do not have data on parents' ages, and include an indicator for such cases.

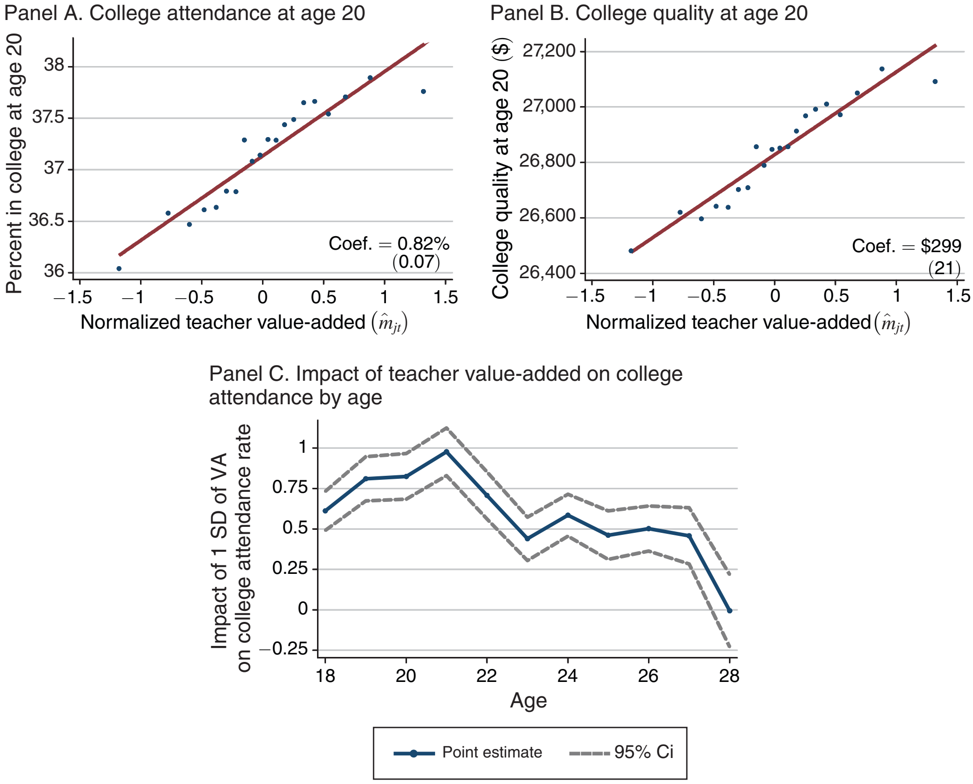

FIGURE 1. EFFECTS OF TEACHER VALUE-ADDED ON COLLEGE OUTCOMES

Notes: These figures are drawn using the linked analysis sample, pooling all grades and subjects, with one observation per student-subject-school year. Panels A and B are binned scatter plots of college attendance rates and college quality versus normalized teacher VA $\hat{m}_{jt}$. These plots correspond to the regressions in columns 1 and 4 of Table 2 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 class-level control vector (defined in the notes to Table 2) separately within each subject by school-level cell, using within-teacher variation to estimate the coefficients on the controls as described in Section IA. We then divide the VA estimates $\hat{m}_{jt}$ into 20 equal-sized groups (vingtiles) and plot the means of the y-variable residuals within each bin against the mean value of $\hat{m}_{jt}$ within each bin. Finally, we add back the unconditional mean of the y variable in the estimation sample to facilitate interpretation of the scale. The solid line shows the best linear fit estimated on the underlying microdata 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. In panel C, we replicate the regression in column 1 of Table 2 (depicted in panel A), varying the age of college attendance from 18 to 28, and plot the resulting coefficients. The dashed lines show the boundaries of the 95 percent confidence intervals for the effect of value-added on college attendance at each age, with standard errors clustered by school-cohort. The coefficients and standard errors from the regressions underlying panel C are reported in online Appendix Table 9.

parent characteristics. Again, the coefficient does not change appreciably.[25] Both parent characteristics and twice-lagged test scores are strong predictors of college attendance rates even conditional on the baseline controls $\mathbf{X}_{ct}$, with F-statistics exceeding 300. Hence, the fact that controlling for these variables does not significantly affect the estimates of κ supports the identification assumption of selection on observables in (7).

[25] The sample in column 3 has fewer observations than in column 1 because twice-lagged test scores are not observed in fourth grade. Replicating the specification in column 1 on exactly the estimation sample used in column 3 yields an estimate of 0.81 percent (0.09).

Table 2— Impacts of Teacher Value-Added on College Attendance

	College at age 20 (%) (1)	College at age 20 (%) (2)	College at age 20 (%) (3)	College quality at age 20 ($) (4)	College quality at age 20 ($) (5)	College quality at age 20 ($) (6)	High quality college (%) (7)	Four or more years of college, ages 18–22 (%) (8)
Teacher VA	0.82 (0.07)	0.71 (0.06)	0.74 (0.09)	298.63 (20.74)	265.82 (18.31)	266.17 (26.03)	0.72 (0.05)	0.79 (0.08)
Mean of dep. var.	37.22	37.22	37.09	26,837	26,837	26,798	13.41	24.59
Baseline controls	X	X	X	X	X	X	X	X
Parent chars. controls		X			X			
Lagged score controls			X			X		
Observations	4,170,905	4,170,905	3,130,855	4,167,571	4,167,571	3,128,478	4,167,571	3,030,878

Notes: Each column reports coefficients from an OLS regression, with standard errors clustered by school-cohort in parentheses. The regressions are estimated on the linked analysis sample (as described in the notes to Table 1). Teacher value-added is estimated using data from classes taught by a teacher in other years, following the procedure described in Section IIIA. The dependent variable in columns 1–3 is an indicator for college attendance at age 20. The dependent variable in columns 4–6 is the earnings-based index of college quality. See notes to Table 1 and Section II for more details on the construction of these variables. The dependent variable in column 7 is an indicator for attending a high-quality college, defined as quality greater than the median college quality among those attending college, which is $43,914. The dependent variable in column 8 is an indicator for attending four or more years of college between the ages of 18 and 22. All columns control for the baseline class-level control vector, which includes: 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 student-level characteristics including ethnicity, gender, age, lagged suspensions and absences, and indicators for grade repetition, special education, free or reduced-price lunch, and limited English; and grade and year dummies. Columns 2 and 5 additionally control for class means of parent characteristics, including 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's household income. Columns 3 and 6 include the baseline controls and class means of twice-lagged test scores. We use within-teacher variation to identify the coefficients on all controls as described in Section IA; the estimates reported are from regressions of outcome residuals on teacher VA with school by subject level fixed effects.

College Quality.—We use the same set of specifications to analyze whether high-VA teachers also improve the quality of colleges that their students attend, as measured by the earnings of students who previously attended the same college (see Section IIB). Students who do not attend college are included in this analysis and assigned the mean earnings of individuals who do not attend college. Panel B of Figure 1 plots the earnings-based index of quality for college attended at age 20 versus teacher VA, using the same baseline controls $\mathbf{X}_{ct}$ and technique as in panel A. Again, there is a highly significant relationship between the quality of colleges students attend and the quality of the teachers they had in grades 4–8 ($t = 14.4$, $p < 0.001$). A one standard deviation improvement in teacher VA raises college quality by $299 (or 1.11 percent) on average, as shown in column 4 of Table 2. Columns 5 and 6 replicate column 4, adding parent characteristics and lagged test score gains to the baseline control vector. As with college attendance, the inclusion of these controls has only a modest effect on the point estimates.

The \$299 estimate in column 4 combines intensive and extensive margin responses because it includes the effect of increased college attendance rates on projected earnings. Isolating intensive margin responses is more complicated because students who are induced to go to college by a high-VA teacher will tend to attend lower-quality colleges, pulling down mean earnings conditional on attendance. We take two approaches to overcome this selection problem and identify intensive-margin effects. First, we define colleges with earnings-based quality above the student-weighted median in our sample (\$43,914) as *high quality*. We regress this high-quality college indicator on teacher VA in the full sample, including students who do not attend college, and find that a one standard deviation increase in teacher VA raises the probability of attending a high-quality college by 0.72 percentage points, relative to a mean of 13.41 percent (column 7 of Table 2). This increase is most consistent with an intensive margin effect, as students would be unlikely to jump from not going to college at all to attending a high-quality college. Second, we derive a lower bound on the intensive margin effect by assuming that those who are induced to attend college attend a college of average quality. The mean college quality conditional on attending college is \$41,756, while the quality for all those who do not attend college is \$17,920. This suggests that at most $(41,756 - 17,920) \times 0.82$ percent $=$ \$195 of the \$299 impact is due to the extensive margin response, confirming that teachers improve the quality of colleges which students attend.

Finally, we analyze the impact of teacher quality on the number of years in college. Column 8 replicates the baseline specification in column 1, replacing the dependent variable with an indicator variable for attending college in at least four years between 18 and 22. A one standard deviation increase in teacher quality increases the fraction of students who spend four or more years in college by 0.79 percentage points (3.2 percent of the mean).[26] While we cannot directly measure college completion in our data, this finding suggests that higher-quality teachers increase not just attendance but also college completion rates.

Panel C of Figure 1 plots the impact of a one standard deviation improvement in teacher quality on college attendance rates at all ages from 18 to 28. We run separate regressions of college attendance at each age on teacher VA, using the same specification as in column 1 of Table 2. As one would expect, teacher VA has the largest impacts on college attendance rates before age 22. However, the impacts remain significant even in the mid-20s, perhaps because of increased attendance of graduate or professional schools. These continued impacts on higher education affect our analysis of earnings impacts, to which we now turn.

C. *Earnings*

The correlation between annual earnings and lifetime income rises rapidly as individuals enter the labor market and begins to stabilize only in their late 20s. We

[26] The magnitude of the four-year attendance impact (0.79 pp) is very similar to the magnitude of the single-year attendance impact (0.82 pp). Since the students who are on the margin of attending for one year presumably do not all attend for four years, this suggests that better teachers increase the number of years that students spend in college on the intensive margin.

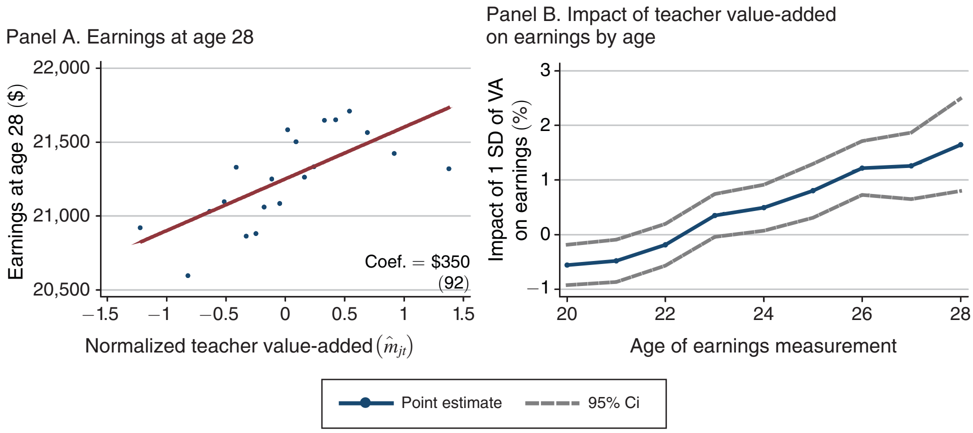

Figure 2. Effect of Teacher Value-Added on Earnings

Notes: Panel A is a binned scatter plot of earnings at age 28 versus normalized teacher VA $\hat{m}_{jt}$. This plot corresponds to the regression in column 1 of Table 3 and uses the same sample restrictions and variable definitions. See notes to Figure 1 for details on the construction of binned scatter plots. In panel B, we replicate the regression in column 1 of Table 3 (depicted in panel A), varying the age at which earnings are measured from 20 to 28. We then plot the resulting coefficients expressed as a percentage of mean wage earnings in the regression sample at each age. The dashed lines show the boundaries of the 95 percent confidence intervals for the effect of value-added on earnings at each age, with standard errors clustered by school-cohort. The coefficients and standard errors from the regressions underlying panel B are reported in online Appendix Table 9.

therefore begin by analyzing the impacts of teacher VA on earnings at age 28, the oldest age at which we have a sufficiently large sample of students to obtain precise estimates. Although individuals' earnings trajectories remain quite steep at age 28, earnings levels at age 28 are highly correlated with earnings at later ages (Haider and Solon 2006), a finding we confirm within the tax data in online Appendix Figure 4.

Panel A of Figure 2 plots individual (W-2) wage earnings at age 28 against VA $\hat{m}_{jt}$, conditioning on the same set of classroom-level controls as above. Being assigned to a higher value-added teacher has a significant impact on earnings, with the null hypothesis of $\kappa = 0$ rejected with $p < 0.01$. A one standard deviation increase in teacher VA in a single grade increases earnings at age 28 by \$350, 1.65 percent of mean earnings in the regression sample.

Columns 1–3 of Table 3 evaluate the robustness of this estimate to the inclusion of parent characteristics and lagged test score gains. These specifications mirror columns 1–3 of Table 2, but use earnings at age 28 as the dependent variable. As with college attendance, controlling for these additional observable characteristics has relatively small effects on the point estimates, supporting the identification assumption in (7). The smallest of the three estimates implies that a one standard deviation increase in teacher VA raises earnings by 1.34 percent.

To interpret the magnitude of this 1.34 percent impact, consider the lifetime earnings gain from having a one standard deviation higher VA teacher in a single grade. Assume that the percentage gain in earnings remains constant at 1.34 percent over the life cycle and that earnings are discounted at a 3 percent real rate (i.e., a 5 percent discount rate with 2 percent wage growth) back to age 12, the mean age in our

TABLE 3—IMPACTS OF TEACHER VALUE-ADDED ON EARNINGS

	Earnings at age 28 ($) (1)	Earnings at age 28 ($) (2)	Earnings at age 28 ($) (3)	Working at age 28 (%) (4)	Total income at age 28 ($) (5)	Wage growth ages 22–28 ($) (6)
Teacher VA	349.84	285.55	308.98	0.38	353.83	286.20
	(91.92)	(87.64)	(110.17)	(0.16)	(88.62)	(81.86)
Mean of dep. var.	21,256	21,256	21,468	68.09	22,108	11,454
Baseline controls	X	X	X	X	X	X
Parent chars. controls		X				
Lagged score controls			X			
Observations	650,965	650,965	510,309	650,965	650,965	650,943

Notes: Each column reports coefficients from an OLS regression, with standard errors clustered by school-cohort in parentheses. The regressions are estimated on the linked analysis sample (as described in the notes to Table 1). There is one observation for each student-subject-school year. Teacher value-added is estimated using data from classes taught by a teacher in other years, following the procedure described in Section IIIA. The dependent variable in columns 1–3 is the individual's wage earnings reported on W-2 forms at age 28. The dependent variable in column 4 is an indicator for having positive wage earnings at age 28. The dependent variable in column 5 is total income (wage earnings plus self-employment income). The dependent variable in column 6 is wage growth between ages 22 and 28. All columns control for the baseline class-level control vector; column 2 additionally controls for parent characteristics, while column 3 additionally controls for twice-lagged test scores (see notes to Table 2 for details). We use within-teacher variation to identify the coefficients on all controls as described in Section IA; the estimates reported are from regressions of outcome residuals on teacher VA with school by subject level fixed effects.

sample. Under these assumptions, the mean present value of lifetime earnings at age 12 in the US population is approximately \$522,000.[27] Hence, the financial value of having a one standard deviation higher VA teacher (i.e., a teacher at the eighty-fourth percentile instead of the median) is 1.34 percent $\times$ \$522,000 $\simeq$ \$7,000 per grade. The undiscounted lifetime earnings gain (assuming a 2 percent growth rate but 0 percent discount rate) is approximately \$39,000 per student.

A second benchmark is the increase in earnings from an additional year of schooling, which is around 9 percent (Gunderson and Oreopoulos 2010, Oreopoulos and Petronijevic 2013). Having a teacher in the first percentile of the value-added distribution (2.33 standard deviations below the mean) is equivalent to missing $\frac{2.33 \times 1.34 \text{ percent}}{9 \text{ percent}}$ = one-third of the school year when taught by a teacher of average quality.

A third benchmark is the cross-sectional relationship between test scores and earnings. A one standard deviation increase in teacher quality raises end-of-year scores by 0.13 standard deviations of the student test score distribution on average across grades and subjects. A one standard deviation increase in student test scores, controlling for the student- and class-level characteristics $\mathbf{X}_{it}$, is associated with a 12 percent increase in earnings at age 28 (online Appendix Table 3, column 3, row 2). The predicted impact of a one standard deviation increase in teacher VA

[27] We calculate this number using the mean wage earnings of a random sample of the US population in 2007 to obtain an earnings profile over the life cycle, and then inflate these values to 2010 US\$. See Chetty et al. (2011) for details.

on earnings is therefore 0.13×12 percent $= 1.55$ percent, similar to the observed impact of 1.34 percent.

Extensive Margin Responses and Other Sources of Income.—The increase in wage earnings comes from a combination of extensive and intensive margin responses. In column 4 of Table 3, we regress an indicator for having positive W-2 wage earnings on teacher VA using the same specification as in column 1. A one standard deviation increase in teacher VA raises the probability of working by 0.38 percent. If the marginal entrant into the labor market were to take a job that paid the mean earnings level in the sample ($21,256), this extensive margin response would raise mean earnings by $81. Since the marginal entrant most likely has lower earnings than the mean, this implies that the extensive margin accounts for at most $81/350 = 23$ percent of the total earnings increase due to better teachers.

Column 5 of Table 3 replicates the baseline specification in column 1 using total income (as defined in Section II) instead of wage earnings. Reassuringly, the point estimate of teachers' impacts changes relatively little with this broader income definition, which includes self-employment and other sources of income. We therefore use wage earnings—which provides an individual rather than household measure of earnings and is unaffected by the endogeneity of filing—for the remainder of our analysis.

Earnings Trajectories.—Next, we analyze how teacher VA affects the trajectory of earnings by examining wage earnings impacts at each age from 20 to 28. We run separate regressions of wage earnings at each age on teacher VA using the same specification as in column 1 of Table 3. Panel B of Figure 2 plots the coefficients from these regressions (which are reported in online Appendix Table 9), divided by average earnings at each age to obtain percentage impacts. The impact of teacher quality on earnings rises almost monotonically with age. At early ages, the impact of higher VA is *negative* and statistically significant, consistent with our finding that higher VA teachers induce their students to go to college. As these students enter the labor force, they have steeper earnings trajectories than students who had lower VA teachers in grades 4–8. Earnings impacts become positive at age 23, become statistically significant at age 24, and grow through age 28, where the earnings impact reaches 1.65 percent, as in Figure 2A.

An alternative way to state the result in panel B of Figure 2 is that better teachers increase the growth rate of students' earnings in their 20s. In column 6 of Table 3, we verify this result directly by regressing the change in earnings from age 22 to age 28 on teacher VA. As expected, a one standard deviation increase in teacher VA increases earnings growth by $286 (2.5 percent) over this period. This finding suggests that teachers' impacts on lifetime earnings could be larger than the 1.34 percent impact observed at age 28.

D. *Other Outcomes*

In this subsection, we analyze the impacts of teacher VA on other outcomes, starting with our "teenage birth" measure, which is an indicator for filing a tax return and claiming a dependent who was born while the mother was a teenager (see Section IIB). Column 1 of Table 4 analyzes the impact of teacher VA on the fraction of female students who have a teenage birth. Having a one standard deviation higher

TABLE 4—IMPACTS OF TEACHER VALUE-ADDED ON OTHER OUTCOMES (*percent*)

	Teenage birth (1)	Percent college grad in zip at age 28 (2)	Have 401(k) at age 28 (3)
Teacher VA	−0.61	0.25	0.55
	(0.06)	(0.04)	(0.16)
Mean of dep. var.	13.24	13.81	19.81
Baseline controls	X	X	X
Observations	2,110,402	468,021	650,965

Notes: Each column reports coefficients from an OLS regression, with standard errors clustered by school-cohort in parentheses. The regressions are estimated on the linked analysis sample (as described in the notes to Table 1). There is one observation for each student-subject-school year. Teacher value-added is estimated using data from classes taught by a teacher in other years, following the procedure described in Section IIIA. The dependent variables in column 1–3 are an indicator for having a teenage birth, the fraction of residents in an individual's zip code of residence at age 28 with a college degree or higher, and an indicator for whether an individual made a contribution to a 401(k) plan at age 28 (see notes to Table 1 and Section II for more details). Column 1 includes only female students. All regressions include the baseline class-level control vector (see notes to Table 2 for details). We use within-teacher variation to identify the coefficients on all controls as described in Section IA; the estimates reported are from regressions of outcome residuals on teacher VA with school by subject level fixed effects.

VA teacher in a single year from grades 4–8 reduces the probability of a teen birth by 0.61 percentage points, a reduction of roughly 4.6 percent, as shown in panel A of Figure 3. This impact is similar to the raw cross-sectional correlation between scores and teenage births (online Appendix Table 3), echoing our results on earnings and college attendance.

Column 2 of Table 4 analyzes the impact of teacher VA on the socioeconomic status of the neighborhood in which students live at age 28, measured by the percent of college graduates living in that neighborhood. A one standard deviation increase in teacher VA raises neighborhood SES by 0.25 percentage points (1.8 percent of the mean) by this metric, as shown in panel B of Figure 3. Column 3 of Table 4 shows that a one standard deviation increase in teacher VA increases the likelihood of saving money in a 401(k) at age 28 by 0.55 percentage points (or 2.8 percent of the mean), as shown in panel C of Figure 3.

Fade-Out of Test Score Impacts.—The final set of outcomes we consider are teachers' impacts on test scores in subsequent grades. Figure 4 plots the impacts of teacher VA on test scores in subsequent years; see online Appendix Table 10 for the underlying coefficients. To construct this figure, we residualize raw test scores $A^*_{i,t+s}$ with respect to the class-level controls $\mathbf{X}_{ct}$ using within-teacher variation and then regress the residuals $A_{i,t+s}$ on $\hat{\mu}_{jt}$. We scale teacher VA in units of student test-score standard deviations in these regressions—by using $\hat{\mu}_{jt}$ as the independent variable instead of $\hat{m}_{jt} = \hat{\mu}_{jt}/\sigma_0$—to facilitate interpretation of the regression coefficients, which are plotted in Figure 4. The coefficient at $s = 0$ is not statistically distinguishable from 1, as shown in our companion paper. Teachers' impacts on

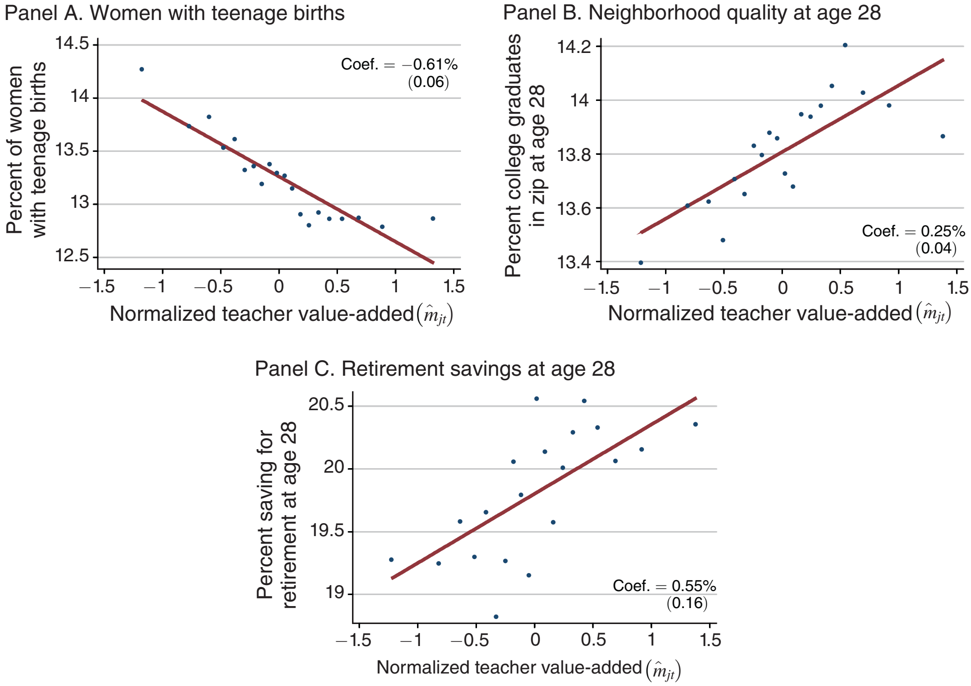

Figure 3. Effects of Teacher Value-Added on Other Outcomes in Adulthood

Notes: These three figures are binned scatter plots corresponding to columns 1–3 of Table 4 and use the same sample restrictions and variable definitions. See notes to Figure 1 for details on the construction of these binned scatter plots.

impacts on test scores fade out rapidly in subsequent years and appear to stabilize at approximately 25 percent of the initial impact after three to four years.[28] This result aligns with existing evidence that improvements in education raise contemporaneous scores, then fade out in later scores, only to reemerge in adulthood (Deming 2009; Heckman et al. 2010b; Chetty et al. 2011).

IV. Research Design 2: Teacher Switching Quasi-Experiments

The estimates in the previous section rely on the assumption that the unobserved determinants of students' long-term outcomes are uncorrelated with teacher quality conditional on observables. In this section, we estimate teachers' long-term impacts using a quasi-experimental design which relaxes and helps validate this identification assumption.

[28] Prior studies (e.g., Kane and Staiger 2008, Jacob, Lefgren, and Sims 2010, Rothstein 2010, Cascio and Staiger 2012) document similar fade-out after one or two years but have not determined whether test score impacts continue to deteriorate after that point. The broader span of our dataset allows us to estimate test score persistence more precisely. For instance, Jacob, Lefgren, and Sims (2010) estimate one-year persistence using 32,422 students and two-year persistence using 17,320 students. We estimate one-year persistence using more than 5.6 million student-year-subject observations and four-year persistence using more than 1.3 million student-year-subject observations.

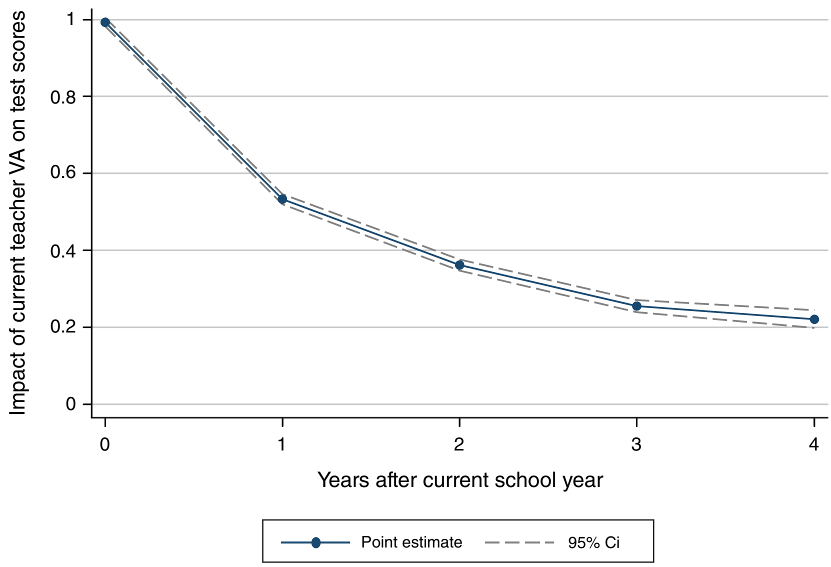

FIGURE 4. EFFECTS OF TEACHER VALUE-ADDED ON FUTURE TEST SCORES

Notes: This figure shows the effect of current teacher VA on test scores at the end of the current and subsequent school years. To construct this figure, we regress end-of-grade test scores in year $t + s$ on teacher VA $\hat{\mu}_{jt}$ in year t, varying s from 0 to 4. As in our companion paper, we scale teacher VA in units of student test-score standard deviations and include all students in the school district data in these regressions, without restricting to the older cohorts that we use to study outcomes in adulthood. We control for the baseline class-level control vector (defined in the notes to Table 2), using within-teacher variation to identify the coefficients on controls as described in Section IA. The dashed lines depict 95 percent confidence intervals on each regression coefficient, with standard errors clustered by school-cohort. The coefficients and standard errors from the underlying regressions are reported in online Appendix Table 10.

A. *Methodology*

Adjacent cohorts of students within a school are exposed to different teachers frequently. We exploit this teacher turnover to obtain a quasi-experimental estimate of teachers' long-term impacts. To understand our research design, 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 is 0.3 higher, so that the mean test-score VA of the teaching staff rises by $0.3/3 = 0.1$. If the distribution of unobserved determinants of students' long-term outcomes does not change between 1994 and 1995, the change in mean college attendance rates between the 1994 and 1995 cohorts of students will reveal the impact of a 0.1 improvement in fourth grade teachers' test-score VA. More generally, we can estimate teachers' long-term impacts by comparing the change in mean student outcomes across cohorts to the change in mean VA driven by teacher turnover provided that student quality is stable over time.

To formalize this approach, let $\hat{m}_{jt}^{-\{t,t-1\}}$ denote the test-score VA estimate for teacher j in school year t constructed as in our companion paper using data from all years except $t-1$ and t. Similarly, let $\hat{m}_{j,t-1}^{-\{t,t-1\}}$ denote the VA estimate for teacher j in school year $t-1$ based on data from all years except $t-1$ and t. Let Q_{sgt} denote the student-weighted mean of $\hat{m}_{jt}^{-\{t,t-1\}}$ across teachers in school s in grade g, which is the average estimated quality of teachers in a given school-grade-year cell; define $Q_{sg,t-1}$ analogously.[29] Let $\Delta Q_{sgt} = Q_{sgt} - Q_{sg,t-1}$ denote the change in mean teacher value-added from year $t - 1$ to year t in grade g in school s. Define mean changes in student outcome residuals ΔY_{sgt} analogously. Note that because we exclude both years t and $t - 1$ when estimating VA, the variation in ΔQ_{sgt} is driven purely by changes in the teaching staff and not by changes in teachers' VA estimates.[30] As above, this leave-out technique ensures that changes in ΔY_{sgt} are not spuriously correlated with ΔQ_{sgt} due to estimation error in VA.

We estimate teachers' long-term impacts by regressing changes in mean outcomes across cohorts on changes in mean test-score VA:

$$\Delta Y_{sgt} = \alpha + \kappa \Delta Q_{sgt} + \Delta \eta'_{sgt}. \tag{9}$$

Note that this specification is the same as the quasi-experimental specification we used to estimate the degree of bias in VA estimates in our companion paper, except that we use long-term outcomes as the dependent variable instead of test scores. The coefficient in (9) identifies the effect of a one standard deviation improvement in teacher quality as defined in (6) under the following assumption.

ASSUMPTION 3 (Teacher Switching as a Quasi-Experiment): *Changes in teacher quality across cohorts within a school-grade are orthogonal to changes in other determinants of student outcomes $\Delta \eta'_{sgt}$ across cohorts:*

$$\mathrm{Cov}(\Delta Q_{sgt}, \Delta \eta'_{sgt}) = 0. \tag{10}$$

This assumption could potentially be violated by endogenous student or teacher sorting. In practice, 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 sharp changes we analyze are likely driven by idiosyncratic shocks such as changes in staffing needs, maternity leaves, or the relocation of spouses. Moreover, in our first paper, we present direct evidence supporting (10) by showing that both prior scores and *contemporaneous* scores in the other subject (e.g., English) are uncorrelated with changes in mean teacher quality in a given subject (e.g., math). We also present additional evidence supporting (10) below.

If observable characteristics $\mathbf{X}_{it}$ are also orthogonal to changes in teacher quality across cohorts (i.e., satisfy Assumption 3), we can implement (9) simply by regressing the change in raw outcomes ΔY^*_{sgt} on ΔQ_{sgt}. We therefore begin with regressions

[29] In our baseline specifications, we impute teacher VA as the sample mean (0) for students for whom we have no leave-out-year VA estimate $\hat{m}_{jt}^{-\{t,t-1\}}$, either because we have no teacher information or because the teacher did not teach in the district outside of years $\{t - 1, t\}$. We show below that we obtain similar results when restricting to the subset of school-grade-subject-year cells with no missing data on teacher VA. See Section V of our companion paper for additional discussion on the effects of this imputation.

[30] Part of the variation in ΔQ_{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 ΔQ_{sgt}. As a result, isolating the variation due purely to teacher switching using an instrumental variables specification yields very similar results (not reported).

of ΔY^{*}_{sgt} on ΔQ_{sgt} and then confirm that changes in observable characteristics $\mathbf{X}_{it}$ across cohorts are uncorrelated with ΔQ_{sgt}.

B. *Results*

Panel A of Figure 5 presents a binned scatter plot of changes in mean college attendance rates ΔY^{*}_{sgt} against changes in mean teacher value-added ΔQ_{sgt} across cohorts. We include year fixed effects (demeaning both the x and y variables by school year), so that the estimate is identified purely from differential changes in teacher value-added across school-grade-subject cells over time. The corresponding regression coefficient, which is based on estimating (9) with year fixed effects but no other controls, is reported in column 1 of panel A of Table 5.

Changes in the quality of the teaching staff have significant impacts on changes in college attendance rates across consecutive cohorts of students in a school-grade-subject cell. The null hypothesis that $\kappa = 0$ is rejected with $p < 0.01$. The point estimate implies that a one standard deviation improvement in teacher quality raises college attendance rates by 0.86 percentage points, with a standard error of 0.23. This estimate is very similar to the estimates of 0.71–0.82 percentage points obtained from the first research design in Table 2. However, the quasi-experimental estimate is much less precise because it exploits less variation.

This analysis identifies teachers' causal impacts provided that (10) holds. One natural concern is that improvements in teacher quality may be correlated with other improvements in a school—such as better resources in other dimensions—which also contribute to students' long-term success and thus lead us to overstate teachers' true impacts. To address this concern, column 2 of panel A of Table 5 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} changes very little relative to the baseline estimate that pools all sources of variation. 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. Controlling for these variables has little impact on the estimate. This result shows that fluctuations in teacher quality relative to trend in specific grades generate significant changes in the affected students' college attendance rates.

In the preceding specifications, we imputed the sample mean of VA (0) for classrooms for which we could not calculate actual VA. This generates downward bias in our estimates because we mismeasure the change in teacher quality ΔQ_{sgt} across cohorts. Column 4 of panel A replicates column 2, limiting the sample to school-grade-subject-year cells in which we can calculate the leave-two-year-out mean for all teachers in the current and preceding year. As expected, the point estimate of a one standard deviation increase in teacher quality increases, but the confidence interval is significantly wider because the sample size is considerably smaller.

Finally, we further evaluate (10) using a series of placebo tests. In column 5 of panel A of Table 5, we replicate column 2, replacing the change in actual college

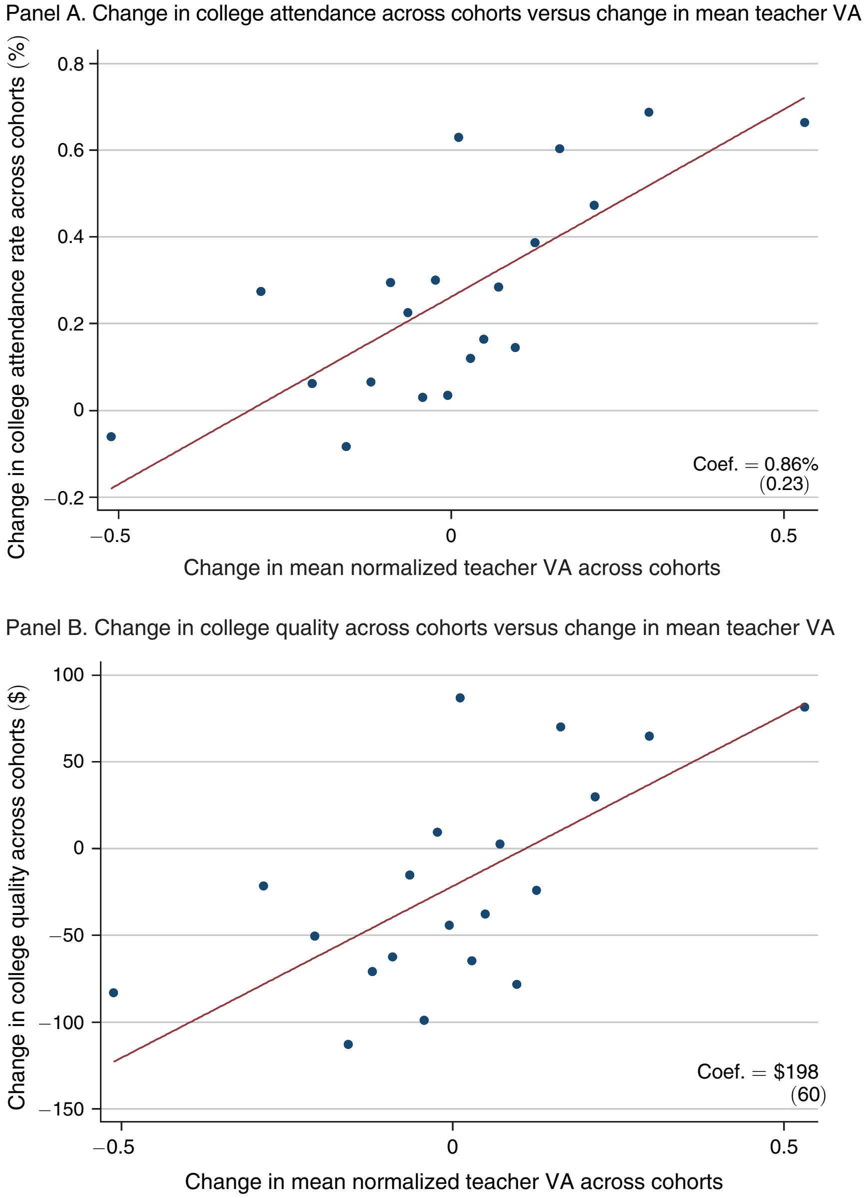

Figure 5. Effects of Changes in Teaching Staff across Cohorts on College Outcomes

Notes: These two figures plot changes in mean college attendance rates (measured in percentage points) and college quality across adjacent cohorts within a school-grade-subject cell against changes in mean teacher VA across those cohorts. These plots correspond to the regressions in column 1 of panels A and B of Table 5 and use the same sample restrictions and variable definitions. To construct these binned scatter plots, 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. Finally, we add back the unconditional (weighted) mean of the x and y variables in the estimation sample.

TABLE 5—IMPACTS OF TEACHER VALUE-ADDED ON COLLEGE OUTCOMES: QUASI-EXPERIMENTAL ESTIMATES

	College attendance (%)				Predicted college attendance (%)
	(1)	(2)	(3)	(4)	(5)
Panel A. College attendance at age 20					
Teacher VA	0.86	0.73	0.67	1.20	0.02
	(0.23)	(0.25)	(0.26)	(0.58)	(0.06)
Year FE	X				
School × year FE		X	X	X	X
Lagged score controls			X		
Lead and lag changes in teacher VA			X		
Number of school× grade × subject × year cells	33,167	33,167	26,857	8,711	33,167
Sample:	Full sample	Full sample	Full sample	No imputed scores	Full sample
	College quality ($)				Predicted college quality ($)
Panel B. College quality at age 20					
Teacher VA	197.64	156.64	176.51	334.52	2.53
	(60.27)	(63.93)	(64.94)	(166.85)	(18.30)
Year FE	X				
School × year FE		X	X	X	X
Lagged score controls			X		
Lead and lag changes in teacher VA			X		
Number of school × grade × subject × year cells	33,167	33,167	26,857	8,711	33,167
Sample:	Full sample	Full sample	Full sample	No imputed scores	Full sample

Notes: Each column reports coefficients from an OLS regression, with standard errors clustered by school-cohort in parentheses. The regressions are estimated on the linked analysis sample (as described in the notes to Table 1), collapsed to school-grade-year-subject means. The independent variable for each regression is the difference in mean teacher value-added between adjacent school-grade-year-subject cells, where we estimate teacher value-added using data which omits both years (see Section IVA for more details). Similarly, dependent variables are defined as changes in means across consecutive cohorts at the school-grade-year-subject level. In panel A, the dependent variable is college attendance at age 20; in panel B, the dependent variable is the earnings-based index of college quality (see Table 1 for details). In column 1 we regress the mean change in the dependent variable on the mean change in teacher value-added, controlling only for year fixed-effects. Column 2 replicates column 1 including school-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. In column 4, we restrict the sample to cells with no imputed VA; other columns impute the sample mean of 0 for classes with missing VA. Column 5 replicates column 2, except that the dependent variable is the predicted value from an individual-level regression of the original dependent variable on the vector of parent characteristics defined in the notes to Table 2.

attendance with the change in predicted college attendance based on parent characteristics. We predict college attendance using an OLS regression of Y_{it}^* on the same five parent characteristics $\mathbf{P}_{it}^*$ used in Section IIIB, with no other control variables. Changes in mean teacher VA have no effect on predicted college attendance rates, supporting the assumption that changes in the quality of the teaching staff are unrelated to changes in student quality at an annual level.

In panel A of Figure 6, we present an alternative set of placebo tests based on the sharp timing of the change in teacher quality. To construct this figure, we replicate

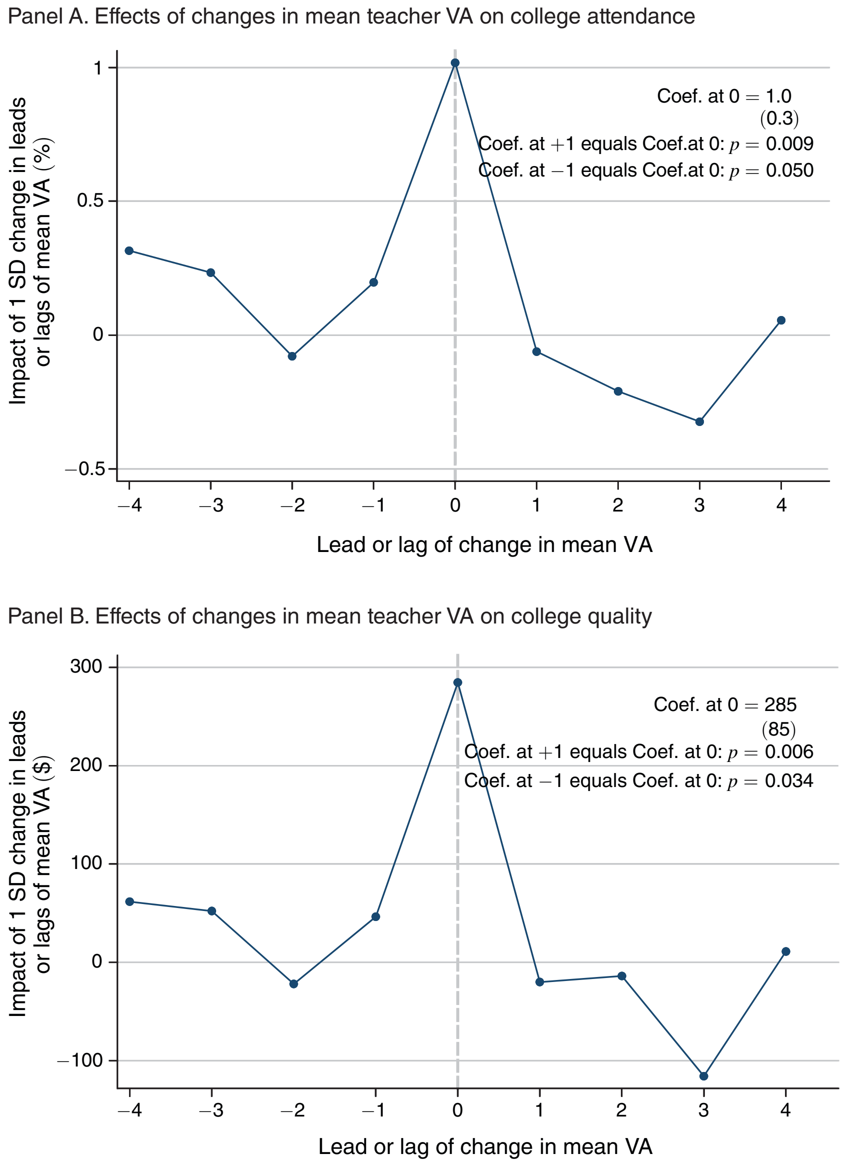

Figure 6. Timing of Changes in Teacher Quality and College Outcomes

Notes: These figures evaluate whether the timing of changes in teacher quality across cohorts aligns with the timing of changes in college outcomes. The point at 0 represents the *treatment effect* of changes in teacher quality on changes in college outcomes for a given group of students; the other points are *placebo tests* that show the impacts of changes in teacher quality for previous and subsequent cohorts on the same set of students. To construct panel A, we regress the change in mean college attendance between adjacent cohorts within a school-grade-subject cell on the change in mean teacher quality across those cohorts as well as four lags and four leads of the change in mean teacher quality within the same school-grade-subject. The regression also includes year fixed effects. Panel A plots the coefficients from this regression. We report the point estimate and standard error on the own-year change in mean teacher quality (corresponding to the value at 0). We also report p-values from hypothesis tests for the equality of the own-year coefficient and the one-year lead or one-year lag coefficients. These standard errors and p-values account for clustering at the school-cohort level. Panel B replicates panel A, replacing the dependent variable with changes in mean college quality across adjacent cohorts.

the specification in column 1 but include changes in mean teacher VA for the four preceding and subsequent cohorts as placebo effects (as well as year fixed effects):

$$\Delta Y^{*}_{sgt} = \alpha_t + \sum_{n=-4}^{4} \kappa_n \Delta Q_{sg,t+n} + \varphi_t. \tag{11}$$

Panel A of Figure 6 plots the vector of coefficients $\kappa = (\kappa_{-4}, \ldots, \kappa_0, \ldots, \kappa_4)$, which represent the impacts of changes in the quality of teaching staff at different horizons on changes in college attendance rates at time 0. As one would expect, κ_0 is positive and highly significant while all the other coefficients are near 0 and statistically insignificant. That is, contemporaneous changes in teacher quality have significant effects on college attendance rates, but past or future changes have no impact, as they do not affect the current cohort of students directly. The placebo tests in panel A support strongly the view that the changes in teacher VA have causal effects on college attendance rates.[31]

Panel B of Figure 5, panel B of Figure 6, and panel B of Table 5 replicate the preceding analysis using the earnings-based index of college quality as the outcome. We find that improvements in teachers' test-score VA across cohorts lead to sharp changes in the quality of colleges that students attend across all of the specifications. The point estimate in the baseline specification (column 1) implies that a one standard deviation improvement in teacher VA raises college quality by \$197 (s.e. = \$60), which is not statistically distinguishable from the estimates of \$266–\$299 obtained using the first research design. We find no evidence that predicted college quality based on parent characteristics is correlated with changes in teacher quality. In addition, changes in teacher VA again affect college quality in the year of the change rather than in preceding or subsequent years, as shown in panel B of Figure 6.

We used specifications analogous to those in Table 5 to investigate the impacts of teaching quality on other outcomes, including earnings at age 28. Unfortunately, our sample size for earnings at age 28 is roughly one-seventh the size of the sample available to study college attendance at age 20. This is both because we have fewer cohorts of students who are currently old enough to be observed at age 28 in the tax data and because we have data on teacher assignments for much fewer schools in the very early years of our school district data. Because of the considerably smaller sample, we obtain very imprecise and fragile estimates of the impacts of teacher quality on earnings using the quasi-experimental design.[32] Hence, we are forced to rely on estimates from the cross-class design in Table 3 to gauge earnings impacts.

[31] In Figure 3 of our companion paper, we directly use event studies around the entry and exit of teachers in the top and bottom 5 percent to demonstrate the impacts of VA on test scores. We do not have adequate power to identify the impacts of these exceptional teachers on college attendance using such event studies. In the cross-cohort regression that pools all teaching staff changes, the t-statistic for college attendance is 3.78 (column 1 of panel A of Table 5 in this paper). The corresponding t-statistic for test scores is 34.0 (column 2 of Table 5 of the first paper). We have much less power here both because the college attendance is only observed for the older half of our sample and because college is a much noisier outcome than end-of-grade test scores.

[32] For instance, estimating the specification in column 1 of Table 5 with earnings at age 28 as the dependent variable yields a confidence interval of (−\$581, \$665), which contains both 0 and values nearly twice as large as the estimated earnings impacts based on our first research design.

However, given that the cross-class and quasi-experimental designs yield very similar estimates of teachers' impacts on test scores, college attendance, and college quality, we expect the cross-class design to yield unbiased estimates of earnings impacts as well.

V. Heterogeneity of Teachers' Impacts

In this section, we analyze whether teachers' impacts vary across demographic groups, subjects, and grades. Because analyzing subgroup heterogeneity requires considerable statistical precision, we use the first research design—comparisons across classrooms conditional on observables—which as noted above relies on a stronger identification assumption but is bolstered by generating results which are similar to the quasi-experimental design in the full sample. We estimate impacts on college quality at age 20 (rather than earnings at age 28) to maximize precision and obtain a quantitative metric based on projected earnings gains.

A. *Demographic Groups*

In panel A of Table 6, we study the heterogeneity of teachers' impacts across demographic subgroups. Each value reported in the first row of the table is a coefficient estimate from a separate regression of college quality on teacher VA conditional on controls. To be conservative, we include both student characteristics $\mathbf{X}_{ct}$ and parent characteristics $\mathbf{P}^*_{ct}$ in the control vector throughout this section and estimate specifications analogous to column 5 of Table 2 on various subsamples. Columns 1 and 2 consider heterogeneity by gender. Columns 3 and 4 consider heterogeneity by parental income, dividing students into groups above and below the median level of parent income in the sample. Columns 5 and 6 split the sample into minority and non-minority students.

Two lessons emerge from panel A of Table 6. First, the point estimates of the impacts of teacher VA are larger for females than males, although we cannot reject equality of the impacts ($p = 0.102$). Second, the impacts are larger for higher-income and non-minority households in absolute terms. For instance, a one standard deviation increase in VA raises college quality by \$190 for children whose parents have below-median income, compared with \$380 for those whose parents have above-median income. However, the impacts are more similar as a percentage of mean college quality: 0.80 percent for low-income students versus 1.25 percent for high-income students.

The larger absolute impact for high socioeconomic students could be driven by two channels: a given increase in teacher VA could have larger impacts on the test scores of high SES students or a given increase in scores could have larger long-term impacts. The second row of coefficient estimates of panel A of Table 6 shows that a one standard deviation increase in teacher VA raises test scores by approximately 0.13 standard deviations on average in all the subgroups, consistent with the findings of Lockwood and McCaffrey (2009). In contrast, the cross-sectional correlation between scores and college quality is significantly larger for higher SES students (online Appendix Table 5). Although not conclusive, these findings suggest that the heterogeneity in teachers' long term impacts is driven by the second

TABLE 6— HETEROGENEITY IN IMPACTS OF TEACHER VALUE-ADDED

College quality at age 20 ($)	Female (1)	Male (2)	Low income (3)	High income (4)	Minority (5)	Non-minority (6)
Panel A. Impacts by demographic group						
Teacher VA	290.65	237.93	190.24	379.89	215.51	441.08
	(23.61)	(21.94)	(19.63)	(27.03)	(17.09)	(42.26)
Mean of dep. var.	27,584	26,073	23,790	30,330	23,831	33,968
Impact as percent of mean	1.05%	0.91%	0.80%	1.25%	0.90%	1.30%
Dep. var.: Test score (SD)						
Teacher VA	0.135	0.136	0.128	0.129	0.136	0.138
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean of dep. var.	0.196	0.158	−0.003	0.331	−0.039	0.651

College quality at age 20 ($)	Elementary school			Middle school	
	(1)	(2)	(3)	(4)	(5)
Panel B. Impacts by subject					
Math teacher VA	207.81		106.34	265.59	
	(21.77)		(28.50)	(43.03)	
English teacher VA		258.16	189.24		521.61
		(25.42)	(33.07)		(63.67)
Control for average VA in other subject				X	X

Notes: In the first row of estimates in panel A, we replicate the specification in column 5 of Table 2 within various population subgroups. In columns 1 and 2, we split the sample between males and females; in columns 3 and 4, we split the sample based on the median parent household income (which is $31,905); in columns 5 and 6, we split the sample based on whether a student belongs to an ethnic minority (Black or Hispanic). In the second row of estimates in panel A, we replicate all of the regressions from the first row replacing college quality with score as the dependent variable. In panel B, we split the sample into elementary schools (where the student is taught by the same teacher for both math and English) and middle schools (which have different teachers for each subject). Columns 1 and 2 replicate the specification in column 5 of Table 2, splitting the sample by subject. In column 3, we regress college quality on measures of math teacher value-added and English teacher value-added together in a dataset reshaped to have one row per student by school year. We restrict the sample so that the number of teacher-year observations is identical in columns 1–3. Columns 4 and 5 replicate column 5 of Table 2 for middle schools with an additional control for the average teacher value-added in the other subject for students in a given class.

mechanism, namely that high SES students' earnings are more sensitive to test score gains.[33] Overall, the heterogeneity in treatment effects on college quality suggests that teacher quality—at least as measured by test-score VA—is complementary to family inputs and resources. This result implies that higher income families should be willing to pay more for high-VA teachers.

B. *Subjects: Math versus English*

Panel B of Table 6 analyzes differences in teachers' impacts across subjects. For these regressions, we split the sample into elementary (columns 1–3) schools and middle (columns 4–5) schools. This distinction is important because students

[33] Importantly, the relationship between college quality and test scores conditional on prior characteristics X_{it} is linear throughout the test score distribution (online Appendix Figure 2b). Hence, the heterogeneity is not due to non-linearities in the relationship between scores and college outcomes but rather the fact that the same increase in scores translates to a bigger change in college outcomes for high-SES families.

have the same teacher for both subjects in elementary school but not middle school.

In column 1, we replicate the baseline specification in column 5 of Table 2, restricting the sample to math classrooms in elementary school. Column 2 repeats this specification for English. In column 3, we include each teacher's math and English VA together in the same specification, reshaping the dataset to have one row for each student-year (rather than one row per student-subject-year, as in previous regressions). Because a given teacher's math and English VA are highly correlated ($r = 0.6$), the magnitude of the two subject-specific coefficients drops by an average of 40 percent when included together in a single regression for elementary school. Intuitively, when math VA is included by itself in elementary school, it partly picks up the effect of having better teaching in English as well.

We find that a one standard deviation increase in teacher VA in English has larger impacts on college quality than a one standard deviation improvement in teacher VA in math. This is despite the fact that the variance of teacher effects in terms of test scores is *larger* in math than English. In Table 2 of our companion paper, we estimated that the standard deviation of teacher effects on student test scores in elementary school is 0.124 in English and 0.163 in math. Using the estimates from column 3 of panel B in Table 6, this implies that an English teacher who raises her students' test scores by one standard deviation raises college quality by $\frac{189/0.124}{106/0.163} = 2.3$ times as much as a math teacher who generates a commensurate test score gain. Hence, the returns to better performance in English are especially large, although it is much harder for teachers to improve students' achievement in English (e.g., Hanushek and Rivkin 2010; Kane et al. 2013).

We find a similar pattern in middle school. In column 4 of panel B, we replicate the baseline specification for the subset of observations in math in middle school. We control for teacher VA in English when estimating this specification by residualizing college quality Y_{it}^* with respect to the student and parent class-level control vectors $\mathbf{X}_{ct}$ and $\mathbf{P}_{it}^*$ as well as $\hat{m}_{jt}$ in English using a regression with math teacher fixed effects as in (3). Column 5 of panel B in Table 6 replicates the same regression for observations in English in middle school, controlling for math teacher VA. A one standard deviation improvement in English teacher quality raises college quality by roughly twice as much as a one standard deviation improvement in math teacher quality. Even though teachers have much smaller impacts on English test scores than math test scores, the small improvements that good teachers generate in English are associated with substantial long-term impacts.

C. *Impacts of Teachers by Grade*

We estimate the impact of a one standard deviation improvement in teacher quality in each grade $g \in [4, 8]$ on college quality (κ_g) by estimating the specification in column 5 of Table 2 but interacting $\hat{m}_{jt}$ with grade dummies. Because the school district data system did not cover many middle schools in the early and mid 1990s, we cannot analyze the impacts of teachers in grades 6–8 for more than half the students who are in grade 4 before 1994. To obtain a more balanced sample for

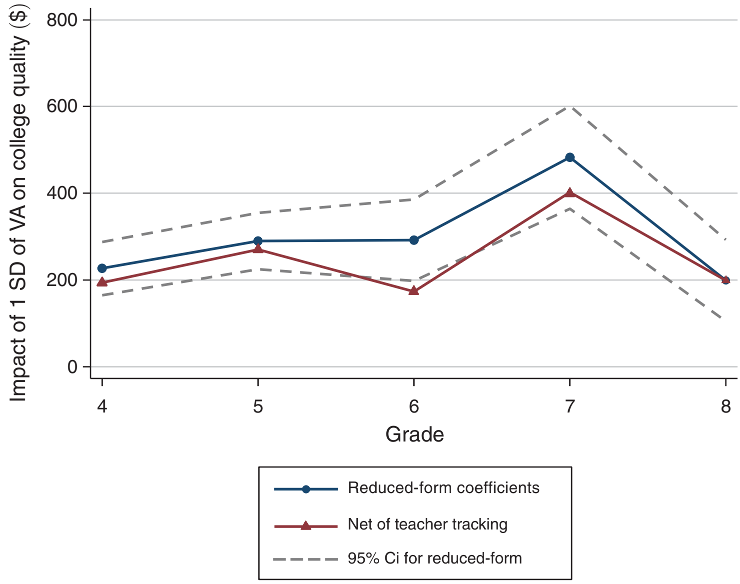

Figure 7. Impacts of Teacher Value-Added on College Quality by Grade

Notes: This figure plots the impact of a one standard deviation increase in teacher VA in each grade from 4 to 8 on our earnings-based index of college quality (defined in the notes to Table 1). The upper (circle) series shows the reduced-form effect of improved teacher quality in each grade, including both the direct impact of the teacher on earnings and the indirect effect through improved teacher quality in future years. To generate this series, we replicate column 5 of Table 2, interacting VA with grade dummies. We restrict the sample to cohorts who would have been in fourth grade during or after 1994 to obtain a balanced sample across grades. The dots in the series plot the coefficients on each grade interaction. The dashed lines show the boundaries of the 95 percent confidence intervals for the reduced-form effects, clustering the standard errors by school-cohort. The lower (triangle) series plots the impact of teachers in each grade on college quality netting out the impacts of increased future teacher quality. We net out the effects of future teachers using the tracking coefficients reported in Appendix Table 12 and solving the system of equations in Section VC. Online Appendix Table 11 reports the reduced-form effects and net-of-future-teachers effects plotted in this figure.

comparisons across grades, we restrict attention to cohorts who would have been in grade 4 during or after 1994 in this subsection.[34]

The series in circles in Figure 7 plots the estimates of κ_g, which are also reported in online Appendix Table 11. We find that teachers' long-term impacts are large and significant in all grades. Although the estimates in each grade have relatively wide confidence intervals, there is no systematic trend in the impacts. This pattern is consistent with the cross-sectional correlations between test scores and adult outcomes, which are also relatively stable across grades (online Appendix Table 6). One issue which complicates cross-grade comparisons is that teachers spend almost the entire school day with their students in elementary school (grades 4–5 as well as 6 in some schools), but only their subject period (math or English) in middle school (grades 7–8). If teachers' skills are correlated across subjects—as is the case with math and English value-added, which have a correlation of 0.6 for elementary

[34] Restricting the sample in the same way does not affect the conclusions above about heterogeneity across subjects or demographic groups, because these groups are balanced across cohorts.

school teachers—then a high-VA teacher should have a greater impact on earnings in elementary school than middle school because they spend more time with the student. Hence, the fact that high-VA math and English teachers continue to have substantial impacts in middle school indicates that education has substantial returns well beyond early childhood.

Tracking and Net Impacts.—The reduced-form estimates of κ_g reported above include the impacts of being tracked to a better teacher in subsequent grades, as discussed in Section I. While a parent may be interested in the reduced-form impact of teacher VA in grade g, a policy that raises teacher quality in grade g will not allow every child to get a better teacher in grade $g + 1$. We now turn to identifying teachers' net impacts $\tilde{\kappa}_g$ in each grade, holding fixed future teachers' test-score VA.

Because we have no data after grade 8, we can only estimate teachers' net effects holding fixed teacher quality up to grade 8. We therefore set $\tilde{\kappa}_8 = \kappa_8$. We recover $\tilde{\kappa}_g$ from estimates of κ_g by subtracting out the impacts of future teachers on earnings iteratively. The net impact of a seventh grade teacher is her reduced-form impact κ_7 minus her indirect impact via tracking to a better eighth grade teacher:

$$\tilde{\kappa}_7 = \kappa_7 - \rho_{78}\tilde{\kappa}_8, \tag{12}$$

where ρ_{78} is the extent to which teacher VA in grade 7 increases teacher VA in grade 8 conditional on controls. If we observed each teacher's true VA, we could estimate ρ_{78} using an OLS regression that parallels (8) with the future teacher's true VA as the dependent variable:

$$m_{j,t_i(8)} = \alpha + \rho_{78}\hat{m}_{j,t_i(7)} + \gamma_1\mathbf{X}_{ct} + \gamma_2\mathbf{P}^*_{ct} + \eta^{\mu}_{ct78}. \tag{13}$$

Since true VA is unobserved, we substitute VA estimates $\hat{m}_{j,t_i(8)}$ for $m_{j,t_i(8)}$ on the left-hand side of (13). This yields an attenuated estimate of ρ_{78} because $\hat{m}_{j,t_i(8)}$ is shrunk toward zero to account for estimation error (see Section IB of our companion paper). If all teachers taught the same number of classes and had the same number of students, the shrinkage factor would not vary across observations. In this case, we could identify ρ_{78} by using $\hat{m}_{j,t_i(8)}$ as the dependent variable in (13) and multiplying the estimate of ρ_{78} by $SD(m_{jt})/SD(\hat{m}_{jt}) = 1/SD(\hat{m}_{jt})$. In the sample for which we observe college attendance, the standard deviation of teacher VA estimates is $SD(\hat{m}_{jt}) = 0.61$. We therefore multiply the estimate of ρ_{78} and all the other tracking coefficients $\rho_{gg'}$ by $1/SD(\hat{m}_{jt}) = 1.63$. This simple approach to correcting for the attenuation bias is an approximation because the shrinkage factor does vary across observations. However, as we discuss below, the magnitude of the tracking coefficients is small and hence further adjusting for the variation in shrinkage factors is unlikely to affect our conclusions.

We estimate ρ_{78} in three steps. First, we residualize $\hat{m}_{j,t_i(8)}$ with respect to the controls by regressing $\hat{m}_{j,t_i(8)}$ on $\mathbf{X}_{ct}$ and $\mathbf{P}^*_{ct}$ with grade 7 teacher fixed effects, as in (3). Second, we run a univariate OLS regression of the residuals on $\hat{m}_{j,t_i(7)}$. Finally, we multiply this coefficient by $1/SD(\hat{m}_{jt}) = 1.63$ to obtain an estimate of $\hat{\rho}_{78}$. Using our estimate of $\hat{\rho}_{78}$, we apply (12) to identify $\tilde{\kappa}_7$ from the reduced-form estimates of

κ_g in Figure 7. Iterating backwards, we calculate κ_6 by estimating $\hat{\rho}_{68}$ and $\hat{\rho}_{67}$ and so on until we obtain the full set of net impacts. We show formally that this procedure recovers net impacts $\tilde{\kappa}_g$ in online Appendix C.

The series in triangles in Figure 7 plots the estimates of the net impacts $\tilde{\kappa}_g$. The differences between the net impacts and reduced-form impacts are modest: the mean difference between κ_g and $\tilde{\kappa}_g$ is 19.6 percent in grades 4–7. This is because the tracking coefficients $\rho_{g,g'}$ are small. A one standard deviation increase in current teacher quality typically raises future teacher quality by about 0.05 standard deviations or less in elementary school and by about 0.25 standard deviations in middle school (online Appendix Table 12). The greater degree of tracking in middle school is consistent with the availability of subject-specific honors classes in grades 7 and 8.

These results suggest that the vast majority of the reduced-form impacts estimated above reflect a teacher's own direct impact rather than the impacts of being tracked to better teachers in later grades. However, we caution that this approach to calculating teachers' net impacts has three important limitations. First, it assumes that all tracking to future teachers occurs exclusively via teachers' test-score VA. We allow students who have high-VA teachers in grade g to be tracked to higher test-score VA (m_{jt}) teachers in grade $g+1$, but *not* to teachers with higher earnings VA μ_{jt}^{Y}. We are forced to make this strong assumption because we have no way to estimate teacher impacts on earnings that are orthogonal to VA, as discussed in Section I. Second, $\tilde{\kappa}_g$ does not net out potential changes in other factors besides teachers, such as peer quality or parental inputs. Hence, $\tilde{\kappa}_g$ cannot be interpreted as the *structural* impact of teacher quality holding fixed all other inputs in a general model of the education production function (e.g., Todd and Wolpin 2003). Finally, our approach assumes that teacher effects are additive across grades. We cannot identify complementarities in teacher VA across grades because our identification strategy forces us to condition on lagged test scores, which are endogenous to the prior teacher's quality. It would be very useful to relax these assumptions in future work to obtain a better understanding of how the sequence of teachers a child has affects her outcomes in adulthood.

VI. Policy Analysis

In this section, we use our estimates to predict the potential earnings gains from selecting and retaining teachers on the basis of their VA. We make four assumptions in our calculations. First, we assume that the percentage impact of a one standard deviation improvement in teacher VA on earnings observed at age 28 is constant at $b = 1.34$ percent (Table 3, column 2) over the life cycle.[35] Second, we ignore general equilibrium effects which may reduce wage rates if all children are better educated. Third, we follow Krueger (1999) and discount earnings gains at a 3 percent real annual rate (consistent with a 5 percent discount rate and 2 percent wage growth) back to age 12, the average age in our sample. Under this assumption, the present value of earnings at age 12 for the average individual in the US population

[35] We have inadequate precision to estimate wage earnings impacts separately by subject and grade level. We therefore assume that a one standard deviation improvement in teacher VA raises earnings by 1.34 percent in all subjects and grade levels in the calculations that follow.

is \$522, 000 in 2010 dollars, as noted above. Finally, we assume that teacher VA m_{jt} is normally distributed.

To quantify the value of improving teacher quality, we evaluate Hanushek's (2009, 2011) proposal to replace teachers whose VA ratings are in the bottom 5 percent of the distribution with teachers of average quality. To simplify exposition, we calculate these impacts for elementary school teachers, who teach one classroom per year. We first calculate the earnings gains from selecting teachers based on their true test-score VA m_{jt} and then calculate the gains from selecting teachers based on VA estimates $\hat{m}_{jt}$. Selection on true VA is informative as a benchmark for the potential gains from improving teacher quality, but is not a feasible policy because we only observe estimates of teacher VA in practice. Hence, it is useful to compare the gains from policies that select teachers based on VA estimates using a few years of performance data with the maximum attainable gain if one were to select teachers based on true VA.

Selection on True VA.—Elementary school teachers teach both math and English and therefore have two separate VA measures on which they could be evaluated. For simplicity, we assume that teachers are evaluated based purely on their VA in one subject (say math) and VA in the other subject is ignored.[36] Consider a student whose teacher's true math VA is Δm_σ standard deviations below the mean. Replacing this teacher with a teacher of mean quality (for a single school year) would raise the student's expected earnings by

$$G = \Delta m_\sigma \times \$522,000 \times b. \tag{14}$$

Under the assumption that m_{jt} is normally distributed, a teacher in the bottom 5 percent of the true VA distribution is on average 2.063 standard deviations below the mean teacher quality. Therefore, replacing a teacher in the bottom 5 percent of the VA distribution with an average teacher generates a present value lifetime earnings gain per student of

$$G = 2.063 \times \$522,000 \times 1.34\% = \$14,500.$$

For a class of average size (28.2), the total NPV (net present value) earnings impact from this replacement is $G_C = \$407,000$. The undiscounted cumulative lifetime earnings gains from deselection are 5.5 times larger than these present value gains (\$80,000 per student and \$2.25 million per classroom), as shown in online Appendix Table 13.[37] These simple calculations show that the potential gains from

[36] School districts typically average math and English VA ratings to calculate a single measure of teacher performance in elementary schools (e.g., DC Public Schools 2012). In online Appendix D, we show that the gains from evaluating teachers based on mean math and English VA are only 12 percent larger than the gains from using information based on only one subject because math and English VA estimates are highly correlated. Therefore, the results we report below slightly underestimate the true gains from selection on mean VA.

[37] These calculations do not account for the fact that deselected teachers may be replaced by rookie teachers, who have lower VA. In our sample, mean test score residuals for students taught by first-year teachers are on average 0.05 standard deviations lower (in units of standardized student test scores) than those taught by more experienced teachers. Given that the median teacher remains in the district for approximately ten years, accounting for the effect of inexperience in the first year would reduce the expected benefits of deselection over a typical

improving the quality of teaching—whether using selection based on VA, teacher training, or other policies—are quite large.[38]

Selection on Estimated VA.—In practice, we can only select teachers on the basis of estimated VA $\hat{m}_{jt}$. This reduces the gains from selection for two reasons: (1) estimation error in VA and (2) drift in teacher quality over time. To quantify the impact of these realities, suppose we use test score data from years $t = 1, \ldots, n$ to estimate teacher VA in school year $n + 1$. The gain in year $n + 1$ from replacing the bottom 5 percent of teachers based on VA estimated using the preceding n years of data is

$$G(n) = -E[m_{j,n+1} \,|\, \hat{m}_{j,n+1} < F^{-1}_{\hat{m}_{j,n+1}}(0.05)] \times \$522,000 \times b, \tag{15}$$

where $E[m_{j,n+1} \,|\, \hat{m}_{j,n+1} < F^{-1}_{\hat{m}_{j,n+1}}(0.05)]$ denotes the expected value of $m_{j,n+1}$ conditional on the teacher's estimated VA falling below the fifth percentile. We calculate this expected value separately for math and English using Monte Carlo simulations as described in online Appendix D.[39]

Panel A of Figure 8 plots the mean gain per classroom $G_C(n) = 28.2 \times G(n)$, averaging over math and English, for $n = 1, \ldots, 10$. The values underlying this figure are reported in online Appendix Table 13. The gain from deselecting teachers based on true VA, $G_C = \$407,000$, is shown by the horizontal line in the figure. The gains from deselecting teachers based on estimated VA are significantly smaller because of noise in VA estimates and drift in teacher quality.[40] With one year of data, the expected gain per class is \$226,000, 56 percent of the gain from selecting on true VA. The gains grow fairly rapidly with more data in the first three years, but the marginal value of additional information is small. With three years of test score data, the gain is \$266,000, but the gain increases to only \$279,000 after ten years. After three years, waiting for one more year would increase the gain by \$4,000 but has an expected cost of \$266,000. The marginal gains from obtaining one more year of data are outweighed by the expected cost of having a low VA teacher on the staff even after the first year (Staiger and Rockoff 2010). Adding data from prior classes yields relatively little information about current teacher quality both because of decreasing returns to additional observations and drift.[41]

horizon by $\frac{0.05/10}{2.063 \times \sigma(m_{jt})} = 2$ percent, where $\sigma(m_{jt}) = 0.14$ is the mean standard deviation of teacher effects across elementary school subjects in our data (see Table 2 of the first paper).

[38] Moreover, these calculations do not include the non-monetary returns to a better education (Oreopoulos and Salvanes 2010), such as lower teenage birth rates.

[39] Without drift, the formula in (15) reduces to $r(n)^{1/2} \times 2.063 \times \$522,000 \times b$, where $r(n)$ denotes the reliability of the VA estimate using n years of data, which is straightforward to calculate analytically. The original working paper version of our study (Chetty, Friedman, and Rockoff 2011b) used this version of the formula and an estimate of b based on a model that did not account for drift.

[40] In panel B of Appendix Table 13, we distinguish these two factors by eliminating estimation error and predicting current VA based on past VA instead of past scores. Without estimation error, $G_C(1) = \$340,000$. Hence, drift and estimation error each account for roughly half of the difference between $G_C(1)$ and G_C.

[41] We also replicated the simulations using VA estimates that do not account for drift. When the estimation window n is short, drift has little impact on the weights placed on test scores across years. As a result, drift-unadjusted measures yield rankings of teacher quality that are very highly correlated with our measures and thus produce similar gains. For instance, selection based on three years of data using VA estimates that do not adjust for drift yields gains that are 98 percent as large as those reported above. Hence, while accounting for drift is important for evaluating out-of-sample forecasts accurately, it may not be critical for practical policy applications for VA.

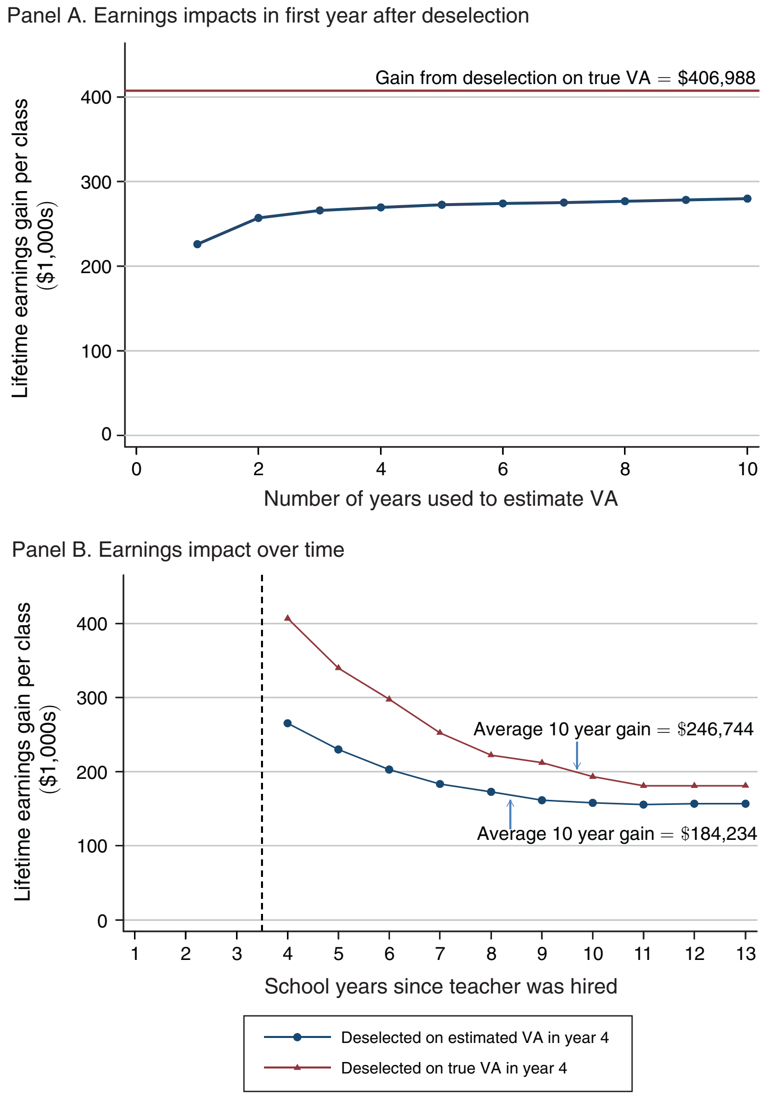

Figure 8. Earnings Impacts of Deselecting Low Value-Added Teachers

Notes: This figure analyzes the impacts of replacing teachers with VA in the bottom 5 percent with teachers of average quality on the present value of lifetime earnings for a single classroom of average size (28.2 students). In panel A, the horizontal line shows the hypothetical gain from deselecting the bottom 5 percent of teachers based on their true VA in the current school year. The series in circles plots the gains from deselecting teachers on estimated VA versus the number of years of prior test score data used to estimate VA. Panel A shows gains for the school year immediately after deselection; panel B shows the gains in subsequent school years, which decay over time due to drift in teacher quality. The lower series in panel B (in circles) plots the earnings gains in subsequent school years from deselecting teachers based on their VA estimate at $t = 4$, constructed using the past three years of data. The first point in this series (at $t = 4$) corresponds to the third point in panel A by construction. The upper series (in triangles) shows the hypothetical gains obtained from deselecting the bottom 5 percent of teachers based on their true VA at $t = 4$; the first dot in this series matches the value in the horizontal line in panel A. For both series in panel B, we also report the unweighted mean gain over the first ten years after deselection. All values in these figures are based on our estimate that a one standard deviation increase in true teacher VA increases earnings by 1.34 percent (column 2 of Table 3). All calculations assume that teachers teach one class per year and report mean values for math and English teachers, which are calculated separately. We calculate earnings gains using Monte Carlo simulations based on our estimates of the teacher VA process as described in Section VI. All values in these figures and their undiscounted equivalents are reported in online Appendix Table 13.

Drift in Quality over Subsequent School Years.—The values in panel A of Figure 8 reflect the gains in the first year after the deselection of teachers, based on $\hat{m}_{j,n+1}$ in school year $n + 1$. Now consider the impacts of such a policy on the earnings of students in a subsequent school year $n + m$:

$$G(m,n) = -E[m_{j,n+m} | \hat{m}_{j,n+1} < F^{-1}_{\hat{m}_{j,n+1}}(0.05)] \times \$522{,}000 \times b,$$

where $E[m_{j,n+m} | \hat{m}_{j,n+1} < F^{-1}_{\hat{m}_{j,n+1}}(0.05)]$ denotes the mean VA of teachers in year $n + m$ conditional on having estimated VA in year $n + 1$ below the fifth percentile.

The lower series in panel B of Figure 8 plots $G_C(m,3) = 28.2 \times G(m,3)$, the gains per class in school year m from deselecting teachers based on their estimated VA for year 4 ($\hat{m}_{j,4}$), constructed using the first three years of data. The first point in this series coincides with the value of \$266,000 in panel A reported for $n = 3$. Because teacher quality drifts over time, the gains fall in subsequent school years, as some of the teachers who were deselected based on their predicted VA in school year n would have reverted toward the mean in subsequent years. Deselection based on VA estimates at the end of year three generates an average gain of \$184,000 per classroom per year over the subsequent ten years, the median survival time in the district for teachers who have taught for three years.

The upper series in panel B of Figure 8 plots the analogous gains when teachers are deselected based on their true VA $m_{j,4}$ in year four instead of their estimated VA $\hat{m}_{j,4}$. The first point in this series coincides with the maximum attainable gain of \$407,000 shown in panel A. The gains again diminish over time because of drift in teacher quality. The average present value gain from deselection based on true VA over the subsequent ten years is approximately \$250,000 per classroom. This corresponds to an undiscounted lifetime earnings gain per classroom of students of approximately \$1.4 million.

The reason that estimation error and drift do not heavily erode the gains from deselection of low VA teachers is that very few of the teachers rated in the bottom 5 percent turn out to be high-VA teachers. For example, among math teachers in elementary school, 3.2 percent of the teachers whose estimated VA based on three years of test score data ($\hat{m}_{j,4}$) is in the bottom 5 percent have true VA $m_{j,4}$ above the median. Nevertheless, because VA estimates are not perfect predictors of m_{jt}, there is still substantial room to use other measures—such as principal evaluations or student surveys—to complement VA estimates and improve predictions of teacher quality.[42]

Costs of Teacher Selection.—The calculations above do not account for the costs associated with a policy which deselects teachers with the lowest performance ratings. First, they ignore downstream costs that may be required to generate earnings gains, most notably the cost associated with higher college attendance rates. Second, and more importantly, they ignore the fact that teachers need to be compensated for the added employment risk they face from such an evaluation system. Rothstein

[42] Of course, these other measures will also be affected by drift and estimation error. For instance, classroom observations have significant noise and may capture transitory fluctuations in teacher quality (Kane et al. 2013).

(2013) estimates the latter cost using a structural model of the labor market for teachers. Rothstein estimates that a policy which fires teachers if their estimated VA after three years falls below the fifth percentile would require a mean salary increase of 1.4 percent to equilibrate the teacher labor market.[43] In our sample, mean teacher salaries were approximately \$50,000, implying that annual salaries would have to be raised by approximately \$700 for all teachers to compensate them for the additional risk. Based on our calculations above, the deselection policy would generate NPV gains of \$184,000 per teacher deselected, or \$9,250 for all teachers on average (because only 1 out of 20 teachers would actually be deselected). Hence, the estimated gains from this policy are more than ten times larger than the costs. Together with the preceding results, Rothstein's (2013) findings imply that deselecting low-VA teachers could be a very cost effective policy if the signal quality of VA does not fall substantially when used for personnel evaluation.[44]

Retention of High-VA Teachers.—An alternative approach to improving teacher quality that may impose lower costs on teachers is to increase the retention of high-VA teachers by paying them bonuses. Using Monte Carlo simulations analogous to those above, we estimate that retaining a teacher at the ninety-fifth percentile of the estimated VA distribution (using three years of data) for an extra year would yield present value earnings gains in the subsequent school year of $\$522{,}000 \times 28.2 \times 1.34$ percent $\times$ $E[m_{j,n+1} \,|\, \hat{m}_{j,n+1} = F^{-1}_{\hat{m}_{j,n+1}}(0.95)] = \$212{,}000$. In our data, roughly 9 percent of teachers in their third year do not return to the school district for a fourth year. The attrition rate is unrelated to teacher VA, consistent with the findings of Boyd et al. (2008). Clotfelter et al. (2008) estimate that a \$1,800 bonus payment in North Carolina reduces attrition rates by 17 percent. Based on these estimates, a one-time bonus payment of \$1,800 to high-VA teachers who return for a fourth year would increase retention rates in the next year by 1.5 percentage points and generate an average benefit of \$3,180. The expected benefit of offering a bonus to even an excellent (ninety-fifth percentile) teacher is only modestly larger than the cost because one must pay bonuses to $(100 - 9)/1.5 \approx 60$ additional teachers for every extra teacher retained.

Replacing ineffective teachers is more cost-effective than attempting to retain high-VA teachers because most teachers stay for the following school year and are relatively inelastic to salary increases. Of course, increasing the salaries of high-VA teachers could attract more talented individuals into teaching to begin with or increase teacher effort. The preceding calculations do not account for these effects.

VII. Conclusion

Our first paper (Chetty, Friedman, and Rockoff 2014) showed that existing test-score value-added measures are a good proxy for a teacher's ability to raise students' test scores. This paper has shown that the same VA measures are also an informative

[43] In the working paper version of his study, Rothstein (2013) calculates the wage gains needed to compensate teachers for a policy that deselects teachers below the twentieth percentile after two years. Jesse Rothstein kindly provided the corresponding estimates for the policy analyzed here in personal correspondence.

[44] Even if there is erosion, as long as the signal quality of VA is a continuous function of its weight in evaluation decisions, one would optimally place non-zero weight on VA, because the net gains would fall from the initial level of \$184,000 in proportion to the weight on VA.

proxy for teachers' long-term impacts. Although these findings are encouraging for the use of value-added metrics, two important issues must be resolved before one can determine how VA should be used for policy.

First, using VA measures to evaluate teachers could induce responses such as teaching to the test or cheating, eroding the signal in VA measures (e.g., Jacob 2005, Neal and Schanzenbach 2010).[45] One can estimate the magnitude of such effects by replicating the analysis in this paper in a district that evaluates teachers based on their VA. If behavioral responses substantially reduce the signal quality of VA, policymakers may need to develop metrics which are more robust to such responses, as in Barlevy and Neal (2012). For instance, districts may also be able to use data on the persistence of test score gains to identify test manipulation and develop a more robust estimate of teacher quality, as in Jacob and Levitt (2003).

Second, one should compare the long-term impacts of evaluating teachers on the basis of VA to other metrics, such as principal evaluations or classroom observation. One can adapt the methods developed in this paper to evaluate these other measures of teacher quality. When a teacher who is rated highly by principals enters a school, do subsequent cohorts of students have higher college attendance rates and earnings? What fraction of a teacher's long-term impact is captured by test-score VA versus other measures of teacher quality? By answering these questions, ultimately one could estimate the optimal weighting of available metrics to identify teachers who are most successful in improving students' long-term outcomes.

More generally, there are many aspects of teachers' long-term impacts that remain to be explored and would be helpful in designing education policy. For example, in this paper we only identified the impact of a single teacher on long-term outcomes. Are teachers' impacts additive over time? Do good teachers complement or substitute for each other across years? Similarly, it would be useful to go beyond the mean treatment effects that we have estimated here and determine whether some teachers are especially effective in improving lower-tail outcomes or producing stars.

Whether or not VA is ultimately used as a policy tool, our results show that parents should place great value on having their child in the classroom of a high value-added teacher. Consider a teacher whose true VA is one standard deviation above the mean who is contemplating leaving a school. Each child would gain approximately \$39,000 in total (undiscounted) lifetime earnings from having this teacher instead of the median teacher. With an annual discount rate of 5 percent, the parents of a classroom of average size should be willing to pay this teacher \$200,000 (\$7,000 per parent) to stay and teach their children during the next school year. Hence, the most important lesson of this study is that improving the quality of teaching—whether via the use of value-added metrics or other policy levers—is likely to have substantial economic and social benefits.

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.

[45] As we noted above, even in the low-stakes regime we study, some unusually high-VA teachers have test score impacts consistent with test manipulation. If such behavior becomes more prevalent when VA is used to evaluate teachers, the predictive content of VA as a measure of true teacher quality could be compromised.

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.

Boyd, Donald, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2008. "Who Leaves? Teacher Attrition and Student Achievement." National Bureau of Economic Research Working Paper 14022.

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

Carrell, Scott E. and James E. West. 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy* 118 (3): 409–32.

Cascio, Elizabeth U., and Douglas O. Staiger. 2012. "Knowledge, Tests, and Fadeout in Educational Interventions." National Bureau of Economic Research Working Paper 18038.

Chamberlain, Gary E. 2013. "Predictive Effects of Teachers and Schools on Test Scores, College Attendance, and Earnings." *Proceedings of the National Academy of Sciences* 110 (43): 17176–82.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126 (4): 1593–660.

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

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." *American Economic Review* 104 (9): 2593–632.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood: Dataset." *American Economic Review.* http://dx.doi.org/10.1257/aer.104.9.2633.

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor. 2008. "Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina." *Journal of Public Economics* 92 (5–6): 1352–70.

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*. Providence: Annenberg Institute for School Reform at Brown University.

Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1 (3): 111–34.

District of Columbia Public Schools. 2012. "IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel." http://www.nctq.org/docs/IMPACT.pdf (accessed June 2014).

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.

Gunderson, Morley K., and Philip Oreopoulos. 2010. "Returns to Education in Developed Countries." In *International Encyclopedia of Education.* Vol. 2, edited by Penelope Peterson, Eva Baker, and Barry McGaw, 298–304. Oxford: Elsevier Ltd.

Haider, Steven, and Gary Solon. 2006. "Life-Cycle Variation in the Association Between Current and Lifetime Earnings." *American Economic Review* 96 (4): 1308–20.

Hanushek, Eric A. 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." *American Economic Review* 61 (2): 280–8.

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.

Hanushek, Eric A. 2011. "The Economic Value of Higher Teacher Quality." *Economics of Education Review* 30 (3): 466–79.

Hanushek, Eric A., and Steven G. Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teaching Quality." *American Economic Review* 100 (2): 267–71.

Heckman, James J., Lena Malofeeva, Rodrigo Pinto, and Peter A. Savelyev. 2010a. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." Unpublished.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010b. "Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program." *Quantitative Economics* 1 (1): 1–46.

Internal Revenue Service (IRS). 2010. *Document 6961: Calendar Year Projections of Information and Withholding Documents for the United States and IRS Campuses 2010-2018*. Washington, DC: IRS Office of Research, Analysis, and Statistics.

Jackson, C. Kirabo. 2013 "Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina." National Bureau of Economic Research Working Paper 18624.

Jacob, Brian A. 2005. "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89 (5–6): 761–96.

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

Jacob, Brian A., and Steven D. Levitt. 2003. "Rotten Apples: An Investigation Of The Prevalence And Predictors Of Teacher Cheating." *The Quarterly Journal of Economics* 118 (3): 843–77.

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

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114 (2): 497–532.

Lockwood, J.R., and Daniel F. McCaffrey. 2009. "Exploring Student-Teacher Interactions in Longitudinal Achievement Data." *Education Finance and Policy* 4 (4): 439–67.

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

Neal, Derek A., and Diane Whitmore Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *Review of Economics and Statistics* 92 (2): 263–83.

Oreopoulos, Philip, and Uros Petronijevic. 2013. "Making College Worth It: A Review of Research on the Returns to Higher Education." National Bureau of Economic Research Working Paper 19053.

Oreopoulos, Philip, and Kjell G. Salvanes. 2010. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25 (1): 159–84.

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

Rothstein, Jesse. 2013. "Teacher Quality Policy When Supply Matters." http://eml.berkeley.edu/~jrothst/workingpapers/.

Staiger, Douglas O., and Jonah E. Rockoff. 2010. "Searching for Effective Teachers with Imperfect Information." *Journal of Economic Perspectives* 24 (3): 97–118.

Todd, Petra E., and Kenneth I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *The Economic Journal* 113 (485): F3–33.

US Census Bureau. 2010. "School Enrollment–Social and Economic Characteristics of Students: October 2008, Detailed." http://www.census.gov/hhes/school/data/cps/index.html.

This article has been cited by:

1. Majah-Leah V. Ravago, Claire Dennis S. Mapa. 2020. Awards and recognition: Do they matter in teachers' income trajectory?. *Studies in Educational Evaluation* **66**, 100901. [Crossref]
2. Louise Beuchert, Tine Louise Mundbjerg Eriksen, Morten Visby Krægpøth. 2020. The impact of standardized test feedback in math: Exploiting a natural experiment in 3rd grade. *Economics of Education Review* **77**, 102017. [Crossref]
3. Yeycol Leiva, Gabriel Pino. 2020. Analysis of the impact of school performance on income inequality in the long run: An application to Chilean municipalities. *Growth and Change* **27**. . [Crossref]
4. Dongwoo Kim. 2020. The relationship between violent crime and urban high school enrollment in the US. *Applied Economics Letters* **27**:12, 961-964. [Crossref]
5. Susan M. Kowalski, Joseph A. Taylor, Karen M. Askinas, Qian Wang, Qi Zhang, William P. Maddix, Elizabeth Tipton. 2020. Examining Factors Contributing to Variation in Effect Size Estimates of Teacher Outcomes from Studies of Science Teacher Professional Development. *Journal of Research on Educational Effectiveness* **13**:3, 430-458. [Crossref]
6. Zachary Griffen, Aaron Panofsky. 2020. VAM on trial: judging science in teacher evaluation lawsuits. *Journal of Cultural Economy* **13**:4, 444-460. [Crossref]
7. Saurabh A. Lall, Li-Wei Chen, Peter W. Roberts. 2020. Are we accelerating equity investment into impact-oriented ventures?. *World Development* **131**, 104952. [Crossref]
8. Serena Canaan. 2020. The long-run effects of reducing early school tracking. *Journal of Public Economics* **187**, 104206. [Crossref]
9. Se Woong Lee. 2020. Keeping an eye on the clock: the role of timing and teacher hiring in the education. *Educational Studies* **114**, 1-16. [Crossref]
10. James Chu, Guirong Li, Prashant Loyalka, Chengfang Liu, Leonardo Rosa, Yanyan Li. 2020. Stuck in Place? A Field Experiment on the Effects of Reputational Information on Student Evaluations. *Social Forces* **98**:4, 1578-1612. [Crossref]
11. Julie Cohen, Vivian Wong, Anandita Krishnamachari, Rebekah Berlin. 2020. Teacher Coaching in a Simulated Environment. *Educational Evaluation and Policy Analysis* **42**:2, 208-231. [Crossref]
12. Are Skeie Hermansen, Nicolai T Borgen, Arne Mastekaasa. 2020. Long-Term Trends in Adult Socio-Economic Resemblance between Former Schoolmates and Neighbouring Children. *European Sociological Review* **36**:3, 366-380. [Crossref]
13. 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]
14. David S. Knight. 2020. Accounting for Teacher Labor Markets and Student Segregation in Analyses of Teacher Quality Gaps. *Educational Researcher* 0013189X2092580. [Crossref]
15. Roddy J. Theobald, Dan D. Goldhaber, Trevor M. Gratz, Kristian L. Holden. 2020. High School English Language Arts Teachers and Postsecondary Outcomes for Students With and Without Disabilities. *Journal of Disability Policy Studies* 104420732091989. [Crossref]
16. PREETIKA JOSHI. 2020. Does Private Country-by-Country Reporting Deter Tax Avoidance and Income Shifting? Evidence from BEPS Action Item 13. *Journal of Accounting Research* **58**:2, 333-381. [Crossref]
17. 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]

18. Stephanie R. Cellini, Rajeev Darolia, Lesley J. Turner. 2020. Where Do Students Go When For-Profit Colleges Lose Federal Aid?. *American Economic Journal: Economic Policy* **12**:2, 46-83. [Abstract] [View PDF article] [PDF with links]
19. Anna Godøy, Ingrid Huitfeldt. 2020. Regional variation in health care utilization and mortality. *Journal of Health Economics* **71**, 102254. [Crossref]
20. Nicholas W. Papageorge, Seth Gershenson, Kyung Min Kang. 2020. Teacher Expectations Matter. *The Review of Economics and Statistics* **102**:2, 234-251. [Crossref]
21. Tenelle Porter, Karina Schumann, Diana Selmeczy, Kali Trzesniewski. 2020. Intellectual humility predicts mastery behaviors when learning. *Learning and Individual Differences* **80**, 101888. [Crossref]
22. Matthew A. Kraft. 2020. Interpreting Effect Sizes of Education Interventions. *Educational Researcher* **49**:4, 241-253. [Crossref]
23. David B. Reid. 2020. Making Sense of Teacher Evaluation Policies and Systems Based on Principals' Experience. *Leadership and Policy in Schools* **19**:2, 304-318. [Crossref]
24. Daniela Torre Gibney, Gary Henry. 2020. Who teaches English learners? A study of the quality, experience, and credentials of teachers of English learners in a new immigrant destination. *Teaching and Teacher Education* **90**, 102967. [Crossref]
25. Michael Bates. 2020. Public and Private Employer Learning: Evidence from the Adoption of Teacher Value Added. *Journal of Labor Economics* **38**:2, 375-420. [Crossref]
26. 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]
27. Emily Rauscher. 2020. Delayed Benefits: Effects of California School District Bond Elections on Achievement by Socioeconomic Status. *Sociology of Education* **93**:2, 110-131. [Crossref]
28. Matthew J. Hirshberg, Lisa Flook, Robert D. Enright, Richard J. Davidson. 2020. Integrating mindfulness and connection practices into preservice teacher education improves classroom practices. *Learning and Instruction* **66**, 101298. [Crossref]
29. 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]
30. Michael Crouch, Tuan D. Nguyen. 2020. Examining Teacher Characteristics, School Conditions, and Attrition Rates at the Intersection of School Choice and Rural Education. *Journal of School Choice* **30**, 1-27. [Crossref]
31. Julie Cohen, Susanna Loeb, Luke C. Miller, James H. Wyckoff. 2020. Policy Implementation, Principal Agency, and Strategic Action: Improving Teaching Effectiveness in New York City Middle Schools. *Educational Evaluation and Policy Analysis* **42**:1, 134-160. [Crossref]
32. Brendan Bartanen. 2020. Principal Quality and Student Attendance. *Educational Researcher* **49**:2, 101-113. [Crossref]
33. Jason M. Miller, Peter Youngs, Frank Perrone, Erin Grogan. 2020. Using Measures of Fit to Predict Beginning Teacher Retention. *The Elementary School Journal* **120**:3, 399-421. [Crossref]
34. Carolyn Abott, Vladimir Kogan, Stéphane Lavertu, Zachary Peskowitz. 2020. School district operational spending and student outcomes: Evidence from tax elections in seven states. *Journal of Public Economics* **183**, 104142. [Crossref]
35. 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]
36. 2020. Book Notes. *Harvard Educational Review* **90**:1, 145-158. [Crossref]
37. Christopher Belfield, Imran Rasul. 2020. Cognitive and Non-Cognitive Impacts of High-Ability Peers in Early Years*. *Fiscal Studies* **41**:1, 65-100. [Crossref]

38. David W Brown, Amanda E Kowalski, Ithai Z Lurie. 2020. Long-Term Impacts of Childhood Medicaid Expansions on Outcomes in Adulthood. *The Review of Economic Studies* **87**:2, 792-821. [Crossref]
39. 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]
40. 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]
41. 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]
42. John P. Papay, Eric S. Taylor, John H. Tyler, Mary E. Laski. 2020. Learning Job Skills from Colleagues at Work: Evidence from a Field Experiment Using Teacher Performance Data. *American Economic Journal: Economic Policy* **12**:1, 359-388. [Abstract] [View PDF article] [PDF with links]
43. Christine M. Dabbs. 2020. Restricting seniority as a factor in public school district layoffs: Analyzing the impact of state legislation on graduation rates. *Economics of Education Review* **74**, 101926. [Crossref]
44. 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]
45. Simon Calmar Andersen, Louise Beuchert, Helena Skyt Nielsen, Mette Kjærgaard Thomsen. 2020. The Effect of Teacher's Aides in the Classroom: Evidence from a Randomized Trial. *Journal of the European Economic Association* **18**:1, 469-505. [Crossref]
46. Yi Wei, Sen Zhou, Yunbo Liu. 2020. The draw of home: How does teacher's initial job placement relate to teacher mobility in rural China?. *PLOS ONE* **15**:1, e0227137. [Crossref]
47. Billy Wong, Yuan-Li Tiffany Chiu. 2020. University lecturers' construction of the 'ideal' undergraduate student. *Journal of Further and Higher Education* **44**:1, 54-68. [Crossref]
48. Evan Totty. 2020. High school value-added and college outcomes. *Education Economics* **28**:1, 67-95. [Crossref]
49. David B. Reid. 2020. Teachers' perceptions of how principals use new teacher evaluation systems. *Teachers and Teaching* **26**:1, 129-144. [Crossref]
50. Sidney Brown. How the Three R's Model (Relationships, Rigor, and Relevance) Addresses School Dropout Issues 94-109. [Crossref]
51. 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]
52. Sarena Goodman, Adam Isen. 2020. Un-Fortunate Sons: Effects of the Vietnam Draft Lottery on the Next Generation's Labor Market. *American Economic Journal: Applied Economics* **12**:1, 182-209. [Abstract] [View PDF article] [PDF with links]
53. David B. Reid. 2020. School principals acting as middle leaders implementing new teacher evaluation systems. *School Leadership & Management* **40**:1, 88-104. [Crossref]
54. Guido Schwerdt, Ludger Woessmann. Empirical methods in the economics of education 3-20. [Crossref]
55. Bjarne Strøm, Torberg Falch. The role of teacher quality in education production 307-319. [Crossref]
56. Eric A. Hanushek. Education production functions 161-170. [Crossref]
57. Jessalynn James, James Wyckoff. Teacher labor markets: An overview 355-370. [Crossref]

58. Martin Salm, Ansgar Wübker. 2020. Sources of regional variation in healthcare utilization in Germany. *Journal of Health Economics* **69**, 102271. [Crossref]
59. Shuqiong Lin, Wen Luo, Fuhui Tong, Beverly J. Irby, Rafael Lara Alecio, Linda Rodriguez, Selena Chapa. 2020. Data-based student learning objectives for teacher evaluation. *Cogent Education* **7**:1. . [Crossref]
60. Christian Michael Smith. 2020. In the Footsteps of Siblings: College Attendance Disparities and the Intragenerational Transmission of Educational Advantage. *Socius: Sociological Research for a Dynamic World* **6**, 237802312092163. [Crossref]
61. Andrew C. Johnston. 2020. Teacher Preferences, Working Conditions, and Compensation Structure. *SSRN Electronic Journal* . [Crossref]
62. Gabriel Lara Ibarra, David McKenzie, Claudia Ruiz-Ortega. 2019. Estimating Treatment Effects with Big Data When Take-up is Low: An Application to Financial Education. *The World Bank Economic Review* **7**. . [Crossref]
63. Allison W. Kenney. 2019. Negotiating Authority in the Ritual of the Public School Board Meeting. *Educational Administration Quarterly* **16**, 0013161X1989122. [Crossref]
64. 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]
65. Kathryn H. Anderson, Damir Esenaliev. 2019. Gender Earnings Inequality and Wage Policy: Teachers, Health Care, and Social Workers in Central Asia. *Comparative Economic Studies* **61**:4, 551-575. [Crossref]
66. Alberto Abadie, Maximilian Kasy. 2019. Choosing Among Regularized Estimators in Empirical Economics: The Risk of Machine Learning. *The Review of Economics and Statistics* **101**:5, 743-762. [Crossref]
67. Harrison J. Kell. 2019. Do Teachers' Personality Traits Predict Their Performance? A Comprehensive Review of the Empirical Literature From 1990 to 2018. *ETS Research Report Series* **2019**:1, 1-27. [Crossref]
68. Maria Liria Dacanay, Mark Lawrence Gale, William Bill Turnbull. Development and Validation of the New Teachers' Behavior Inventory 632-639. [Crossref]
69. Ron Shadbegian, Dennis Guignet, Heather Klemick, Linda Bui. 2019. Early childhood lead exposure and the persistence of educational consequences into adolescence. *Environmental Research* **178**, 108643. [Crossref]
70. Heather C. Hill, Charalambos Y. Charalambous, Mark J. Chin. 2019. Teacher Characteristics and Student Learning in Mathematics: A Comprehensive Assessment. *Educational Policy* **33**:7, 1103-1134. [Crossref]
71. Janet Penner-Williams, Eva I. Diaz, Diana Gonzales Worthen. 2019. Sustainability of teacher growth from professional development in culturally and linguistically responsive instructional practices. *Teaching and Teacher Education* **86**, 102891. [Crossref]
72. Eunice S. Han, Thomas N. Maloney. 2019. Teacher Unionization and Student Academic Performance: Looking beyond Collective Bargaining. *Labor Studies Journal* **57**, 0160449X1988337. [Crossref]
73. Laura K. Rogers, Sy Doan. 2019. Late to Class. *The Elementary School Journal* 000-000. [Crossref]
74. Ben Kelcey, Heather C. Hill, Mark J. Chin. 2019. Teacher mathematical knowledge, instructional quality, and student outcomes: a multilevel quantile mediation analysis. *School Effectiveness and School Improvement* **30**:4, 398-431. [Crossref]

75. Jose Eos Trinidad. 2019. Teacher Response Process to Bureaucratic Control: Individual and Group Dynamics Influencing Teacher Responses. *Leadership and Policy in Schools* **18**:4, 533-543. [Crossref]
76. Tim Kaiser, Lukas Menkhoff. 2019. Financial education in schools: A meta-analysis of experimental studies. *Economics of Education Review* 101930. [Crossref]
77. Ishuan Li. 2019. Book Review: Education and the commercial mindset , by Abrams, S. E. *The American Economist* **64**:2, 327-329. [Crossref]
78. Dan Goldhaber, Cory Koedel. 2019. Public Accountability and Nudges: The Effect of an Information Intervention on the Responsiveness of Teacher Education Programs to External Ratings. *American Educational Research Journal* **56**:5, 1557-1589. [Crossref]
79. 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]
80. Susanna Loeb, Michael S. Christian, Heather Hough, Robert H. Meyer, Andrew B. Rice, Martin R. West. 2019. School Differences in Social–Emotional Learning Gains: Findings From the First Large-Scale Panel Survey of Students. *Journal of Educational and Behavioral Statistics* **44**:5, 507-542. [Crossref]
81. Kathleen Lynch, Heather C. Hill, Kathryn Gonzalez, Cynthia Pollard. 2019. Strengthening STEM Instruction in Schools: Learning From Research. *Policy Insights from the Behavioral and Brain Sciences* **6**:2, 236-242. [Crossref]
82. Kathleen H. Corriveau, Marcus A. Winters. 2019. Trusting Your Teacher: Implications for Policy. *Policy Insights from the Behavioral and Brain Sciences* **6**:2, 123-129. [Crossref]
83. Bladimir Carrillo. 2019. Present Bias and Underinvestment in Education? Long-run Effects of Childhood Exposure to Booms in Colombia. *Journal of Labor Economics* . [Crossref]
84. Dania V. Francis, Angela C. M. de Oliveira, Carey Dimmitt. 2019. Do School Counselors Exhibit Bias in Recommending Students for Advanced Coursework?. *The B.E. Journal of Economic Analysis & Policy* **19**:4. . [Crossref]
85. Amelie Schiprowski. 2019. The Role of Caseworkers in Unemployment Insurance: Evidence from Unplanned Absences. *Journal of Labor Economics* . [Crossref]
86. Ozkan Eren. 2019. Teacher Incentives and Student Achievement: Evidence from an Advancement Program. *Journal of Policy Analysis and Management* **38**:4, 867-890. [Crossref]
87. 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]
88. Brendan Bartanen, Jason A. Grissom, Laura K. Rogers. 2019. The Impacts of Principal Turnover. *Educational Evaluation and Policy Analysis* **41**:3, 350-374. [Crossref]
89. Grieve Chelwa, Miquel Pellicer, Mashekwa Maboshe. 2019. Teacher Pay and Educational Outcomes: Evidence from the Rural Hardship Allowance in Zambia. *South African Journal of Economics* **87**:3, 255-282. [Crossref]
90. Shannon W. Anderson, Amanda Kimball. 2019. Evidence for the Feedback Role of Performance Measurement Systems. *Management Science* **65**:9, 4385-4406. [Crossref]
91. Billy Wong, Yuan-Li Tiffany Chiu. 2019. Exploring the concept of 'ideal' university student. *Studies in Higher Education* **36**, 1-12. [Crossref]
92. Corey A. DeAngelis. 2019. Divergences between effects on test scores and effects on non-cognitive skills. *Educational Review* **64**, 1-12. [Crossref]
93. W. Bentley MacLeod, Miguel Urquiola. 2019. Is Education Consumption or Investment? Implications for School Competition. *Annual Review of Economics* **11**:1, 563-589. [Crossref]

94. Michael F. Lovenheim, Alexander Willén. 2019. The Long-Run Effects of Teacher Collective Bargaining. *American Economic Journal: Economic Policy* **11**:3, 292-324. [Abstract] [View PDF article] [PDF with links]
95. Michela Carlana. 2019. Implicit Stereotypes: Evidence from Teachers' Gender Bias*. *The Quarterly Journal of Economics* **134**:3, 1163-1224. [Crossref]
96. Derek C. Briggs, Jessica L. Alzen. 2019. Making Inferences About Teacher Observation Scores Over Time. *Educational and Psychological Measurement* **79**:4, 636-664. [Crossref]
97. 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]
98. Paul Hanselman. 2019. Access to Effective Teachers and Economic and Racial Disparities in Opportunities to Learn. *The Sociological Quarterly* **60**:3, 498-534. [Crossref]
99. Prashant Loyalka, Anna Popova, Guirong Li, Zhaolei Shi. 2019. Does Teacher Training Actually Work? Evidence from a Large-Scale Randomized Evaluation of a National Teacher Training Program. *American Economic Journal: Applied Economics* **11**:3, 128-154. [Abstract] [View PDF article] [PDF with links]
100. John A. Williams, Laura C. Hart, Bob Algozzine. 2019. Perception vs. reality: edTPA perceptions and performance for teacher candidates of color and White candidates. *Teaching and Teacher Education* **83**, 120-133. [Crossref]
101. M. Niaz Asadullah, Md. Abdul Alim, M. Anowar Hossain. 2019. Enrolling girls without learning: Evidence from public schools in Afghanistan. *Development Policy Review* **37**:4, 486-503. [Crossref]
102. Matthew Davis, Blake Heller. 2019. No Excuses Charter Schools and College Enrollment: New Evidence from a High School Network in Chicago. *Education Finance and Policy* **14**:3, 414-440. [Crossref]
103. Heewon Jang, Sean F. Reardon. 2019. States as Sites of Educational (In)Equality: State Contexts and the Socioeconomic Achievement Gradient. *AERA Open* **5**:3, 233285841987245. [Crossref]
104. Jia Wu, Xiangdong Wei, Hongliang Zhang, Xiang Zhou. 2019. Elite schools, magnet classes, and academic performances: Regression-discontinuity evidence from China. *China Economic Review* **55**, 143-167. [Crossref]
105. 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]
106. Barbara Bruns, Isabel Harbaugh Macdonald, Ben Ross Schneider. 2019. The politics of quality reforms and the challenges for SDGs in education. *World Development* **118**, 27-38. [Crossref]
107. 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]
108. 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]
109. Xiaoxia Xie, Chien-Chung Huang, Yafan Chen, Feng Hao. 2019. Intelligent robots and rural children. *Children and Youth Services Review* **100**, 283-290. [Crossref]
110. Thomas Huizen, Lisa Dumhs, Janneke Plantenga. 2019. The Costs and Benefits of Investing in Universal Preschool: Evidence From a Spanish Reform. *Child Development* **90**:3. . [Crossref]
111. Dan Goldhaber, Vanessa Quince, Roddy Theobald. 2019. Teacher quality gaps in U.S. public schools: Trends, sources, and implications. *Phi Delta Kappan* **100**:8, 14-19. [Crossref]

112. Elizabeth Ross. 2019. Ensuring equitable access to great teachers: State policy priorities. *Phi Delta Kappan* **100**:8, 20-26. [Crossref]

113. David B Reid. 2019. Shared leadership: A comparative case study of two first year US principals' socialization around teacher evaluation policy. *Educational Management Administration & Leadership* **47**:3, 369-382. [Crossref]

114. 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]

115. Audrey Amrein-Beardsley, Kevin Close. 2019. Teacher-Level Value-Added Models on Trial: Empirical and Pragmatic Issues of Concern Across Five Court Cases. *Educational Policy* **17570**, 089590481984359. [Crossref]

116. Sule Alan, Teodora Boneva, Seda Ertac. 2019. Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit*. *The Quarterly Journal of Economics* **126**. . [Crossref]

117. Sy Doan, Jonathan D. Schweig, Kata Mihaly. 2019. The Consistency of Composite Ratings of Teacher Effectiveness: Evidence From New Mexico. *American Educational Research Journal* 000283121984136. [Crossref]

118. Griet Vanwynsberghe, Gudrun Vanlaar, Jan Van Damme, Bieke De Fraine. 2019. Long-term effects of first-grade teachers on students' achievement: a replication study. *School Effectiveness and School Improvement* **30**:2, 177-193. [Crossref]

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

120. Jason A. Grissom, Brendan Bartanen, Hajime Mitani. 2019. Principal Sorting and the Distribution of Principal Quality. *AERA Open* **5**:2, 233285841985009. [Crossref]

121. 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]

122. Yi Wei, Sen Zhou. 2019. Are Better Teachers More Likely To Move? Examining Teacher Mobility In Rural China. *The Asia-Pacific Education Researcher* **28**:2, 171-179. [Crossref]

123. Alejandra Mizala, Ben Schneider. 2019. Promoting quality education in Chile: the politics of reforming teacher careers. *Journal of Education Policy* **9**, 1-27. [Crossref]

124. 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]

125. 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]

126. Dan Goldhaber. 2019. Evidence-Based Teacher Preparation: Policy Context and What We Know. *Journal of Teacher Education* **70**:2, 90-101. [Crossref]

127. Lisa E. Kim, Verena Jörg, Robert M. Klassen. 2019. A Meta-Analysis of the Effects of Teacher Personality on Teacher Effectiveness and Burnout. *Educational Psychology Review* **31**:1, 163-195. [Crossref]

128. 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]

129. Eric J Johnson, Stephan Meier, Olivier Toubia. 2019. What's the Catch? Suspicion of Bank Motives and Sluggish Refinancing. *The Review of Financial Studies* **32**:2, 467-495. [Crossref]

130. Boukje Compen, Kristof De Witte, Wouter Schelfhout. 2019. The role of teacher professional development in financial literacy education: A systematic literature review. *Educational Research Review* **26**, 16-31. [Crossref]

131. Cory Koedel, Jiaxi Li, Matthew G. Springer, Li Tan. 2019. Teacher Performance Ratings and Professional Improvement. *Journal of Research on Educational Effectiveness* **12**:1, 90-115. [Crossref]
132. Nathan Burroughs, Jacqueline Gardner, Youngjun Lee, Siwen Guo, Israel Touitou, Kimberly Jansen, William Schmidt. A Review of the Literature on Teacher Effectiveness and Student Outcomes 7-17. [Crossref]
133. Ludger Woessmann. Ludger Woessmann Recommends "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood" by Raj Chetty, John N. Friedman, and Jonah E. Rockoff 157-159. [Crossref]
134. Şefika Şule Erçetin, Mehmet Sabir Çevik. A Study on Validity and Reliability of Teacher Efficacy Scale 411-430. [Crossref]
135. Michael D. Makowsky, Thomas Stratmann, Alex Tabarrok. 2019. To Serve and Collect: The Fiscal and Racial Determinants of Law Enforcement. *The Journal of Legal Studies* **48**:1, 189-216. [Crossref]
136. Corey DeAngelis, Patrick Wolf. 2019. Private School Choice and Character: More Evidence from Milwaukee. *SSRN Electronic Journal* . [Crossref]
137. Matias D. Cattaneo, Richard K. Crump, Max Farrell, Yingjie Feng. 2019. On Binscatter. *SSRN Electronic Journal* . [Crossref]
138. Steffen Andersen, Cristian Badarinza, Lu Liu, Julie Marx, Tarun Ramadorai. 2019. Reference Points in the Housing Market. *SSRN Electronic Journal* . [Crossref]
139. Nirav Mehta. 2019. Measuring quality for use in incentive schemes: The case of "shrinkage" estimators. *Quantitative Economics* **10**:4, 1537-1577. [Crossref]
140. Tirthatanmoy Das, Solomon W. Polachek. Microfoundations of Earnings Differences 9-76. [Crossref]
141. E. Jason Baron. 2019. Union Reform, Performance Pay, and New Teacher Supply: Evidence from Wisconsin's Act 10. *SSRN Electronic Journal* . [Crossref]
142. E. Jason Baron. 2019. School Spending and Student Outcomes: Evidence from Revenue Limit Elections in Wisconsin. *SSRN Electronic Journal* . [Crossref]
143. Sheldon Lewis Eakins. A School Model for Developing Access to Higher Education for African American 167-189. [Crossref]
144. 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]
145. Shujie Liu, Xianxuan Xu, James Stronge. 2018. The influences of teachers' perceptions of using student achievement data in evaluation and their self-efficacy on job satisfaction: evidence from China. *Asia Pacific Education Review* **19**:4, 493-509. [Crossref]
146. E. Jason Baron. 2018. The Effect of Teachers' Unions on Student Achievement in the Short Run: Evidence from Wisconsin's Act 10. *Economics of Education Review* **67**, 40-57. [Crossref]
147. Fritz Schiltz, Paolo Sestito, Tommaso Agasisti, Kristof De Witte. 2018. The added value of more accurate predictions for school rankings. *Economics of Education Review* **67**, 207-215. [Crossref]
148. Ryuichi Tanaka, Kazumi Ishizaki. 2018. Do teaching practices matter for students' academic achievement? A case of linguistic activity. *Journal of the Japanese and International Economies* **50**, 26-36. [Crossref]
149. Mette Gørtz, Eva Rye Johansen, Marianne Simonsen. 2018. Academic achievement and the gender composition of preschool staff. *Labour Economics* **55**, 241-258. [Crossref]
150. David C. Owens, Troy D. Sadler, Christopher D. Murakami, Chia-Lin Tsai. 2018. Teachers' views on and preferences for meeting their professional development needs in STEM. *School Science and Mathematics* **118**:8, 370-384. [Crossref]

151. Sule Alan, Seda Ertac, Ipek Mumcu. 2018. Gender Stereotypes in the Classroom and Effects on Achievement. *The Review of Economics and Statistics* **100**:5, 876-890. [Crossref]

152. James P. Spillane, Matthew Shirrell, Samrachana Adhikari. 2018. Constructing "Experts" Among Peers: Educational Infrastructure, Test Data, and Teachers' Interactions About Teaching. *Educational Evaluation and Policy Analysis* **40**:4, 586-612. [Crossref]

153. Mark Olofson, David Knight. 2018. Does the Middle School Model Make a Difference? Relating Measures of School Effectiveness to Recommended Best Practices. *Education Sciences* **8**:4, 160. [Crossref]

154. Rodney Hughes, Lauren Dahlin, Tara Tucci. 2018. An Investigation of a Multiple-Measures Teaching Evaluation System and Its Relationship With Students' College-Going Outcomes. *Educational Policy* 089590481881330. [Crossref]

155. Luke N. Condra, James D. Long, Andrew C. Shaver, Austin L. Wright. 2018. The Logic of Insurgent Electoral Violence. *American Economic Review* **108**:11, 3199-3231. [Abstract] [View PDF article] [PDF with links]

156. Jose M. Cordero, María Gil-Izquierdo. 2018. The effect of teaching strategies on student achievement: An analysis using TALIS-PISA-link. *Journal of Policy Modeling* **40**:6, 1313-1331. [Crossref]

157. Frances Woolley. 2018. The political economy of university education in Canada. *Canadian Journal of Economics/Revue canadienne d'économique* **51**:4, 1061-1087. [Crossref]

158. Haoying Sun, Lizhen Xu. 2018. Online Reviews and Collaborative Service Provision: A Signal-Jamming Model. *Production and Operations Management* **27**:11, 1960-1977. [Crossref]

159. Bruce Saddler, Kristie Asaro-Saddler, Tammy Ellis-Robinson, Matthew LaFave. 2018. Where Do They Come From and How Can We Find More? Recruiting Teacher Candidates During Lean Times. *Excelsior: Leadership in Teaching and Learning* **11**:1. . [Crossref]

160. . Building Human Capital 49-68. [Crossref]

161. Patrícia Martinková, Dan Goldhaber, Elena Erosheva. 2018. Disparities in ratings of internal and external applicants: A case for model-based inter-rater reliability. *PLOS ONE* **13**:10, e0203002. [Crossref]

162. Eric Chyn. 2018. Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children. *American Economic Review* **108**:10, 3028-3056. [Abstract] [View PDF article] [PDF with links]

163. 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]

164. Peter Bergman, Matthew J. Hill. 2018. The effects of making performance information public: Regression discontinuity evidence from Los Angeles teachers. *Economics of Education Review* **66**, 104-113. [Crossref]

165. 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]

166. M. Maria Glymour, Jennifer J. Manly. 2018. Compulsory Schooling Laws as quasi-experiments for the health effects of education: Reconsidering mechanisms to understand inconsistent results. *Social Science & Medicine* **214**, 67-69. [Crossref]

167. Eric C. Sun, Thomas R. Miller, Jasmin Moshfegh, Laurence C. Baker. 2018. Anesthesia Care Team Composition and Surgical Outcomes. *Anesthesiology* **129**:4, 700-709. [Crossref]

168. Rajeev Darolia, Cory Koedel. 2018. HIGH SCHOOLS AND STUDENTS' INITIAL COLLEGES AND MAJORS. *Contemporary Economic Policy* **36**:4, 692-710. [Crossref]

169. Prakarsh Singh, William A. Masters. 2018. Performance bonuses in the public sector: Winner-take-all prizes versus proportional payments to reduce child malnutrition in India. *Journal of Development Economics* 102295. [Crossref]

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

171. Katharine O. Strunk, Dan Goldhaber, David S. Knight, Nate Brown. 2018. Are There Hidden Costs Associated With Conducting Layoffs? The Impact of Reduction-in-Force and Layoff Notices on Teacher Effectiveness. *Journal of Policy Analysis and Management* **37**:4, 755-782. [Crossref]

172. Deven Carlson, Stéphane Lavertu. 2018. School Improvement Grants in Ohio: Effects on Student Achievement and School Administration. *Educational Evaluation and Policy Analysis* **40**:3, 287-315. [Crossref]

173. Kirabo Jackson, Alexey Makarin. 2018. Can Online Off-the-Shelf Lessons Improve Student Outcomes? Evidence from a Field Experiment. *American Economic Journal: Economic Policy* **10**:3, 226-254. [Abstract] [View PDF article] [PDF with links]

174. James Cowan, Dan Goldhaber. 2018. Do bonuses affect teacher staffing and student achievement in high poverty schools? Evidence from an incentive for national board certified teachers in Washington State. *Economics of Education Review* **65**, 138-152. [Crossref]

175. 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]

176. Nolan Kopkin, Mirinda L Martin, Danielle Hollar. 2018. Improvements in standardised test scores from a multi-component nutrition and healthy living intervention in a US elementary-school setting. *Health Education Journal* **77**:5, 527-541. [Crossref]

177. 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]

178. Matthew A. Kraft, David Blazar, Dylan Hogan. 2018. The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence. *Review of Educational Research* **88**:4, 547-588. [Crossref]

179. Rebecca Allen, Sam Sims. 2018. Do pupils from low-income families get low-quality teachers? Indirect evidence from English schools. *Oxford Review of Education* **44**:4, 441-458. [Crossref]

180. 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]

181. 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]

182. 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]

183. Tuan D. Nguyen, Christopher Redding. 2018. Changes in the Demographics, Qualifications, and Turnover of American STEM Teachers, 1988–2012. *AERA Open* **4**:3, 233285841880279. [Crossref]

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

185. Paul T. von Hippel, Laura Bellows. 2018. How much does teacher quality vary across teacher preparation programs? Reanalyses from six states. *Economics of Education Review* **64**, 298-312. [Crossref]

186. 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]

187. Mette Trier Damgaard, Helena Skyt Nielsen. 2018. Nudging in education. *Economics of Education Review* **64**, 313-342. [Crossref]
188. 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]
189. 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]
190. NaYoung Hwang. 2018. Suspensions and Achievement: Varying Links by Type, Frequency, and Subgroup. *Educational Researcher* 0013189X1877957. [Crossref]
191. 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]
192. David B. Reid. 2018. How schools train principals to use new teacher evaluation systems: A comparative case study of charter school and traditional public school principals. *Journal of School Choice* **12**:2, 237-253. [Crossref]
193. 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]
194. Barbara Masi. 2018. A ticket to ride: The unintended consequences of school transport subsidies. *Economics of Education Review* **63**, 100-115. [Crossref]
195. Irina Horoi, Moiz Bhai. 2018. NEW EVIDENCE ON NATIONAL BOARD CERTIFICATION AS A SIGNAL OF TEACHER QUALITY. *Economic Inquiry* **56**:2, 1185-1201. [Crossref]
196. Martha Cecilia Bottia, Roslyn Arlin Mickelson, Jason Giersch, Elizabeth Stearns, Stephanie Moller. 2018. The role of high school racial composition and opportunities to learn in students' STEM college participation. *Journal of Research in Science Teaching* **55**:3, 446-476. [Crossref]
197. Anna J. Egalite, Brian Kisida. 2018. The Effects of Teacher Match on Students' Academic Perceptions and Attitudes. *Educational Evaluation and Policy Analysis* **40**:1, 59-81. [Crossref]
198. 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]
199. Ayesha K. Hashim, Katharine O. Strunk, Julie A. Marsh. 2018. The new school advantage? Examining the effects of strategic new school openings on student achievement. *Economics of Education Review* **62**, 254-266. [Crossref]
200. Jamie L. Taxer, Anne C. Frenzel. 2018. Inauthentic expressions of enthusiasm: Exploring the cost of emotional dissonance in teachers. *Learning and Instruction* **53**, 74-88. [Crossref]
201. Sam Sims, Rebecca Allen. 2018. Identifying Schools With High Usage and High Loss of Newly Qualified Teachers. *National Institute Economic Review* **243**:1, R27-R36. [Crossref]
202. 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]
203. 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]
204. Li Feng, Tim R. Sass. 2018. The Impact of Incentives to Recruit and Retain Teachers in "Hard-to-Staff" Subjects. *Journal of Policy Analysis and Management* **37**:1, 112-135. [Crossref]
205. John Ewing. Who Are the Experts? 173-181. [Crossref]

206. Stephen L. Morgan, Daniel T. Shackelford. School and Teacher Effects 513-534. [Crossref]
207. Ethan L. Hutt, Jessica Gottlieb, Julia J. Cohen. 2018. Diffusion in a vacuum: edTPA, legitimacy, and the rhetoric of teacher professionalization. *Teaching and Teacher Education* **69**, 52-61. [Crossref]
208. Peter Hull. 2018. Estimating Hospital Quality with Quasi-Experimental Data. *SSRN Electronic Journal* . [Crossref]
209. Richard DiSalvo. 2018. Grade Configurations and Student Test Performance: Evidence from Recent National Data. *SSRN Electronic Journal* . [Crossref]
210. Shannon W. Anderson, Amanda Kimball. 2018. Evidence for the Feedback Role of Performance Measurement Systems. *SSRN Electronic Journal* . [Crossref]
211. Federico Gonzalez. 2018. Impactos De Corto Plazo Del Programa Extracurricular De Refuerzo Escolar 'Con Las Manos' En Un Colegio De Bogott (The Short-Term Effects of the Extracurricular Program Con Las Manos in a School in Bogott, Colombia). *SSRN Electronic Journal* . [Crossref]
212. Rebecca DizonnRoss. 2018. How Does School Accountability Affect Teachers? Evidence from New York City. *SSRN Electronic Journal* . [Crossref]
213. Marcus A. Winters. 2018. The Costs and Benefits of Test-Based Promotion. *SSRN Electronic Journal* . [Crossref]
214. Richard Murphy, Felix Weinhardt, Gill Wyness. 2018. Who Teaches the Teacher? A RCT of Peer-to-Peer Observation and Feedback in 181 Schools. *SSRN Electronic Journal* . [Crossref]
215. Prakarsh Singh, William A. Masters. 2018. Performance Bonuses in the Public Sector: Winner-Take-All Prizes Versus Proportional Payments to Reduce Child Malnutrition in India. *SSRN Electronic Journal* . [Crossref]
216. Corey DeAngelis. 2018. Divergences between Effects on Test Scores and Effects on Non-Cognitive Skills. *SSRN Electronic Journal* . [Crossref]
217. Jason Ward. 2018. Licensure Requirements and Regional Equilibrium in Teacher Labor Markets. *SSRN Electronic Journal* . [Crossref]
218. My Nguyen. 2018. The Relationship between Race-Congruent Students and Teachers: Does Racial Discrimination Exist?. *SSRN Electronic Journal* . [Crossref]
219. Alejandro Ome, Alicia Menendez, Huyen Elise Le. 2017. Improving teaching quality through training: Evidence from the Caucasus. *Economics of Education Review* **61**, 1-8. [Crossref]
220. Dario Sansone. 2017. Why does teacher gender matter?. *Economics of Education Review* **61**, 9-18. [Crossref]
221. F. Chris Curran. 2017. Income-Based Disparities in Early Elementary School Science Achievement. *The Elementary School Journal* **118**:2, 207-231. [Crossref]
222. 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]
223. 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]
224. Prita Nurmalia Kusumawardhani. 2017. Does teacher certification program lead to better quality teachers? Evidence from Indonesia. *Education Economics* **25**:6, 590-618. [Crossref]
225. Joshua Hyman. 2017. Does Money Matter in the Long Run? Effects of School Spending on Educational Attainment. *American Economic Journal: Economic Policy* **9**:4, 256-280. [Abstract] [View PDF article] [PDF with links]

226. 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]
227. Alejandra Mizala, Florencia Torche. 2017. Means-Tested School Vouchers and Educational Achievement: Evidence from Chile's Universal Voucher System. *The ANNALS of the American Academy of Political and Social Science* **674**:1, 163-183. [Crossref]
228. Margot I. Jackson, Susan L. Moffitt. 2017. The State of Unequal Educational Opportunity: Conclusion to the Special Issue on the Coleman Report 50 Years Later. *The ANNALS of the American Academy of Political and Social Science* **674**:1, 281-285. [Crossref]
229. 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]
230. Yyannú Cruz-Aguayo, Pablo Ibarrarán, Norbert Schady. 2017. Do tests applied to teachers predict their effectiveness?. *Economics Letters* **159**, 108-111. [Crossref]
231. Santiago Pereda-Fernández. 2017. Social Spillovers in the Classroom: Identification, Estimation and Policy Analysis. *Economica* **84**:336, 712-747. [Crossref]
232. 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]
233. 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]
234. Tyler J. VanderWeele. 2017. On the promotion of human flourishing. *Proceedings of the National Academy of Sciences* **114**:31, 8148-8156. [Crossref]
235. David Lazer, Jason Radford. 2017. Data ex Machina: Introduction to Big Data. *Annual Review of Sociology* **43**:1, 19-39. [Crossref]
236. Deven Carlson, Matthew M. Chingos, David E. Campbell. 2017. The Effect of Private School Vouchers on Political Participation: Experimental Evidence From New York City. *Journal of Research on Educational Effectiveness* **10**:3, 545-569. [Crossref]
237. Andrew S. Griffen, Petra E. Todd. 2017. Assessing the Performance of Nonexperimental Estimators for Evaluating Head Start. *Journal of Labor Economics* **35**:S1, S7-S63. [Crossref]
238. Ahmed Elhakeem, Rebecca Hardy, David Bann, Rishi Caleyachetty, Theodore D Cosco, Richard PG Hayhoe, Stella G Muthuri, Rebecca Wilson, Rachel Cooper. 2017. Intergenerational social mobility and leisure-time physical activity in adulthood: a systematic review. *Journal of Epidemiology and Community Health* **71**:7, 673-680. [Crossref]
239. Li Feng, Tim R. Sass. 2017. Teacher Quality and Teacher Mobility. *Education Finance and Policy* **12**:3, 396-418. [Crossref]
240. Halil Dundar, Benoît Millot, Michelle Riboud, Mari Shojo, Harsha Aturupane, Sangeeta Goyal, Dhushyanth Raju. Primary and Secondary Education: The Quality Challenge 81-121. [Crossref]
241. Pablo A. Acosta, Rita Almeida, Thomas Gindling, Christine Lao Peña. Coverage and Targeting 37-67. [Crossref]
242. Jesse Rothstein. 2017. Measuring the Impacts of Teachers: Comment. *American Economic Review* **107**:6, 1656-1684. [Abstract] [View PDF article] [PDF with links]
243. 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]
244. 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]

245. 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]
246. Galit Shmueli. 2017. Research Dilemmas with Behavioral Big Data. *Big Data* **5**:2, 98-119. [Crossref]
247. Ciro Avitabile. The Rapid Expansion of Higher Education in the New Century 47-75. [Crossref]
248. 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]
249. Joana Monteiro, Rudi Rocha. 2017. Drug Battles and School Achievement: Evidence from Rio de Janeiro's Favelas. *The Review of Economics and Statistics* **99**:2, 213-228. [Crossref]
250. 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]
251. David Blazar, David Braslow, Charalambos Y. Charalambous, Heather C. Hill. 2017. Attending to General and Mathematics-Specific Dimensions of Teaching: Exploring Factors Across Two Observation Instruments. *Educational Assessment* **22**:2, 71-94. [Crossref]
252. Tatiana Melguizo, Gema Zamarro, Tatiana Velasco, Fabio J. Sanchez. 2017. The Methodological Challenges of Measuring Student Learning, Degree Attainment, and Early Labor Market Outcomes in Higher Education. *Journal of Research on Educational Effectiveness* **10**:2, 424-448. [Crossref]
253. Orazio Attanasio, Costas Meghir, Emily Nix, Francesca Salvati. 2017. Human capital growth and poverty: Evidence from Ethiopia and Peru. *Review of Economic Dynamics* **25**, 234-259. [Crossref]
254. 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]
255. Cory Koedel, Jiaxi Li, Matthew G. Springer, Li Tan. 2017. The Impact of Performance Ratings on Job Satisfaction for Public School Teachers. *American Educational Research Journal* **54**:2, 241-278. [Crossref]
256. Dan Goldhaber, John M. Krieg, Roddy Theobald. 2017. Does the Match Matter? Exploring Whether Student Teaching Experiences Affect Teacher Effectiveness. *American Educational Research Journal* **54**:2, 325-359. [Crossref]
257. Sarah K. Burns, James P. Ziliak. 2017. Identifying the Elasticity of Taxable Income. *The Economic Journal* **127**:600, 297-329. [Crossref]
258. Cory Koedel, P. Brett Xiang. 2017. Pension Enhancements and the Retention of Public Employees. *ILR Review* **70**:2, 519-551. [Crossref]
259. Dhushyanth Raju. 2017. Public School Teacher Management in Sri Lanka. *South Asia Economic Journal* **18**:1, 39-63. [Crossref]
260. Melinda Adnot, Thomas Dee, Veronica Katz, James Wyckoff. 2017. Teacher Turnover, Teacher Quality, and Student Achievement in DCPS. *Educational Evaluation and Policy Analysis* **39**:1, 54-76. [Crossref]
261. 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]
262. Jinjing Li, Riyana Miranti, Yogi Vidyattama. 2017. What matters in education: a decomposition of educational outcomes with multiple measures. *Educational Research and Evaluation* **23**:1-2, 3-25. [Crossref]
263. 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]

264. Kimberlee C. Everson. 2017. Value-Added Modeling and Educational Accountability. *Review of Educational Research* **87**:1, 35-70. [Crossref]
265. Drew Bailey, Greg J. Duncan, Candice L. Odgers, Winnie Yu. 2017. Persistence and Fadeout in the Impacts of Child and Adolescent Interventions. *Journal of Research on Educational Effectiveness* **10**:1, 7-39. [Crossref]
266. 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]
267. Henry M. Levin. Empirical Puzzles on Effective Teachers: U.S. Research 189-204. [Crossref]
268. Dominik Becker, Klaus Birkelbach. Bildungsungleichheit durch Schul- und Schulklasseneffekte 179-210. [Crossref]
269. R.G. Fryer. The Production of Human Capital in Developed Countries 95-322. [Crossref]
270. Pelin Akyol, Kala Krishna. 2017. Preferences, selection, and value added: A structural approach. *European Economic Review* **91**, 89-117. [Crossref]
271. Margaret Brehm, Scott A. Imberman, Michael F. Lovenheim. 2017. Achievement effects of individual performance incentives in a teacher merit pay tournament. *Labour Economics* **44**, 133-150. [Crossref]
272. 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]
273. Teevrat Garg, Maulik Jagnani, Vis Taraz. 2017. Human Capital Costs of Climate Change: Evidence from Test Scores in India. *SSRN Electronic Journal* . [Crossref]
274. Barbara Biasi. 2017. Unions, Salaries, and the Market for Teachers: Evidence from Wisconsin. *SSRN Electronic Journal* . [Crossref]
275. E. Jason Baron. 2017. The Effect of Teachers' Unions on Student Achievement: Evidence from Wisconsin's Act 10. *SSRN Electronic Journal* . [Crossref]
276. Irina Horoi, Moiz Bhai. 2017. New Evidence on National Board Certification as a Signal of Teacher Quality. *SSRN Electronic Journal* . [Crossref]
277. Jan Feld, Nicolas Salamanca, Ulf ZZlitz. 2017. Students Are Almost as Effective as Professors in University Teaching. *SSRN Electronic Journal* . [Crossref]
278. Johannes S. Kunz, Kevin E. Staub, Rainer Winkelmann. 2017. Estimating Fixed Effects: Perfect Prediction and Bias in Binary Response Panel Models, with an Application to the Hospital Readmissions Reduction Program. *SSRN Electronic Journal* . [Crossref]
279. Evan Totty. 2017. High School Value-Added and Labor Market Outcomes. *SSRN Electronic Journal* . [Crossref]
280. Kumar Aniket. 2017. Man Versus Machine: Automation, Market Structure and Skills. *SSRN Electronic Journal* . [Crossref]
281. Rebecca Jorgensen, Charles C. Moul. 2017. Consequences of Wisconsin's Act 10: Teacher Compensation and Employment. *SSRN Electronic Journal* . [Crossref]
282. Paul T. von Hippel, Laura Bellows. 2017. How Much Does Teacher Quality Vary Across Teacher Preparation Programs? Reanalyzing Estimates from 6 States. *SSRN Electronic Journal* . [Crossref]
283. Bart H. H. Golsteyn, Stan Vermeulen, Inge de Wolf. 2016. Teacher Literacy and Numeracy Skills: International Evidence from PIAAC and ALL. *De Economist* **164**:4, 365-389. [Crossref]
284. 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]

285. Han Feng, Jiayao Li. 2016. Head teachers, peer effects, and student achievement. *China Economic Review* **41**, 268-283. [Crossref]
286. 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]
287. David J. Deming, Sarah Cohodes, Jennifer Jennings, Christopher Jencks. 2016. School Accountability, Postsecondary Attainment, and Earnings. *Review of Economics and Statistics* **98**:5, 848-862. [Crossref]
288. Laila S. Lomibao. 2016. Enhancing mathematics teachers' quality through Lesson Study. *SpringerPlus* **5**:1. . [Crossref]
289. 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]
290. John Garen. 2016. Assessing the literature on school reform from an entrepreneurship perspective. *Journal of Entrepreneurship and Public Policy* **5**:3, 383-403. [Crossref]
291. Stéphane Lavertu. 2016. We All Need Help: "Big Data" and the Mismeasure of Public Administration. *Public Administration Review* **76**:6, 864-872. [Crossref]
292. Amy Finkelstein, Matthew Gentzkow, Heidi Williams. 2016. Sources of Geographic Variation in Health Care: Evidence From Patient Migration*. *The Quarterly Journal of Economics* **131**:4, 1681-1726. [Crossref]
293. 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]
294. 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]
295. John P. Papay, Matthew A. Kraft. 2016. The Productivity Costs of Inefficient Hiring Practices: Evidence From Late Teacher Hiring. *Journal of Policy Analysis and Management* **35**:4, 791-817. [Crossref]
296. Deven Carlson, Stéphane Lavertu. 2016. Charter school closure and student achievement: Evidence from Ohio. *Journal of Urban Economics* **95**, 31-48. [Crossref]
297. Katharine O. Strunk, Julie A. Marsh, Ayesha K. Hashim, Susan Bush-Mecenas. 2016. Innovation and a Return to the Status Quo. *Educational Evaluation and Policy Analysis* **38**:3, 549-577. [Crossref]
298. Dan Goldhaber, Katharine O. Strunk, Nate Brown, David S. Knight. 2016. Lessons Learned From the Great Recession. *Educational Evaluation and Policy Analysis* **38**:3, 517-548. [Crossref]
299. 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]
300. Julie Cohen, Dan Goldhaber. 2016. Building a More Complete Understanding of Teacher Evaluation Using Classroom Observations. *Educational Researcher* **45**:6, 378-387. [Crossref]
301. 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]
302. 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]
303. Nathaniel G. Hilger. 2016. Parental Job Loss and Children's Long-Term Outcomes: Evidence from 7 Million Fathers' Layoffs. *American Economic Journal: Applied Economics* **8**:3, 247-283. [Abstract] [View PDF article] [PDF with links]

304. 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]

305. 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]

306. Joseph Engelberg, Raymond Fisman, Jay C. Hartzell, Christopher A. Parsons. 2016. Human Capital and the Supply of Religion. *Review of Economics and Statistics* **98**:3, 415-427. [Crossref]

307. Nicola Fuchs-Schündeln, Paolo Masella. 2016. Long-Lasting Effects of Socialist Education. *Review of Economics and Statistics* **98**:3, 428-441. [Crossref]

308. Tim R. Sass, Ron W. Zimmer, Brian P. Gill, T. Kevin Booker. 2016. Charter High Schools' Effects on Long-Term Attainment and Earnings. *Journal of Policy Analysis and Management* **35**:3, 683-706. [Crossref]

309. Seth Gershenson, Stephen B. Holt, Nicholas W. Papageorge. 2016. Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review* **52**, 209-224. [Crossref]

310. Matthew G. Springer, Walker A. Swain, Luis A. Rodriguez. 2016. Effective Teacher Retention Bonuses. *Educational Evaluation and Policy Analysis* **38**:2, 199-221. [Crossref]

311. Raj Chetty, Nathaniel Hendren, Lawrence F. Katz. 2016. The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review* **106**:4, 855-902. [Abstract] [View PDF article] [PDF with links]

312. Nolan G. Pope. 2016. How the Time of Day Affects Productivity: Evidence from School Schedules. *Review of Economics and Statistics* **98**:1, 1-11. [Crossref]

313. Dan Goldhaber, Joe Walch. 2016. Teacher tenure. *Phi Delta Kappan* **97**:6, 8-15. [Crossref]

314. C. Kirabo Jackson, Rucker C. Johnson, Claudia Persico. 2016. The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms *. *The Quarterly Journal of Economics* **131**:1, 157-218. [Crossref]

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

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

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

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

319. 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]

320. DAN GOLDHABER, CYRUS GROUT. 2016. Which plan to choose? The determinants of pension system choice for public school teachers. *Journal of Pension Economics and Finance* **15**:1, 30-54. [Crossref]

321. Sean Corcoran. Section I Discussion: How Do Educators Use Student Growth Measures in Practice? 153-165. [Crossref]

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

323. Josh Kinsler. 2016. Teacher Complementarities in Test Score Production: Evidence from Primary School. *Journal of Labor Economics* **34**:1, 29-61. [Crossref]

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

325. 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]

326. Lindsay Fox. 2016. Seeing Potential. *AERA Open* **2**:1, 233285841562375. [Crossref]

327. 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]

328. Galit Shmueli. 2016. Analyzing Behavioral Big Data: Methodological, Practical, Ethical, and Moral Issues. *SSRN Electronic Journal* . [Crossref]

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

330. Albert Cheng, Gema Zamarro. 2016. Measuring Teacher Conscientiousness and its Impact on Students: Insight from the Measures of Effective Teaching Longitudinal Database. *SSRN Electronic Journal* . [Crossref]

331. 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]

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

333. Joel Mittleman. 2016. What's in a Match? Disentangling the Significance of Teacher Race/Ethnicity. *SSRN Electronic Journal* . [Crossref]

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

335. Galit Shmueli. 2016. Beauty Is in the Eye of the (Data) Beholder: Researchers' Dilemmas Using Behavioral Big Data. *SSRN Electronic Journal* . [Crossref]

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

337. Thomas Schwab. 2016. Spillover from the Haven: Cross-Border Externalities of Patent Box Regimes Within Multinational Firms. *SSRN Electronic Journal* . [Crossref]

338. 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]

339. Zeyu Xu, Umut Özek, Michael Hansen. 2015. Teacher Performance Trajectories in High- and Lower-Poverty Schools. *Educational Evaluation and Policy Analysis* **37**:4, 458-477. [Crossref]

340. Mehtabul Azam, Geeta Gandhi Kingdon. 2015. Assessing teacher quality in India. *Journal of Development Economics* **117**, 74-83. [Crossref]

341. 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]

342. David Blazar. 2015. Effective teaching in elementary mathematics: Identifying classroom practices that support student achievement. *Economics of Education Review* **48**, 16-29. [Crossref]

343. Cory Koedel, Eric Parsons, Michael Podgursky, Mark Ehlert. 2015. Teacher Preparation Programs and Teacher Quality: Are There Real Differences Across Programs?. *Education Finance and Policy* **10**:4, 508-534. [Crossref]

344. Allison Atteberry, Susanna Loeb, James Wyckoff. 2015. Do First Impressions Matter? Predicting Early Career Teacher Effectiveness. *AERA Open* **1**:4, 233285841560783. [Crossref]

345. Chandra L. Muller. 2015. Measuring School Contexts. *AERA Open* **1**:4, 233285841561305. [Crossref]

346. Christian P. Wilkens, Jennifer Randhare Ashton, Diane M. Maurer, Shelly Smith. 2015. Some of This Is Not Your Fault. *Schools* **12**:2, 329-341. [Crossref]

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

348. Matthew D. Hendricks. 2015. Towards an optimal teacher salary schedule: Designing base salary to attract and retain effective teachers. *Economics of Education Review* **47**, 143-167. [Crossref]

349. Jane Hannaway. 2015. Shifting Boundaries and Shady Borders: A Call for Research on the Political Economy of Education Reform. *Education Finance and Policy* **10**:3, 301-313. [Crossref]

350. 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]

351. Cary Lou, Adam Thomas. 2015. The Relationship Between Academic Achievement And Nonmarital Teenage Childbearing: Evidence from the Panel Study of Income Dynamics. *Perspectives on Sexual and Reproductive Health* **47**:2, 91-98. [Crossref]

352. Cary Lou, Adam Thomas. 2015. The Relationship Between Academic Achievement And Nonmarital Teenage Childbearing: Evidence from the Panel Study of Income Dynamics. *Perspectives on Sexual and Reproductive Health* **50**, n/a-n/a. [Crossref]

353. Susanna Loeb, Luke C. Miller, James Wyckoff. 2015. Performance Screens for School Improvement. *Educational Researcher* **44**:4, 199-212. [Crossref]

354. Dan Goldhaber. 2015. Exploring the Potential of Value-Added Performance Measures to Affect the Quality of the Teacher Workforce. *Educational Researcher* **44**:2, 87-95. [Crossref]

355. Eric Parsons, Cory Koedel, Michael Podgursky, Mark Ehlert, P. Brett Xiang. 2015. Incorporating End-of-Course Exam Timing Into Educational Performance Evaluations. *Journal of Research on Educational Effectiveness* **8**:1, 130-147. [Crossref]

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

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

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

359. Haoying Sun, Lizhen Xu. 2015. Online Reviews and Collaborative Service Provision: A Signal-Jamming Model. *SSRN Electronic Journal* . [Crossref]

360. Matthew Davis, Blake Heller. 2015. Do 'No Excuses' Charter Schools Raise More than Test Scores? College-Going Impacts of a High School Network in Chicago. *SSRN Electronic Journal* . [Crossref]

361. Michael Callen, Saad Gulzar, Syed Ali Hasanain, Muhammad Yasir Khan, Arman Rezaee. 2015. Personalities and Public Sector Performance: Evidence from a Health Experiment in Pakistan. *SSRN Electronic Journal* . [Crossref]

362. Simon Calmar Andersen, Louise Beuchert, Helena Skyt Nielsen, Mette Kjjrgaard Thomsen. 2015. The Effect of Teacher's Aides in the Classroom: Evidence from a Randomized Trial. *SSRN Electronic Journal* . [Crossref]
363. John T. Dalton, Tin Cheuk Leung. 2015. Being Bad by Being Good: Captains, Managerial Ability, and the Slave Trade. *SSRN Electronic Journal* . [Crossref]
364. Seth Gershenson, Stephen Holt, Nicholas Papageorge. 2015. Who Believes in Me? The Effect of Student-Teacher Demographic Match on Teacher Expectations. *SSRN Electronic Journal* . [Crossref]
365. Tatiana Melguizo, Gema Zamarro, Tatiana Velasco, Fabio SSnchez. 2015. How Can We Accurately Measure Whether Students are Gaining Relevant Outcomes in Higher Education?. *SSRN Electronic Journal* . [Crossref]
366. Jane Dokko, Geng Li, Jessica Hayes. 2015. Credit Scores and Committed Relationships. *SSRN Electronic Journal* . [Crossref]
367. Felipe Goncalves. 2015. The Effects of School Construction on Student and District Outcomes: Evidence from a State-Funded Program in Ohio. *SSRN Electronic Journal* . [Crossref]
368. Jonathan T. Rothwell. 2015. Defining Skilled Technical Work. *SSRN Electronic Journal* . [Crossref]
369. Lucia Rizzica. 2015. The Use of Fixed-Term Contracts and the (Adverse) Selection of Public Sector Workers. *SSRN Electronic Journal* . [Crossref]
370. Barbara Biasi. 2015. School Finance Equalization and Intergenerational Income Mobility: Does Equal Spending Lead to Equal Opportunities?. *SSRN Electronic Journal* . [Crossref]
371. L. Einav, J. Levin. 2014. Economics in the age of big data. *Science* **346**:6210, 1243089-1243089. [Crossref]
372. Raj Chetty, Nathaniel Hendren, Patrick Kline, Emmanuel Saez. 2014. Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *. *The Quarterly Journal of Economics* **129**:4, 1553-1623. [Crossref]
373. Raj Chetty, John N. Friedman, Jonah E. Rockoff. 2014. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review* **104**:9, 2593-2632. [Abstract] [View PDF article] [PDF with links]
374. 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]
375. Peter Hinrichs. 2014. What Kind of Teachers are Schools Looking for? Evidence from a Randomized Field Experiment. *SSRN Electronic Journal* . [Crossref]
376. Massimo G. Colombo, Samuele Murtinu. 2014. Venture Capital Investments in Europe and Portfolio Firms' Economic Performance: Independent versus Corporate Investors. *SSRN Electronic Journal* . [Crossref]
377. Michela Braga, Marco Paccagnella, MIchele Pellizzari. 2014. The Academic and Labor Market Returns of University Professors. *SSRN Electronic Journal* . [Crossref]
378. Matthew D. Hendricks. 2013. Towards an Optimal Teacher Salary Schedule: Designing Base Salary to Attract and Retain Effective Teachers. *SSRN Electronic Journal* . [Crossref]
379. E. Glen Weyl. 2007. Psychic Income, Taxes and the Allocation of Talent. *SSRN Electronic Journal* . [Crossref]