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

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

Online Appendix

Online Appendix A: Structural Interpretation of Reduced-Form Parameters

This appendix formalizes how the reduced-form parameters we estimate should be interpreted in a stylized dynamic model of the education production function. The model we outline below follows previous work (e.g., Todd and Wolpin 2003), except that we focus exclusively on the role of teachers, abstracting from other inputs to the education production function, such as peers or parental investment.

Dynamic Model: Setup

The model is characterized by a specification for scores, a specification for earnings (or other adult outcomes), and a rule that governs student and teacher assignment to classrooms.

Classroom Assignment Rule. School principals assign student i in school year t to a classroom $c = c(i,t)$ based on observed and unobserved determinants of student achievement. Principals then assign a teacher j to each classroom c based on classroom and teacher characteristics. For simplicity, we assume that each teacher teaches one class per year, as in elementary schools.

Test Scores. Let $j = j(c(i,t))$ denote student i's teacher in school year t. Let μ_{jt} represent teacher j's "test-score value-added" in year t.[45] We scale μ_{jt} in student test-score SDs 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. Let $t_i(g)$ denote the calendar year in which student i reaches grade g; $t_i(0)$ denotes the year in which the student starts school (Kindergarten). As in Section I, student i' stest score in year t, ${A^*}_{it}$, is

$$A^*_{it} = \beta X_{it} + \mu_{jt} + \varepsilon_{it} \tag{16}$$

where X_{it} denotes observable determinants of student achievement, such as lagged test scores and family characteristics and ε_{it} is an idiosyncratic student-level error.

Earnings. Earnings are a function of teacher quality over all years in school, up to grade $G = 12$. Let τ_{jt} represent teacher j's "direct earnings value-added," i.e. the direct impact of teacher j on earnings holding other factors fixed. We scale τ_{jt} so that the average teacher has $\tau_{jt} = 0$ and the standard deviation of τ_{jt} is 1. We assume that a teacher's direct earnings value-added τ_{jt} is linearly related to her test-score value-added μ_{jt}:

$$\tau_{jt} = \phi\mu_{jt} + \tau^{\perp}_{jt}$$

where ϕ measures the relationship between earnings- and test-score VA and $\tau^{\perp}_{jt}$ represents the portion of a teacher's earnings impact that is orthogonal to her test-score impact.

For a student in grade h, we model earnings Y^*_i as

$$Y^*_i = \beta^Y X_{it} + \sum_{g=h}^{G} \gamma_g \tau_{j,t_i(g)} + \varepsilon^Y_{it} \tag{17}$$

where γ_g measures the effect of teacher quality in grade g on earnings. The error term ε^Y_{it} reflects individual heterogeneity in earnings ability, which may be correlated with academic ability ε_{it}. The error ε^Y_{it} may also be correlated with μ_{jt} and τ_{jt} because the principal may systematically sort

[45] To simplify notation, we write $\mu_{j(i,t),t}$ as μ_{jt} and always denote by j the teacher who taught student i in the relevant year t. For instance, $\mu_{j,t-s}$ denotes the value-added in year $t-s$ of the teacher j who taught student i in year $t-s$. We adopt the same convention with τ_{jt} below as well.

certain types of students to certain teachers. Accounting for such selection is the key challenge in obtaining unbiased estimates of teachers' causal impacts.

In the statistical model in (3), the teacher fixed effect $\alpha_j = \sum_{g=h}^{G} \gamma_g \tau_{j,t_i(g)}$ combines the effects of the current and subsequent teachers. We show how this affects the interpretation of the reduced-form treatment effects we estimate in the next subsection.

Treatment Effects in the Dynamic Model

We define two notions of a teacher's treatment effect on earnings: total earnings value-added and the impact of test-score VA on earnings.

Total Earnings Value-Added. One natural definition of a teacher's impact on earnings is the effect of changing the teacher of class c in grade g from j to j' in year t on expected earnings:

$$\mu^Y_{jt} - \mu^Y_{j't} = \mathbb{E}Y_{it}(j(i,t)) - \mathbb{E}Y_{it}(j'(i,t)) \quad (18)$$

$$= \gamma_g \left(\tau^Y_{jt} - \tau^Y_{j't}\right) + \sum_{s=g+1}^{G} \gamma_s \left(\mathbb{E}\left[\tau^Y_{j,t_i(s)} \mid j(i,t)\right] - \mathbb{E}\left[\tau^Y_{j,t_i(s)} \mid j'(i,t)\right]\right). \quad (19)$$

Being assigned teacher j instead of j' affects earnings through two channels. The first term in (19) represents the direct impact of the change in teachers on earnings. The second term represents the indirect impact via changes in the expected quality of subsequent teachers to which the student is assigned. For example, a higher achieving student may be tracked into a more advanced sequence of classes taught by higher quality teachers. In a more general model, other determinants of earnings such as parental effort or peer quality might also respond endogenously to the change in teachers. We refer to μ^Y_{jt} as a teacher's "total earnings value-added," or "earnings VA" for short.

In principle, one can estimate teachers' total earnings VA using an approach identical to the one we used to estimate teachers' test-score VA in our first paper. We could predict a teacher's earnings VA $\widehat{\mu}^Y_{jt}$ in year t based on mean residual earnings for students in other years with observational data. Such a prediction would yield unbiased forecasts of teachers' impacts on earnings if

$$\frac{Cov\left(\mu^Y_{jt}, \widehat{\mu}^Y_{jt}\right)}{Var\left(\widehat{\mu}^Y_{jt}\right)} = 1 \Rightarrow \frac{Cov\left(\varepsilon^Y_{it}, \widehat{\mu}^Y_{jt}\right)}{Var\left(\widehat{\mu}^Y_{jt}\right)} = 0. \quad (20)$$

This condition requires that unobserved determinants of students'earnings are orthogonal to earnings VA estimates. Although conceptually analogous to the requirement for forecast unbiasedness of test-score VA stated in Assumption 1, (20) turns out not to hold in practice. Tests for sorting on pre-determined characteristics analogous to those in Section IV of our first paper reveal that (20) is violated for earnings VA estimates based on the same control vector (prior test scores and student and classroom demographics) that we used to estimate test score VA. In particular, we find substantial "effects" of earnings VA estimates on parent income and family characteristics, indicating that our baseline control vector is unable to fully account for sorting when estimating earnings VA.

Why are we able to construct unbiased estimates of test score VA but not earnings VA? Controlling for lagged test scores effectively absorbs most unobserved determinants of student achievement on which students are sorted to classrooms, but does not account for unobserved determinants of earnings. To see how this can occur, let ζ_i denote a student's academic ability, which affects both test scores and earnings, and ζ^Y_i denote determinants of earnings that are orthogonal to academic achievement, such as family connections. Suppose students are sorted to teachers on the basis of both of these characteristics. The key difference between the two characteristics is that latent student ability ζ_i appears directly in $A_{i,t-1}$, whereas latent student earnings ability ζ_i^Y does not directly appear in $A_{i,t-1}$. As a result, variation in academic ability ζ_i can be largely purged from the error term $\widetilde{\varepsilon}'_{it}$ in the specification for test scores in (16) by controlling for $A_{i,t-1}$.[46] In contrast, family connections are not reflected in $A_{i,t-1}$ and therefore appear in the error term ε_{it}^Y in the specification for earnings in (17). Under such a data generating process, we would be able to identify teachers' causal impacts on test scores by controlling for $A_{i,t-1}$, but would not be able to identify teachers' causal impacts on earnings because there is systematic variation across teachers in students' earnings purely due to variation in family connections ζ_i^Y even conditional on $A_{i,t-1}$.

Consistent with this reasoning, we showed in our first paper that the key to obtaining forecast unbiased estimates of test-score VA was to control for prior test scores, $A_{i,t-1}$. If we observed an analog of lagged scores such as lagged expected earnings, we could effectively control for ζ_i^Y and potentially satisfy (20). Lacking such a control in our data, we cannot identify teachers' total earnings VA and defer this task to future work.

Impacts of Test-Score VA on Earnings. An alternative objective, which we focus on in our empirical analysis, is to identify the impacts of teachers' test-score based VA μ_{jt} on earnings. Let σ_μ denote the standard deviation of teachers' test-score VA. The reduced-form earnings impact of having a 1 SD better teacher, as measured by test-score VA, in grade g is

$$\kappa_g = \mathbb{E}[Y_{it} \mid \mu_{j't} = \mu_{jt} + \sigma_\mu] - \mathbb{E}[Y_{it} \mid \mu_{jt}] \tag{21}$$

$$= \gamma_g \phi \sigma_\mu + \sum_{s=g+1}^{G} \gamma_s \left(\mathbb{E}\left[\tau^Y_{j,t_i(s)} \mid \mu_{j't}\right] - \mathbb{E}\left[\tau^Y_{j,t_i(s)} \mid \mu_{jt}\right] \right). \tag{22}$$

The reduced-form impact κ_g consists of two terms. The first is the direct effect of having a better teacher in grade g in school year T, which is attenuated by $\phi = \frac{Cov(\tau_{jt}, \mu_{jt})}{Var(\mu_{jt})}$ because we only pick up the portion of earnings impacts that projects onto test-score VA. The second is the impact of having different teachers in subsequent grades.

Impacts of Multiple Teachers. Let $\tilde{\kappa}_g$ denote the impact of teachers' test-score VA in grade g on earnings holding fixed teachers' test-score VA in other grades. An intuitive specification to identify $\tilde{\kappa}_g$ is to generalize (5) and regress earnings on teacher VA in all grades simultaneously:

$$Y_i^* = \sum_{g=0}^{G} \left[\tilde{\kappa}_g \widehat{m}_{j,t_i(g)} + \tilde{\beta}_g X_{i,t_i(g)} \right] + \varepsilon_i^m. \tag{23}$$

Identifying $\{\tilde{\kappa}_g\}$ in (23) requires the orthogonality condition $Cov\left(\widehat{m}_{j,t_i(g)}, \varepsilon_i^m\right) = 0$. This orthogonality condition is violated if we do not include grade $g-1$ test scores $A_{i,g-1}$ in the control vector X because teacher assignment is correlated with lagged test scores and other factors that directly affect earnings, as shown in Table 7 of our companion paper. But $A_{i,g-1}$ is endogenous to grade $g-1$ teacher VA $\widehat{m}_{j,t_i(g-1)}$, implying that we cannot identify $\tilde{\kappa}_g$ by directly estimating (23). To address this problem, we instead estimate the degree of teacher tracking and then recover the net impacts $\kappa\tilde{}_g$ from our reduced form estimates κ_g using the iterative method in Section V.C.

[46] In general, controlling for lagged test scores need not completely account for the variation in ζ_i because lagged test scores are noisy measures of latent ability. The fact that controlling for $A_{i,t-1}$ does eliminate bias in practice (as shown in our first paper) suggests that students are allocated to classrooms based on factors highly correlated with $A_{i,t-1}$ and other factors that directly affect earnings (ζ_i^Y).

Conceptually, estimating the effects of multiple teachers requires simultaneous quasi-random assignment of teachers in multiple grades. Our primary research design, which requires conditioning on lagged test scores, only yields quasi-random variation in teacher assignment one grade at a time. This is why we cannot directly estimate (23) and also cannot identify the substitutability or complementarity of teachers' impacts across grades.

Online Appendix B: 1098-T Data and College Quality Index

Quality of 1098-T Data. We evaluate whether the 1098-T data capture college enrollment accurately in three ways.[47] First, we find that the correlation between enrollment counts for students age 18-21 based on 1098-T's and enrollment counts for colleges listed in the IPEDS dataset from the Department of Education exceeds 0.95. Second, the aggregate counts of students enrolled in college are aligned with estimates based on the CPS. In 2009, 27.4 million 1098-T forms were issued (Internal Revenue Service, 2010). According to the Current Population Survey (US Census Bureau, 2010, Tables V and VI), in October 2008, there were 22.6 million students in the U.S. (13.2 million full time, 5.4 million part-time, and 4 million vocational). As an individual can be a student at some point during the year but not in October and can receive a 1098-T form from more than one institution, the number of 1098-T forms for the calendar year should indeed be higher than the number of students as of October. Third, two independent evaluations of the Project STAR class size experiment using data from 1098-T's (Chetty et al. 2011) and the National Student Clearinghouse (Dynarski et al. 2013) obtained nearly identical point estimates of the impacts of class size on college attendance.

College Quality Index. Our index of college quality is based on the average earnings of the individuals who attend each college. The construction of such an index requires several choices, including (1) the age at which college attendance is measured, (2) the age at which earnings are measured, (3) the cohort of students used, and (4) the definition of earnings. In what follows, we assess the stability of rankings of colleges with respect to these four choices.

We begin by constructing measures of college quality that vary the four parameters above. In each case, we first identify all individuals who are U.S. citizens as of February 19, 2013 to remove those who were temporarily in the United States for college and for whom we do not have post-college earnings data.[48] We group individuals by the higher education institution they attended and by age of attendance, as measured on December 31 of each year.[49] We group individuals not enrolled at a higher education institution at a given age (i.e., those who have no 1098-T form filed on their behalf during the tax year) in a separate "no college" category. For each college (including the "no college" group), we then compute earnings of the students at various ages (in real 2010 dollars). We begin by defining earnings based on individual W-2 wage earnings and then consider broader income measures. We top code individual earnings at $10 million to reduce the influence of outliers and we include only those who are alive at the age at which we measure earnings.

We first evaluate the stability of rankings of college quality with respect to the age at which we measure earnings. Appendix Figure 1a plots the percentile ranking of colleges based on earnings measured at age 23 (one-year after most students graduate from 4 year colleges) and age 27 (five-years post-college) vs. the oldest age at which we can measure earnings of college graduates in our

[47] Legally, colleges are not required to file 1098-T forms for students whose qualified tuition and related expenses are waived or paid entirely with scholarships or grants. However, the forms appear to be available even for such cases, perhaps because of automated reporting to the IRS by universities.

[48] Only current citizenship status is recorded in the database. As a result, the date at which we determine citizenship is simply the date we accessed the data.

[49] We include the small fraction of students who attend more than one college in a single year in the determination of college quality for each unique institution to which they are matched.

sample, which is 32 (ten-years post-college). We hold the age of college attendance constant at 20 and focus on the cohort of students born in 1979. To construct this figure, we bin colleges into 100 percentiles based on their ranking using age 32 earnings (without any weighting) and compute the mean percentile ranking based on earnings at age 23 and 27 within each bin. Rankings at age 27 are very well aligned with rankings at age 32, but rankings at age 23 are very poorly aligned with measures based on older data. Colleges that have the highest-earning graduates at age 32 are commonly ranked average or even below-average based on data at age 23.

In Appendix Figure 1b, we extend this analysis to cover all ages from 23-32. This figure plots the rank correlation between college quality measured at age 32 with college quality measured using earnings at earlier ages. Each point shows the correlation of an earnings-based percentile ranking at a given age with the ranking based on earnings at age 32. The correlation is very low at age 23 and rises steeply at first before asymptoting to 1. At age 28 and after, the correlations are all above 0.95, implying that we obtain similar rankings irrespective of what age one uses to measure earnings of college graduates beyond this point. The stability of the index starting in the late 20's is consistent with evidence from other studies that annual earnings starting in the late 20's are quite highly correlated with lifetime earnings (Haider and Solon 2006).

Panel A of Appendix Table 2 presents the rank correlations to corresponding to Appendix Figure 1b. The rest of Appendix Table 2 studies the rank correlations between college quality measures as we vary the other parameters. In Panel B, we vary the age at which we measure college attendance from 18 to 25, holding fixed the age of earnings measurement at 30 for the cohort born in 1981 (which is the oldest cohort for which we can measure college attendance at age 18). When we measure attendance between 18 and 22, college quality rankings are very highly correlated with each other. The correlations begin to fall when we measure attendance at later ages. This is intuitive, as ages 18-22 correspond to those at which most students would attend a 4-year college if they proceed through school at a normal pace.

Panel C of Appendix Table 2 varies the cohort of students included in the measure, including students born between 1979 and 1981. We hold fixed the ages of college attendance and earnings measurement at 20 and 30, respectively. The measures are very highly correlated across cohorts, showing that the reliability of our index of college quality is quite high.

Finally, Panel D of Appendix Table 2 shows the relationship between college quality measures based on alternative definitions of earnings. In addition to our baseline measure of mean W-2 earnings, we consider median W-2 earnings and mean total income (W-2 wages plus self-employment income from Form 1040). In each case we hold fixed age in college at 20, age of earnings measurement at 30, and focus on the 1979 cohort. The correlation between these measures exceeds 0.94, showing that the rankings are not sensitive to the concept of income used to measure earnings. We view W-2 earnings as the preferred measure because it is unaffected by marriage and the endogeneity of filing.

Based on these results, we construct our preferred measure of college quality measuring college attendance at age 20 and mean W-2 earnings at age 31. These choices allow us to combine data from two cohorts — students born in 1979 and 1980 — for whom we measure earnings in 2010 and 2011, respectively. We code college quality as missing for a small number of institutions with fewer than 100 students across the two cohorts and for institutions founded in or after 2001.[50] If students attended two or more colleges in a given year, we assign them the maximum college quality across all colleges attended.

[50] As a result, 0.21% of students who attend college at age 20 in our sample are missing information on college quality.

Online Appendix C: Identifying Teachers' Net Impacts

This appendix shows that the iterative method described in Section V.C recovers the net impacts of teacher VA, $\tilde{\kappa}_g$, defined as the impact of raising teacher VA in grade g on earnings, holding fixed VA in subsequent grades. The derivation below assumes that true VA is observed; when using VA estimates instead of true VA, one must account for attenuation due to shrinkage, as discussed in the text.

To simplify notation, we omit controls in this derivation; in practice, we residualize all the dependent variables in the regressions below with respect to the standard control vector. Furthermore, we replace the year subscript t with a grade subscript g, so that $m_{jg} = m_{j,t_i(g)}$.

We begin by estimating the following equations using OLS for $g \in [4, 8]$:

$$Y_{ig} = \kappa_g m_{jg} + \varepsilon_{ig}^m \tag{24}$$

$$m_{jg'} = \rho_{gg'} m_{jg} + \eta_{igg'}^{\rho} \quad \forall g' > g \tag{25}$$

The first set of equations identifies the reduced-form impact of teacher VA in grade g on earnings. The second set of equations identifies the impact of teacher VA in grade g on teacher VA in future grade g'. Note that identification of the tracking coefficients $\rho_{gg'}$ using (13) requires the following variant of Assumption 2:

Assumption 2A Teacher value-added in grade g is orthogonal to unobserved determinants of future teacher value-added conditional on controls:

$$Cov\left(m_{jg}, \eta_{igg'}^{\rho}\right) = 0.$$

After estimating $\{\kappa_g\}$ and $\{\rho_{gg'}\}$, we recover the net impacts $\tilde{\kappa}_g$ as follows. In the additive model of teacher effects in Appendix A, earnings Y_{ig} for a student in grade g can be written as $\sum_{g'=4}^{8} \tilde{\kappa}_{g'} m_{jg'} + \varepsilon_{ig}^m$. Substituting this definition of Y_{ig} into (24) and noting that $\rho_{gg'} = Cov\left(m_{jg'}, m_{jg}\right)/Var\left(m_{jg}\right)$ yields

$$\kappa_g = \frac{Cov\left(\sum_{g'=4}^{8} \tilde{\kappa}_{g'} m_{jg'} + \varepsilon_{ig'}^m, m_{jg}\right)}{Var\left(m_{jg}\right)} = \sum_{g'=4}^{8} \rho_{gg'} \tilde{\kappa}_{g'}.$$

One implication of Assumption 2, the orthogonality condition needed to identify earnings impacts, is that

$$Cov\left(m_{jg'}, m_{jg}\right) = 0 \quad \text{for } g' < g$$

since past teacher quality $m_{j(i,g')}$ is one component of the error term ε_{igt}^{μ} in (24). Combined with the fact that $\rho_{gg} = 1$ by definition, these equations imply that

$$\begin{aligned} \kappa_g &= \tilde{\kappa}_g + \sum_{g'=g+1}^{8} \rho_{gg'} \tilde{\kappa}_{g'} \quad \forall g < 8 \\ \kappa_8 &= \tilde{\kappa}_8. \end{aligned}$$

Rearranging this triangular set of equations yields the following system of equations, which can be solved by iterating backwards as in Section V.C:

$$\begin{aligned} \tilde{\kappa}_8 &= \kappa_8 \\ \tilde{\kappa}_g &= \kappa_g - \sum_{g'=g+1}^{8} \rho_{gg'} \tilde{\kappa}_{g'} \quad \forall g < 8. \end{aligned} \tag{26}$$

Online Appendix D: Policy Simulations

Deselection based on Average VA in Math and English. Suppose we deselect the 5% of elementary school teachers with the lowest mean standardized VA across math and English. Simulating a bivariate normal distribution with a within-year correlation between $\widehat{m}_{jt}$ across math and English of $r = 0.6$, we calculate that teachers whose mean VA across subjects is in the bottom 5% have a standardized VA that is $\Delta m_\sigma = 1.84$ SD below the mean in both math and English.

To calculate the long-term earnings impact of replacing such teachers, we must identify the impacts of changes in VA in one subject holding fixed VA in the other subject. Given the between-subject VA correlation of $r = 0.6$, our earnings impact estimate of $b = 1.34\%$ reflects the effect of a 1 SD improvement in a given subject (e.g. math) combined with a 0.6 SD improvement in the other subject (English). Under the simplifying assumption that earnings impacts do not vary across subjects, the impact of a 1 SD improvement in VA in a given subject is $b_s = \frac{b}{1+0.6} = 0.84\%$. Therefore, replacing a teacher with mean VA in the bottom 5% with an average teacher for one school year in elementary school increases the present value of a student's earnings by

$$G' = \$522,000 \times 2 \times 1.84 \times 0.84\% = \$16,100$$

and yields total gains of $\$454,000$ for an average-sized classroom.

Deselection on Estimated VA: Monte-Carlo Simulations. To calculate the integral in (15), we first construct Σ_A, the VCV matrix of $\overrightarrow{A}_j^{-t}$, the vector of past class average scores, using the parameters of the autocovariance vector of test scores reported in Columns 1 and 2 of Table 2 of our companion paper. We define the off-diagonal elements of Σ_A directly based on the autocovariances σ_{As} reported in Table 2 of our first paper, setting the autocovariance $\sigma_{As} = \sigma_{A7}$ for $s > 7$. We define the diagonal elements of Σ_A as the variance of mean class test scores, which we compute based on the estimates in Table 2 as (Class+Teacher Level SD)^2 + (Individual-Level SD)^2/28.2, where 28.2 is the average number of students per class.

We then simulate draws of average class scores from a $N(0, \Sigma_A)$ distribution for one million teachers and calculate $\widehat{m}_{j,n+1}$ based on scores from the first n periods using the same method used to construct the VA estimates in our companion paper. Finally, we calculate the conditional expectation in (15) as the mean test score in year $n+1$ for teachers with $\widehat{m}_{j,n+1}$ in the bottom 5% of the distribution.

We calculate the gains from deselection based on true VA in Figure 8b using analogous Monte Carlo simulations, except that we draw scores from the VCV matrix of true VA Σ_μ instead of test scores Σ_A. The off-diagonal elements of the two matrices are identical, but the diagonal elements of Σ_μ reflect only the variance of teacher quality σ_μ^2. We use the quadratic estimates of σ_μ reported in the last row of Table 2 in our companion paper for this simulation.

Note that if one's goal is to maximize expected gains over a teacher's tenure, one should deselect teachers after n years based on mean predicted VA over all future years, discounted by the survival probabilities. We find that this more complex policy increases gains by less than 1% over 10 years relative to the policy of deselecting teachers based on VA estimates at the end of year 3 that we simulate in Figure 8b. Intuitively, because the VA drift process is close to an AR(1) process, the relative weights on average scores from a teacher's first three years do not change much when projecting beyond year 4.

The calculations we report in the text assume that VA estimates have zero forecast bias. While the estimates in our first paper do not reject this hypothesis, the upper bound on the 95% confidence interval for our quasi-experimental estimate of forecast bias is 9%, which would imply $\mathbb{E}\left[m_{j,n+1} \mid \widehat{m}_{j,n+1}\right] = 0.91\widehat{m}_{j,n+1}$. This degree of forecast bias has modest impacts on the gains from deselection: for instance, the earnings gains per class in year 4 based on 3 years of test score data are $G_C(3) = \$242,000$.

APPENDIX TABLE 1
Structure of Linked Analysis Dataset

Student	Subject	Year	Grade	Class	Teacher	Test Score	Matched to Tax Data?	Earnings at Age 28
Bob	Math	1992	4	1	Jones	0.5	1	$35K
Bob	English	1992	4	1	Jones	-0.3	1	$35K
Bob	Math	1993	5	2	Smith	0.9	1	$35K
Bob	English	1993	5	2	Smith	0.1	1	$35K
Bob	Math	1994	6	3	Harris	1.5	1	$35K
Bob	English	1994	6	4	Adams	0.5	1	$35K
Nancy	Math	2002	3	5	Daniels	0.4	0	.
Nancy	English	2002	3	5	Daniels	0.2	0	.
Nancy	Math	2003	4	6	Jones	-0.1	0	.
Nancy	English	2003	4	6	Jones	0.1	0	.

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

APPENDIX TABLE 2
Correlation of College Rankings Based on Alternative Measures

Panel A: Correlation of College Rankings Across Ages at Which Earnings are Measured

	Age 23	Age 24	Age 25	Age 26	Age 27	Age 28	Age 29	Age 30	Age 31	Age 32
Age 23	1.000									
Age 24	0.858	1.000								
Age 25	0.747	0.949	1.000							
Age 26	0.676	0.901	0.967	1.000						
Age 27	0.614	0.852	0.928	0.972	1.000					
Age 28	0.577	0.822	0.900	0.950	0.979	1.000				
Age 29	0.553	0.802	0.882	0.934	0.962	0.983	1.000			
Age 30	0.519	0.774	0.860	0.916	0.948	0.968	0.980	1.000		
Age 31	0.505	0.761	0.848	0.905	0.937	0.958	0.971	0.986	1.000	
Age 32	0.495	0.750	0.838	0.897	0.930	0.952	0.964	0.977	0.987	1.000

Panel B: Correlation of College Rankings Across Ages at Which College Attendance is Measured

	Age 18	Age 19	Age 20	Age 21	Age 22	Age 23	Age 24	Age 25
Age 18	1.000							
Age 19	0.948	1.000						
Age 20	0.930	0.975	1.000					
Age 21	0.909	0.947	0.972	1.000				
Age 22	0.880	0.914	0.940	0.968	1.000			
Age 23	0.850	0.886	0.909	0.933	0.960	1.000		
Age 24	0.803	0.830	0.851	0.873	0.893	0.932	1.000	
Age 25	0.766	0.790	0.806	0.830	0.851	0.883	0.935	1.000

Panel C: Correlation of College Rankings Across Birth Cohorts

	Cohort 1979	Cohort 1980	Cohort 1981
Cohort 1979	1.000		
Cohort 1980	0.931	1.000	
Cohort 1981	0.933	0.942	1.000

Panel D: Correlation of College Rankings Across Earnings Definitions

	Mean W-2	Median W-2	Mean W-2 + S-E
Mean W-2 Earnings	1.000		
Median W-2 Earnings	0.960	1.000	
Mean W-2 + Self-Employment Income	0.989	0.943	1.000

Notes: This table displays Spearman rank correlations between alternative earnings-based indices of college quality, each of which is defined by four characteristics: age of earnings measurement, age of college attendance, cohort of students, and definition of earnings. Throughout this table, we construct college quality measures from only a single birth cohort of students; however, the preferred measure used in the text combines two cohorts. Panel A varies the age of earnings measurement from 23 to 32, holding fixed the age of college attendance at 20, using only the 1979 cohort of students, and using mean W-2 wage earnings. Panel B varies the age of college attendance from 18 to 25, holding fixed the age of earnings measurement at 30, using only the 1981 cohort, and using mean W-2 wage earnings. Panel C varies the birth cohort of students, holding fixed the age of college attendance at 20, the age of earnings measurement at 30, and using mean W-2 wage earnings. Panel D varies the measure of earnings between the baseline (mean W-2 wage earnings by college) and two alternatives (median W-2 wage earnings by college and mean total income by college), holding fixed the age of college attendance at 20, the age of earnings measurement at 30, and using only the 1979 cohort of students.

APPENDIX TABLE 3

Cross-Sectional Correlations Between Outcomes in Adulthood and Test Scores

Dep. Var.:	College at Age 20	College Quality at Age 20	Earnings at Age 28	Teenage Birth	Percent College Grads in ZIP at Age 28
	(%) (1)	($) (2)	($) (3)	(%) (4)	(%) (5)
No Controls	18.37	6,366	7,709	-6.57	1.87
	(0.02)	(6)	(23)	(0.02)	(0.01)
With Controls	5.54	2,114	2,585	-1.58	0.34
	(0.04)	(11)	(59)	(0.05)	(0.01)
Math	6.04	2,295	2,998	-1.21	0.31
Full Controls	(0.06)	(16)	(83)	(0.07)	(0.02)
English	5.01	1,907	2,192	-2.01	0.37
Full Controls	(0.06)	(16)	(88)	(0.06)	(0.02)
Mean of Dep. Var.	37.71	26,963	21,622	13.25	13.43

Notes: Each cell reports coefficients from a separate OLS regression of an outcome in adulthood on test scores measured in standard deviation units, with standard errors reported in parentheses. The regressions are estimated on observations from the linked analysis sample (as described in the notes to Table 1). There is one observation for each student-subject-school year, and we pool all subjects and grades in estimating these regressions. The dependent variable is an indicator for attending college at age 20 in column 1, our earnings-based index of college quality in column 2, wage earnings at age 28 in column 3, an indicator for having a teenage birth (defined for females only) in column 4, and the fraction of residents in an individual's zip code of residence with a college degree or higher at age 28 in column 5. See notes to Table 1 for definitions of these variables. The regressions in the first row include no controls. The regressions in the second row include the full vector of student- and class-level controls used to estimate the baseline value-added model described in Section III.A, as well as teacher fixed effects. The regressions in the third and fourth row both include the full vector of controls and split the sample into math and English test score observations. The final row displays of the mean of the dependent variable in the sample for which we have the full control vector (i.e., the sample used in the 2nd row).

APPENDIX TABLE 4
Cross-Sectional Correlations Between Test Scores and Earnings by Age

	Dependent Variable: Earnings ($)								
Age:	20 (1)	21 (2)	22 (3)	23 (4)	24 (5)	25 (6)	26 (7)	27 (8)	28 (9)
No Controls	889 (20)	1,098 (25)	1,864 (28)	3,592 (34)	4,705 (39)	5,624 (44)	6,522 (48)	7,162 (51)	7,768 (54)
With Controls	392 (64)	503 (79)	726 (91)	1,372 (110)	1,759 (125)	1,971 (139)	2,183 (152)	2,497 (161)	2,784 (171)
Mean Earnings	6,484	8,046	9,559	11,777	14,004	16,141	18,229	19,834	21,320
Pct. Effect (with controls)	6.1%	6.2%	7.6%	11.6%	12.6%	12.2%	12.0%	12.6%	13.1%

Notes: Each cell in the first two rows reports coefficients from a separate OLS regression of earnings at a given age on test scores measured in standard deviation units, with standard errors in parentheses. See notes to Table 1 for our definition of earnings. We restrict this table to students born in cohorts 1979 and 1980, so that regressions are estimated on a constant subsample of the linked analysis sample. There is one observation for each student-subject-school year, and we pool all subjects and grades in estimating these regressions. The first row includes no controls; the second includes the full vector of student- and class-level controls used to estimate the baseline value-added model described in Section III.A as well as teacher fixed effects. Means of earnings for the estimation sample with controls are shown in the third row. The last row divides the coefficient estimates from the specification with controls by the mean earnings to obtain a percentage impact by age.

APPENDIX TABLE 5
Heterogeneity in Cross-Sectional Correlations Across Demographic Groups

Dependent Variable:	Earnings at Age 28	College at at Age 20	College Quality Age 20	Teenage Birth
	($)	(%)	($)	(%)
	(1)	(2)	(3)	(4)
Male	2,408	5.36	1,976	n/a
	(88)	(0.06)	(16)	
	[22,179]	[34.24]	[26,205]	
Female	2,735	5.74	2,262	-1.58
	(80)	(0.06)	(17)	(0.05)
	[21,078]	[41.07]	[27,695]	[13.25]
Non-minority	2,492	5.11	2,929	-0.72
	(139)	(0.08)	(27)	(0.04)
	[31,587]	[59.67]	[34,615]	[2.82]
Minority	2,622	5.65	1,734	-1.96
	(62)	(0.05)	(12)	(0.06)
	[17,644]	[28.98]	[23,917]	[17.20]
Low Parent Inc.	2,674	5.14	1,653	-1.72
	(85)	(0.06)	(15)	(0.07)
	[18,521]	[26.91]	[23,824]	[16.67]
High Parent Inc.	2,573	5.73	2,539	-1.29
	(92)	(0.06)	(18)	(0.06)
	[26,402]	[49.92]	[30,420]	[9.21]

Notes: Each column reports coefficients from an OLS regression, with standard errors in parentheses and the mean of the dependent variable in the estimation sample in brackets. These regressions replicate the second row (full sample, with controls and teacher fixed effects) of estimates in Columns 1-4 of Appendix Table 3, splitting the sample based on student demographic characteristics. The demographic groups are defined in exactly the same way as in Panel A of Table 6. We split rows 1 and 2 by the student's gender. We split the sample in rows 3 and 4 based on whether a student belongs to an ethnic minority (black or hispanic). We split the sample in rows 5 and 6 based on whether a student's parental income is higher or lower than median in the sample, which is $31,905.

APPENDIX TABLE 6

Cross-Sectional Correlations between Test Scores and Outcomes in Adulthood by Grade

Dep. Variable:	Earnings at Age 28	College at Age 20	College Quality at Age 20	Earnings at Age 28	College at Age 20	College Quality at Age 20
	No Controls			With Controls		
	($)	(%)	($)	($)	(%)	($)
	(1)	(2)	(3)	(4)	(5)	(6)
Grade 4	7,561	18.29	6,378	2,970	6.78	2,542
	(57)	(0.05)	(13)	(122)	(0.09)	(23)
Grade 5	7,747	18.27	6,408	2,711	5.28	2,049
	(50)	(0.05)	(13)	(108)	(0.08)	(23)
Grade 6	7,524	17.95	6,225	2,395	4.92	1,899
	(51)	(0.05)	(14)	(140)	(0.10)	(27)
Grade 7	7,891	18.23	6,197	2,429	4.48	1,689
	(54)	(0.05)	(14)	(198)	(0.11)	(29)
Grade 8	7,795	19.10	6,596	2,113	5.43	2,106
	(48)	(0.05)	(13)	(141)	(0.11)	(28)

Notes: Each column reports coefficients from an OLS regression, with standard errors in parentheses. The regressions in the first three columns replicate the first row (full sample, no controls) of estimates in Columns 1-3 of Appendix Table 3, splitting the sample by grade. The regressions in the second set of three columns replicate the second row (full sample, with controls and teacher fixed effects) of estimates in Columns 1-3 of Appendix Table 3, again splitting the sample by grade.

APPENDIX TABLE 7

Robustness of Baseline Results to Student-Level Controls, Clustering, and Missing Data

Dep. Var.:	College at Age 20 (%) (1)	College Quality at Age 20 ($) (2)	Earnings at Age 28 ($) (3)
Panel A: Individual Controls			
Teacher VA, with baseline controls	0.825	299	350
	(0.072)	(21)	(92)
Observations	4,170,905	4,167,571	650,965
Teacher VA, with additional individual controls	0.873	312	357
	(0.072)	(21)	(90)
Observations	4,170,905	4,167,571	650,965
Panel B: Clustering			
Teacher VA, school clustered	0.825	299	350
	(0.115)	(36)	(118)
Observations	4,170,905	4,167,571	650,965
Panel C: Missing Data			
Teacher VA, cells > 95% VA coverage	0.819	277	455
	(0.090)	(26)	(202)
Observations	2,238,143	2,236,354	363,392
Teacher VA, cells > median match rate	0.912	345	563
	(0.094)	(28)	(203)
Observations	2,764,738	2,762,388	278,119

Notes: The table presents robustness checks of the main results in Tables 2 and 3. In Panel A, the first row replicates Columns 1 and 4 of Table 2 and Column 1 of Table 3 as a reference. The second row adds individual controls, so that the control vector exactly matches that used to estimate the value-added model (see Section III.A for details). Panel B clusters standard errors by school. In Panel C, the first row limits the sample to those school-grade-year-subject cells in which we are able to calculate teacher value-added for at least 95% of students. The second row limits the sample to those school-grade-year-subject cells in which the rate at which we are able to match observations to the tax data is more than the school-level-subject-specific median across cells.

APPENDIX TABLE 8

Impacts of Teacher Value-Added: Sensitivity to Trimming

	Percent Trimmed in Upper Tail						Bottom and Top 1%	Jacob and Levitt Proxy
	5%	4%	3%	2%	**1%**	0%		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test Score	0.972	0.975	0.981	0.987	0.993	1.005	0.996	0.918
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	0.006	(0.006)
College at Age 20	0.93	0.90	0.88	0.86	0.82	0.72	0.79	1.10
	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.09)
College Quality at Age 20	329	320	315	307	299	276	292	371
	(23)	(22)	(22)	(21)	(21)	(21)	(21)	(25)
Earnings at Age 28	404	405	390	356	350	248	337	391
	(102)	(100)	(99)	(96)	(92)	(91)	(94)	(118)

Notes: This table presents results that use alternative approaches to trimming the tails of the distribution of teacher VA. Each coefficient reports the coefficient on teacher VA from a separate OLS regression, with standard errors clustered by school-cohort in parentheses. The regressions in the first row replicate the baseline specification used in Column 1 of Table 3 in our companion paper (using VA scaled in units of student test-score SDs), except that we include only the class-level controls that correspond to the baseline set of controls in this paper (as in Section III.A). The regressions in rows 2-4 replicate the baseline specification used in Columns 1 and 4 of Table 2 and Column 1 of Table 3. Columns 1-6 report results for trimming the upper tail at various cutoffs. Column 7 shows estimates when both the bottom and top 1% of VA outliers are excluded. Finally, Column 8 excludes teachers who have more than one classroom that is an outlier according to Jacob and Levitt's (2003) proxy for cheating. Jacob and Levitt define an outlier classroom as one that ranks in the top 5% of a test-score change metric defined in the notes to Appendix Figure 3. The results in Column 5 (1% trimming) correspond to those reported in the main text.

APPENDIX TABLE 9
Impacts of Teacher Value-Added on Outcomes by Age

Panel A: College Attendance

Dependent Variable:	College Attendance (%)										
Age:	18	19	20	21	22	23	24	25	26	27	28
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Teacher Value-Added	0.61	0.81	0.82	0.98	0.71	0.44	0.58	0.46	0.50	0.46	-0.01
	(0.06)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.09)	(0.11)
Mean Attendance Rate	29.4	36.8	37.2	35.7	32.2	24.4	20.31	17.3	15.7	13.9	12.3

Panel B: Wage Earnings

Dependent Variable:	Earnings ($)								
Age:	20	21	22	23	24	25	26	27	28
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Teacher Value-Added	-32	-35	-18	44	74	141	230	254	350
	(11)	(14)	(18)	(25)	(32)	(44)	(47)	(63)	(92)
Mean Earnings	5,696	7,293	9,473	12,582	15,080	17,547	18,833	20,229	21,256

Notes: These results present the regression estimates underlying the results in Panel C of Figure 1 (in Panel A) and Panel B of Figure 2 (in Panel B). The regressions in Panel A match the specification from Column 1 of Table 2, with college attendance measured at different ages; those in Panel B match the specification from Column 1 of Table 3.

APPENDIX TABLE 10

Impacts of Teacher Value-Added on Current and Future Test Scores

Dep. Var.:	Test Score (SD)				
	t	t+1	t+2	t+3	t+4
	(1)	(2)	(3)	(4)	(5)
Teacher VA	0.993	0.533	0.362	0.255	0.221
	(0.006)	(0.007)	(0.007)	(0.008)	(0.012)
Observations	7,401,362	5,603,761	4,097,344	2,753,449	1,341,266

Notes: This table presents the regression estimates plotted in Figure 4; see notes to that figure for details.

APPENDIX TABLE 11
Impacts of Value-Added on College Quality by Grade

	College Quality at Age 20				
Panel A: Reduced-Form Coefficients					
	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Teacher Value-Added	226	289	292	482	198
	(31)	(33)	(48)	(61)	(48)
Panel B: Coefficients Net of Teacher Tracking					
	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Teacher Value-Added	194	270	173	402	198

Notes: This table presents the regression estimates plotted in Figure 7; see notes to that figure for details.

APPENDIX TABLE 12

Tracking: Impact of Current Teacher VA on Future Teachers' VA

	Future Teacher Value-Added			
	Grade 5	Grade 6	Grade 7	Grade 8
Grade 4 Teacher VA	0.028	0.057	0.024	0.027
	(0.007)	(0.006)	(0.004)	(0.005)
Grade 5 Teacher VA		0.059	0.014	0.019
		(0.006)	(0.005)	(0.014)
Grade 6 Teacher VA			0.198	0.196
			(0.012)	(0.014)
Grade 7 Teacher VA				0.405
				(0.017)

Notes: Each cell reports the coefficient from a separate regression of teacher value-added in a subsequent grade on teacher value-added in the current grade, with standard errors at the school-cohort level. As in Figure 7, we first residualize each dependent variable (i.e. lead VA, two-year lead VA, etc.) with respect to the classroom-level baseline control vector (see notes to Table 2 for more details). We then regress residualized future VA on current VA interacted with grade. We multiply the resulting regression coefficients by 1.63 to account for the attenuation bias due to using VA estimates instead of true VA as the dependent variable (see text for details). All regressions are estimated using observations in the linked analysis sample for which the student is progressing through grades at normal pace (e.g., the student is in sixth grade two years after fourth grade).

APPENDIX TABLE 13

Earnings Impacts of Replacing Teachers Below 5th Percentile with Average Teachers

	Panel A: Impacts in First Year After Deselection			
	Selection on Estimated VA		Selection on True VA	
Years Used to Estimate VA	Present Value of Earnings Gain per Class	Undiscounted Sum of Earnings Gain per Class	Present Value of Earnings Gain per Class	Undiscounted Sum of Earnings Gain per Class
1	$225,843	$1,249,636	$406,988	$2,251,954
2	$256,651	$1,420,105		
3	$265,514	$1,469,147		
4	$269,297	$1,490,081		
5	$272,567	$1,508,174		
6	$274,143	$1,516,891		
7	$275,232	$1,522,918		
8	$276,665	$1,530,845		
9	$278,112	$1,538,851		
10	$279,406	$1,546,013		

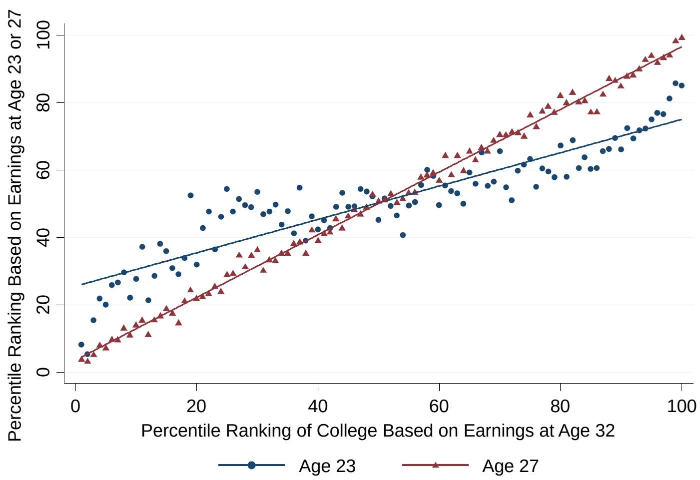
	Panel B: Impacts in Subsequent School Years			
	Selection on Estimated VA in Yr. 4		Selection on True VA in Yr. 4	
School Years Since Teacher was Hired	Present Value of Earnings Gain per Class	Undiscounted Sum of Earnings Gain per Class	Present Value of Earnings Gain per Class	Undiscounted Sum of Earnings Gain per Class
4	$265,514	$1,469,147	$406,988	$2,251,954
5	$229,923	$1,272,213	$339,870	$1,880,574
6	$202,631	$1,121,202	$297,569	$1,646,511
7	$183,538	$1,015,557	$252,422	$1,396,703
8	$172,867	$956,509	$222,339	$1,230,251
9	$161,575	$894,032	$212,185	$1,174,067
10	$157,812	$873,209	$193,255	$1,069,324
11	$155,349	$859,581	$180,876	$1,000,824
12	$156,582	$866,400	$180,909	$1,001,007
13	$156,547	$866,206	$181,027	$1,001,662
Avg. Gain	$184,234	$1,019,405	$246,744	$1,365,288

Notes: In Panel A, we present the earnings impacts per classroom of a policy that deselects the bottom 5% of teachers after *N* years and replaces them with a teacher of median quality, where we vary *N* from 1 to 10. We calculate these values using the methods described in Section VI. The first column presents estimates of the NPV earnings gains of deselection based on teacher value-added that is estimated from *N* years of observing a single average-sized (28.2 students) classroom per year of student scores. The third column shows the theoretical gain from deselecting teachers based on current true value-added; this value does not vary across years. Panel B presents the per class impacts of deselecting teachers (after 3 years of observation) in subsequent school years. Column 1 reports the present value of earnings gains in the ten years (i.e., years 4-13) after deselecting teachers based on their VA estimate in year 4, constructed using the past three years of data. The first number in Column 1 of Panel B matches the 3rd number in Column 1 of Panel A. Column 3 presents analogous values, instead deselecting teachers based on true value-added in year 4, so that the gains in the year 4 match the gains reported in Column 3 of Panel A. Columns 2 and 4 in each panel replicate Columns 1 and 3, presenting the undiscounted sum of future earnings impacts instead of present values. The bottom row in the table reports the unweighted means of the estimates from years 4-13 in Panel B for each column; these are the values reported in the introduction of the paper.

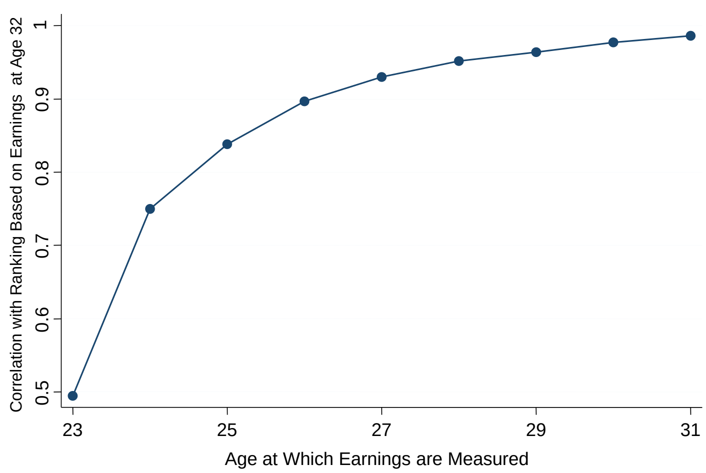
APPENDIX FIGURE 1

Stability of College Rankings by Age of Earnings Measurement

a) Rankings of Colleges Based on Earnings at Ages 23 and 27 vs. Age 32



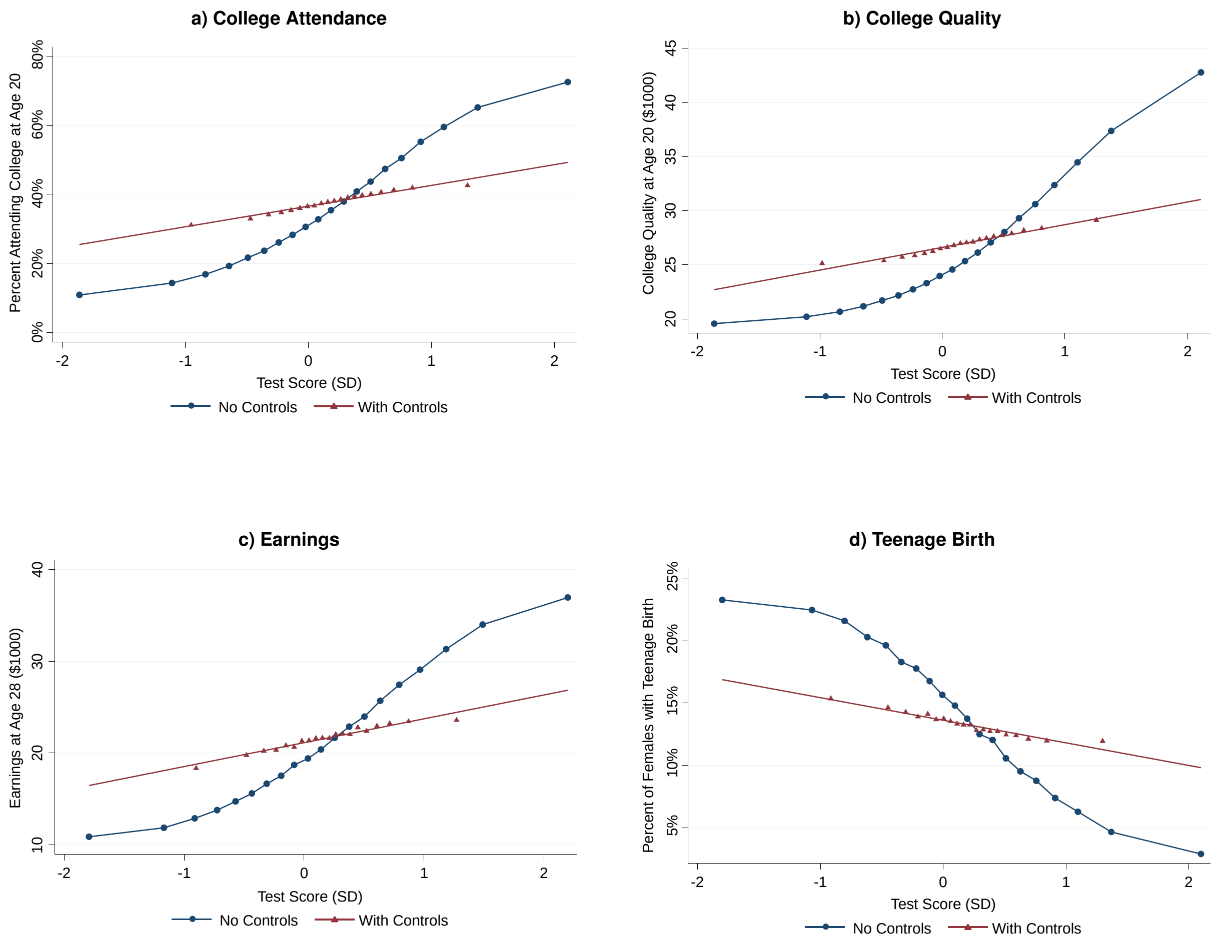
b) Correlation of College Rankings Based on Earnings at Age 32
With Rankings Based on Earnings at Earlier Ages



Notes: In Panel A, we take all college attendees in 1999 at age 20, as recorded by 1098-T forms, and construct three separate college quality indices by averaging W-2 earnings by college at ages 23, 27, and 32. We convert each college quality measure into a percentile rank based on the within-age distribution of college quality. We then bin colleges into 100 equal-sized (percentile) bins using the college quality measure based on age 32 earnings and plot the mean percentile rank of colleges in each bin using the age 23 (in circles) and age 27 (in triangles) measures. The best-fit lines are estimated from an unweighted OLS regression of percentile ranks run at the college level. In Panel B, we take the same college attendees and calculate ten separate college quality measures by averaging W-2 earnings by college at each age from 23-32. We then plot the Spearman rank correlation between each college quality measure based on earnings at ages 23-31 and the college quality measure based on earnings at age 32.

APPENDIX FIGURE 2

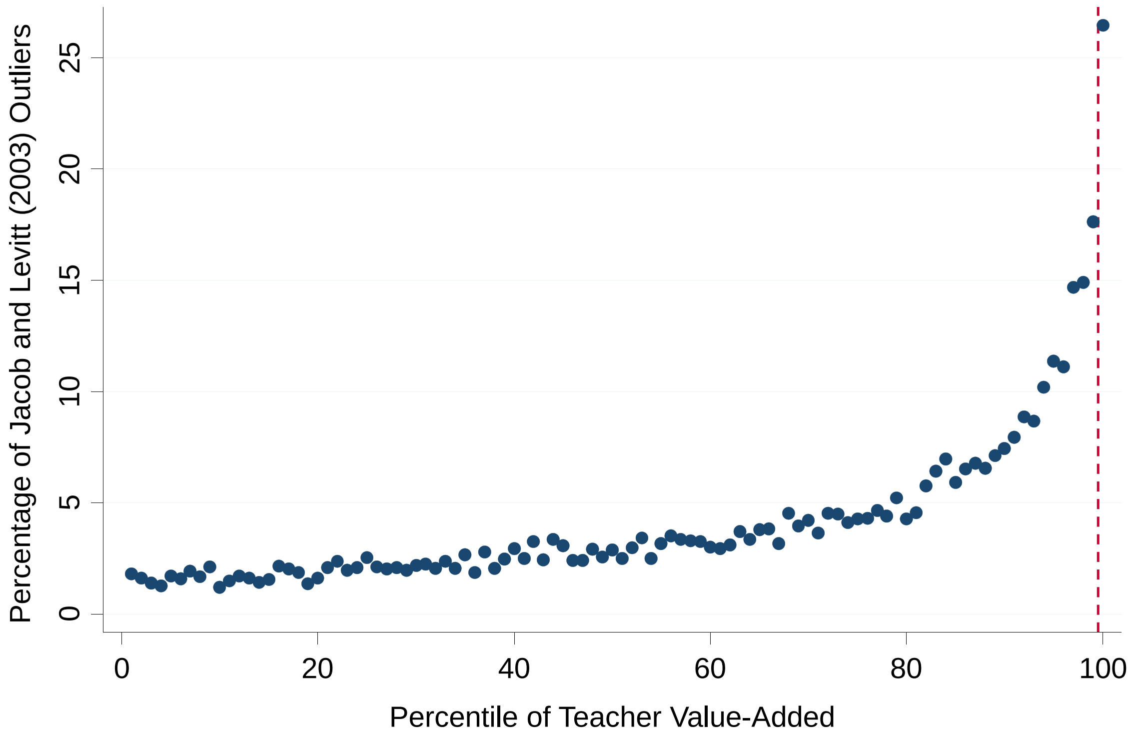
Correlations Between Outcomes in Adulthood and Test Scores



Notes: These figures present binned scatter plots corresponding to the cross-sectional regressions of outcomes in adulthood on test scores presented in Columns 1-4 of Appendix Table 3. See notes to Table 1 and Appendix Table 3 for further information on the variable definitions and sample specification. In each panel, the series in circles corresponds to the first row of estimates, without controls. The series in triangles corresponds to the second row of estimates, which includes the full control vector used to estimate the value-added model. To construct the series in circles, we bin raw test scores into twenty equal-sized groups (vingtiles) and plot the means of the outcome within each bin against the mean test score within each bin. To construct the series in triangles, we first regress both the test scores and adult outcomes on the individual and class controls and teacher fixed effects and compute residuals of both variables. We then divide the test score residuals into twenty equal-sized groups and plot the means of the outcome residuals within each bin against the mean test score residuals within each bin. Finally, we add back the unconditional mean of both test scores and the adult outcome in the estimation sample to facilitate interpretation of the scale. We connect the dots in the non-linear series without controls and show a best-fit line for the series with controls, estimated using an OLS regression on the microdata.

APPENDIX FIGURE 3

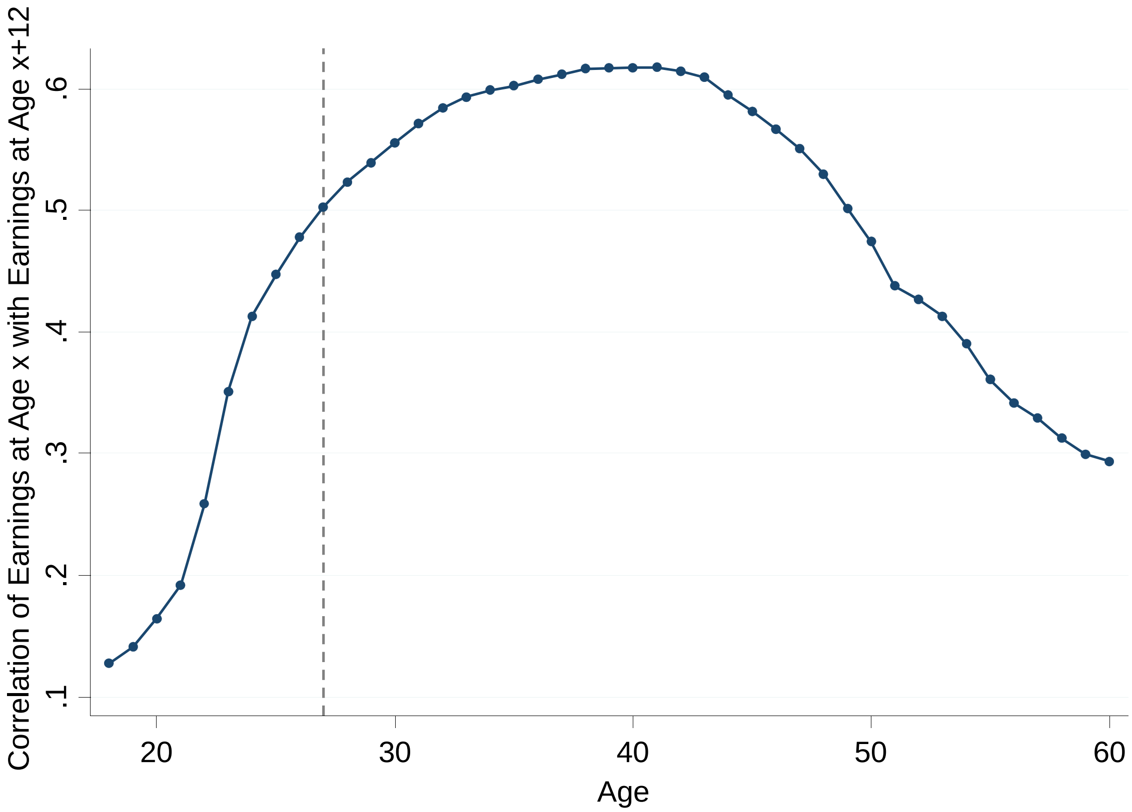
Jacob and Levitt (2003) Proxy for Test Manipulation vs. Value-Added Estimates



Notes: This figure plots the relationship between our leave-out-year measure of teacher value added and Jacob and Levitt's proxy for cheating. 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 III.A. The y-axis variable is constructed as follows: Let $\Delta\bar{A}_{c,t} = \bar{A}_{c,t} - \bar{A}_{c,t-1}$ denote the change in mean test scores from year $t-1$ and t for students in classroom c. Let $R_{c,t}$ denote the ordinal rank of classroom c in $\Delta\bar{A}_{c,t}$ among classrooms in its grade, subject, and school year and $r_{c,t}$ the ordinal rank as a fraction of the total number of classrooms in that grade, subject, and school year. Jacob and Levitt's (2003) measure for cheating in each classroom is $JL_c = (r_{c,t})^2 + (1 - r_{c,t+1})^2$. Higher values of this proxy indicate very large test score gains followed by very large test score losses, which Jacob and Levitt show to be correlated with a higher chance of having suspicious patterns of answers indicative of cheating. Following Jacob and Levitt, we define a classroom as an outlier if its value of JL_c falls within the top 5% of classrooms in the data. To construct the binned scatter plot, we group classrooms into percentiles based on their teacher's estimated value-added, ranking classrooms separately by school-level and subject. We then compute the percentage of Jacob-Levitt outliers within each percentile bin and scatter these fractions vs. the percentiles of teacher VA. Each point thus represents the fraction of Jacob-Levitt outliers at each percentile of teacher VA. The dashed vertical line depicts the 99th percentile of the value-added distribution. We exclude classrooms with estimated VA above this threshold in our baseline specifications because they have much higher frequencies of Jacob-Levitt outliers. See Appendix Table 8 for results with trimming at other cutoffs.

APPENDIX FIGURE 4

Correlation of Earnings Over the Lifecyle



Notes: This figure plots the correlation of wage earnings at each age x with wage earnings at age $x + 12$. We calculate wage earnings as the sum of earnings reported on all W-2 forms for an individual in a given year. Individuals with no W-2 are assigned 0 wage earnings. Earnings at age x are calculated in 1999, the first year in which we have W-2 data, and earnings at age $x+12$ are calculated in 2011, the last year of our data. We calculate these correlations using the population of current U.S. citizens. The dashed vertical line denotes age 28, the age at which we measure earnings in our analysis of teachers' impacts.