The Long-Term Impacts of Teachers: Teacher Value-Added and Students’ Outcomes in Adulthood

Raj Chetty, Harvard
John N. Friedman, Harvard
Jonah Rockoff, Columbia

January 2014
How can we measure and improve the quality of teaching in elementary schools?


Rate teachers based on their students’ test score gains

School districts have started to use VA to evaluate teachers, leading to considerable debate

Ex: Washington D.C. lays off teachers and offers bonuses using a metric that puts 50% weight on VA measures
Debate About Teacher Value-Added

- Debate stems primarily from two intellectual issues:
  1. Disagreement about whether VA measures are biased
     [Kane and Staiger 2008, Rothstein 2010]
     - Do differences in test-score gains across teachers capture causal impacts of teachers or are they driven by student sorting?
     - If VA estimates are biased, they will incorrectly reward or penalize teachers for the mix of students they get
  2. Lack of evidence on teachers’ long-term impacts
     - Do teachers who raise test scores improve students’ long-term outcomes or are they simply better at teaching to the test?
Objectives of This Project

- This study answers these two questions by tracking one million children from childhood to early adulthood
  - Develop new quasi-experimental tests for bias in VA estimates
  - Test whether children who get high VA teachers have better outcomes in adulthood

- Results also shed light on broader issues in the economics of education
  - What are the long-run returns to investments in better teaching?
  - Are impacts on scores a good proxy for long-term impacts of educational interventions?
Outline

1. Data

2. Construction of Value-Added Estimates with Drift

3. Evaluating Bias in Value-Added Estimates

4. Long-Term Impacts

5. Policy Implications

Results drawn from two papers: *Measuring the Impacts of Teachers: I & II*, NBER Working Papers 19423, 19424
Dataset 1: School District Data

- Teacher and class assignments from 1991-2009 for 2.5 million children

- Test scores from 1989-2009
  - Scaled scores standardized by grade and subject (math/reading)
  - 18 million test scores, grades 3-8

- Exclude students in special ed. schools and classrooms (6% of obs.)
Selected data from U.S. federal income tax returns from 1996-2010

- Includes non-filers via information forms (e.g. W-2’s)

Student outcomes: earnings, college, teenage birth, neighborhood quality

Parent characteristics: household income, 401k savings, home ownership, marital status, age at child birth

- Omitted variables from standard VA models

Approximately 90% of student records matched to tax data

- Data were analyzed as part of a broader project on tax policy

- Research based purely on statistics aggregating over thousands of individuals, not on individual data
## Data Structure

<table>
<thead>
<tr>
<th>Student</th>
<th>Subject</th>
<th>Year</th>
<th>Grade</th>
<th>Class</th>
<th>Teacher</th>
<th>Test Score</th>
<th>Age 28 Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tom</td>
<td>Math</td>
<td>1992</td>
<td>4</td>
<td>1</td>
<td>Samuelson</td>
<td>0.5</td>
<td>$22K</td>
</tr>
<tr>
<td>Tom</td>
<td>English</td>
<td>1992</td>
<td>4</td>
<td>1</td>
<td>Samuelson</td>
<td>1.3</td>
<td>$22K</td>
</tr>
<tr>
<td>Tom</td>
<td>Math</td>
<td>1993</td>
<td>5</td>
<td>2</td>
<td>Solow</td>
<td>0.9</td>
<td>$22K</td>
</tr>
<tr>
<td>Tom</td>
<td>English</td>
<td>1993</td>
<td>5</td>
<td>2</td>
<td>Solow</td>
<td>0.1</td>
<td>$22K</td>
</tr>
<tr>
<td>Tom</td>
<td>Math</td>
<td>1994</td>
<td>6</td>
<td>3</td>
<td>Arrow</td>
<td>1.5</td>
<td>$22K</td>
</tr>
<tr>
<td>Tom</td>
<td>English</td>
<td>1994</td>
<td>6</td>
<td>4</td>
<td>Stigler</td>
<td>0.5</td>
<td>$22K</td>
</tr>
</tbody>
</table>

One observation per student-subject-year
## Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (1)</th>
<th>S.D. (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student Data:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class size (not student-weighted)</td>
<td>28.2</td>
<td>5.8</td>
</tr>
<tr>
<td>Test score (SD)</td>
<td>0.12</td>
<td>0.91</td>
</tr>
<tr>
<td>Female</td>
<td>50.4%</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>11.7</td>
<td>1.6</td>
</tr>
<tr>
<td>Free lunch eligible (1999-2009)</td>
<td>77.1%</td>
<td></td>
</tr>
<tr>
<td>Minority (Black or Hispanic)</td>
<td>72.1%</td>
<td></td>
</tr>
<tr>
<td>English language learner</td>
<td>4.9%</td>
<td></td>
</tr>
<tr>
<td>Special education</td>
<td>3.1%</td>
<td></td>
</tr>
<tr>
<td>Repeating grade</td>
<td>2.7%</td>
<td></td>
</tr>
<tr>
<td>Number of subject-school years per student</td>
<td>6.25</td>
<td>3.18</td>
</tr>
<tr>
<td>Student match rate to adult outcomes</td>
<td>89.2%</td>
<td></td>
</tr>
<tr>
<td>Student match rate to parent chars.</td>
<td>94.8%</td>
<td></td>
</tr>
</tbody>
</table>
## Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (1)</th>
<th>S.D. (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult Outcomes:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual wage earnings at age 20</td>
<td>5,670</td>
<td>7,773</td>
</tr>
<tr>
<td>Annual wage earnings at age 25</td>
<td>17,194</td>
<td>19,889</td>
</tr>
<tr>
<td>Annual wage earnings at age 28</td>
<td>20,885</td>
<td>24,297</td>
</tr>
<tr>
<td>In college at age 20</td>
<td>35.6%</td>
<td></td>
</tr>
<tr>
<td>In college at age 25</td>
<td>16.5%</td>
<td></td>
</tr>
<tr>
<td>College Quality at age 20</td>
<td>26,408</td>
<td>13,461</td>
</tr>
<tr>
<td>Contribute to a 401(k) at age 25</td>
<td>19.1%</td>
<td></td>
</tr>
<tr>
<td>ZIP code % college graduates at age 25</td>
<td>13.7%</td>
<td></td>
</tr>
<tr>
<td>Had a child while a teenager (for women)</td>
<td>14.3%</td>
<td></td>
</tr>
<tr>
<td><strong>Parent Characteristics:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household income (child age 19-21)</td>
<td>40,808</td>
<td>34,515</td>
</tr>
<tr>
<td>Ever owned a house (child age 19-21)</td>
<td>34.8%</td>
<td></td>
</tr>
<tr>
<td>Contributed to a 401k (child age 19-21)</td>
<td>31.3%</td>
<td></td>
</tr>
<tr>
<td>Ever married (child age 19-21)</td>
<td>42.2%</td>
<td></td>
</tr>
<tr>
<td>Age at child birth</td>
<td>28.3</td>
<td>7.8</td>
</tr>
<tr>
<td>Predicted Score</td>
<td>0.17</td>
<td>0.26</td>
</tr>
</tbody>
</table>
Simplest case: teachers teach one class per year with $N$ students

All teachers have test score data available for $t$ previous years

Objective: predict test scores for students taught by teacher $j$ in year $t+1$ using test score data from previous $t$ years
Constructing Value-Added Estimates

- Three steps to estimate VA in year $t+1$

1. Form residual test scores, controlling for observables
   - Regress test scores $A_{is}$ on observable student characteristics $X_{is}$, including prior test scores $A_{i,s-1}$

2. Regress mean class-level test score residuals in year $t$ on class-level test score residuals in years $0$ to $t-1$

3. Use estimated coefficients $\psi_1, \ldots, \psi_t$ to predict VA in year $t+1$ based on mean test score residuals in years $1$ to $t$ for each teacher $j$

- Paper generalizes this approach to allow for variation in numbers of students and classes across teachers
Practical complications: number of students varies across classes, number of years varies across teachers, multiple classes per year in middle school.

Generalize regression approach by estimating an autocorrelation vector and assume stationarity of teacher VA process.

Then form a prediction for VA in each teacher-year using data from all other years using autocorrelation vector.

STATA ado file to implement this procedure on the web.
Two special cases:

1. Forecast VA in year $t$ using data from only year $t-s$:

$$
\hat{\mu}_{jt} = r_s \bar{A}_{j,t-s} \text{ where } r_s = Corr(\bar{A}_t, \bar{A}_{t-s})
$$

2. Without drift, put equal weight on all prior scores. Formula collapses to standard shrinkage estimator [e.g., Kane and Staiger 2008]

$$
\hat{\mu}_{jt} = \bar{A}_j^{-t} \frac{\sigma^2_{\mu}}{\sigma_{\mu}^2 + (\sigma_{\theta}^2 + \sigma_{\xi}^2/n)/T}
$$
Autocorrelation Vector in Elementary School for English and Math Scores

Years Between Classes

Correlation ($r_s$)

English

Math
Empirical Distribution of Estimated Teacher Effects in Elementary School

SD for English = 0.080
SD for Math = 0.116
Autocorrelation Vector in Middle School for English and Math Scores

Years Between Classes

Correlation

English

Math
Empirical Distribution of Estimated Teacher Effects in Middle School

SD for English = 0.042
SD for Math = 0.092
Test Scores vs. Teacher Value-Added

Coef. = 0.998
(0.006)
Part I: Bias in VA Estimates
Question 1: Are VA Estimates Unbiased?

- Teachers’ estimated VA may reflect unobserved differences in type of students they get rather than causal impact of teacher.

- We evaluate whether VA measures provided unbiased forecasts of teachers’ causal impacts in two ways.

  First test: are observable characteristics excluded from VA model are correlated with VA estimates?

  - Ex: parent income is a strong predictor of test scores even conditional on control vector used to estimate VA.

  - Do high VA teachers have students from higher-income families?

  - Combine parental background characteristics into a single predicted score using a cross-sectional regression.
Predicted Score Based on Twice-Lagged Score vs. Current Teacher VA

Predicted Score in Year $t$ vs. Teacher Value-Added

Coef. = 0.022 (0.002)
VA measures orthogonal to predictors of scores such as parent income

But selection on unobservables could still be a problem (Rothstein 2010)

Ideal test: out-of-sample forecasts in experiments (Kane and Staiger 2008)

Does a student who is randomly assigned to a teacher previously estimated to be high VA have higher test score gains?

We use teacher switching as a quasi-experimental analog
### Teacher Switchers in School-Grade-Subject-Year Level Data

<table>
<thead>
<tr>
<th>School</th>
<th>Grade</th>
<th>Subject</th>
<th>Year</th>
<th>Teachers</th>
<th>Mean Score</th>
<th>Mean Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1992</td>
<td>Smith, Hanushek, …</td>
<td>-.09</td>
<td>$15K</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1993</td>
<td>Smith, Hanushek, …</td>
<td>-.04</td>
<td>$17K</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1994</td>
<td>Smith, Hanushek, …</td>
<td>-.05</td>
<td>$16K</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1995</td>
<td>Black, Hanushek, …</td>
<td>0.01</td>
<td>$18K</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1996</td>
<td>Black, Hanushek, …</td>
<td>0.04</td>
<td>$17K</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1997</td>
<td>Black, Hanushek, …</td>
<td>0.02</td>
<td>$18K</td>
</tr>
</tbody>
</table>

Smith switches to a different school in 1995; Black replaces him
Impact of High Value-Added Teacher Entry on Cohort Test Scores

School-Grade-Cohort Mean Test Score

Year Relative to Entry of High Value-Added Teacher

Score in Current Grade
Impact of High Value-Added Teacher Entry on Cohort Test Scores

<table>
<thead>
<tr>
<th>School-Grade-Cohort Mean Test Score</th>
<th>Year Relative to Entry of High Value-Added Teacher</th>
<th>Score in Previous Grade</th>
<th>Score in Current Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Impact of High Value-Added Teacher Entry on Cohort Test Scores

\[ \Delta \text{Score} = 0.035 \quad (0.008) \]
\[ \Delta \text{TVA} = 0.042 \quad (0.002) \]

\[ p [\Delta \text{score} = 0] < 0.001 \]
\[ p [\Delta \text{score} = \Delta \text{TVA}] = 0.34 \]

Number of Events = 1135

School-Grade-Cohort Mean Test Score vs. Year Relative to Entry of High Value-Added Teacher

Score in Current Grade

Score in Previous Grade
Impact of High Value-Added Teacher Exit on Cohort Test Scores

$\Delta$ Score = -0.045 (0.008)
$\Delta$ TVA = -0.042 (0.002)

$p [\Delta \text{ score} = 0] < 0.001$
$p [\Delta \text{ score} = \Delta \text{ TVA}] = 0.66$

Number of Events = 1115

Year Relative to Departure of High Value-Added Teacher

Score in Current Grade

Score in Previous Grade
Impact of Low Value-Added Teacher Entry on Cohort Test Scores

<table>
<thead>
<tr>
<th>Year Relative to Entry</th>
<th>School-Grade-Cohort Mean Test Score</th>
<th>∆ Score</th>
<th>(Standard Error)</th>
<th>∆ TVA</th>
<th>(Standard Error)</th>
<th>p [∆ score = 0]</th>
<th>p [∆ score = ∆ TVA]</th>
<th>Number of Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>Score in Current Grade</td>
<td>-0.021</td>
<td>(0.007)</td>
<td>-0.033</td>
<td>(0.002)</td>
<td>&lt; 0.01</td>
<td>0.09</td>
<td>1148</td>
</tr>
<tr>
<td>-2</td>
<td>Score in Previous Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>Score in Current Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>Score in Previous Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Score in Current Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Score in Previous Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of Events = 1148
Impact of Low Value-Added Teacher Exit on Cohort Test Scores

\[ \Delta \text{Score} = 0.034 \]
\[ \Delta \text{TVA} = 0.034 \]

\[ p \left[ \Delta \text{score} = 0 \right] < 0.001 \]
\[ p \left[ \Delta \text{score} = \Delta \text{TVA} \right] = 0.99 \]

Number of Events = 1089
Teacher Switchers Design: Changes in Scores vs. Changes in Mean Teacher VA

Changes in Scores vs. Changes in Mean Teacher VA

Coef. = 0.974 (0.033)
Changes in Predicted Scores vs. Changes in Mean Teacher VA

Changes in Predicted Scores

Changes in Mean Teacher VA Value-Added

Coef. = 0.004 (0.005)
Changes in Other-Subject Scores vs. Changes in Mean Teacher VA
Middle Schools Only

Coef. = 0.038
(0.083)
# Estimates of Forecast Bias Using Parent Characteristics and Lagged Scores

<table>
<thead>
<tr>
<th>Dep. Var.:</th>
<th>Score in Year t</th>
<th>Pred. Score using Parent Chars.</th>
<th>Score in Year t</th>
<th>Pred. Score using Year t-2 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher VA</td>
<td>0.998</td>
<td>0.002</td>
<td>0.996</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.0057)</td>
<td>(0.0003)</td>
<td>(0.0057)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>Parent Chars. Controls</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>6,942,979</td>
<td>6,942,979</td>
<td>6,942,979</td>
<td>5,096,518</td>
</tr>
</tbody>
</table>
### Estimates of Forecast Bias with Alternative Control Vectors

<table>
<thead>
<tr>
<th>Control Vector</th>
<th>Quasi-Experimental Estimate of Bias (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>2.58 (3.34)</td>
</tr>
<tr>
<td>Student-level lagged scores</td>
<td>4.83 (3.29)</td>
</tr>
<tr>
<td>Non-score controls only</td>
<td>45.39 (2.26)</td>
</tr>
<tr>
<td>No controls</td>
<td>65.58 (3.73)</td>
</tr>
</tbody>
</table>
Rothstein result 1: Students are sorted into classrooms based on pre-determined variables such as grade $g-2$ test scores

- We confirm this result in our data

Rothstein result 2: Selection on observables is minimal conditional on grade $g-1$ controls

- Controlling for grade $g-2$ score does not affect VA estimates
- Consistent with our findings that VA does not predict $g-2$ score

$\Rightarrow$ Rothstein notes that his findings do not imply bias in VA estimates

- But they raise concerns about potential selection on unobservables
- Our quasi-experimental teacher switcher tests indicate that selection on unobservables turns out to be modest in practice
Part II: Long-Term Impacts
Fade-Out of Teachers’ Impacts on Test Scores in Subsequent Grades

Impact of Current Teacher VA on Test Scores

Years After Current School Year

Point Estimate

95% CI
Impacts on Outcomes in Adulthood

- Do teachers who raise test scores also improve long-term outcomes?

- Regress residualized long-term outcomes on teacher-level VA estimates

\[ Y_{it} = \alpha + \kappa \hat{m}_{jt} + \eta_{it} \]

- Then validate OLS estimates using cross-cohort switchers design

- Interpretation of these reduced-form coefficients [Todd and Wolpin 2003]

  - Impact of having better teacher, as measured by VA, for a single year during grades 4-8 on earnings

  - Includes benefit of better teachers, peers, etc. in later grades via tracking, as well as any complementarity with future teacher quality
College Attendance at Age 20 vs. Teacher Value-Added

Normalized Teacher Value Added ($\hat{m}_{jt}$)

Coef. = 0.82% (0.07)
Change in College Attendance Rate Across Cohorts vs. Change in Mean Teacher VA

Change in College Attendance Rate Across Cohorts (%)

Change in Mean Normalized Teacher VA Across Cohorts

Coef. = 0.86%

(0.23)
Event Study of Coefficients on College Attendance

Impact of 1 SD Change in Leads or Lags of Mean VA (%)

Coef. at 0 = 1.0 (0.3)

Coef. at +1 equals Coef. at 0: p=0.009

Coef. at -1 equals Coef. at 0: p=0.050
## Impacts of Teacher Value-Added on College Attendance

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>College at Age 20 (%)</th>
<th>College at Age 20 (%)</th>
<th>College at Age 20 (%)</th>
<th>College Quality at Age 20 ($)</th>
<th>College Quality at Age 20 ($)</th>
<th>College Quality at Age 20 ($)</th>
<th>High Quality College (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value-Added</td>
<td>0.82 (0.07)</td>
<td>0.71 (0.06)</td>
<td>0.74 (0.09)</td>
<td>298.63 (20.74)</td>
<td>265.82 (18.31)</td>
<td>266.17 (26.03)</td>
<td>0.72 (0.05)</td>
</tr>
<tr>
<td>Mean of Dep. Var.</td>
<td>37.22</td>
<td>37.22</td>
<td>37.09</td>
<td>26,837</td>
<td>26,837</td>
<td>26,798</td>
<td>13.41</td>
</tr>
<tr>
<td>Baseline Controls</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Parent Chars. Controls</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Lagged Score Controls</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Observations</td>
<td>4,170,905</td>
<td>4,170,905</td>
<td>3,130,855</td>
<td>4,167,571</td>
<td>4,167,571</td>
<td>3,128,478</td>
<td>4,167,571</td>
</tr>
</tbody>
</table>
Normalized Teacher Value Added ($m_{jt}$)

College Quality (Projected Earnings) at Age 20 vs. Teacher Value-Added

Coef. = $299 (21)$
Earnings at Age 28 vs. Teacher Value-Added

\[ \text{Normalized Teacher Value Added (\( m_{jt} \))} \]

\[ \begin{align*}
& \begin{array}{c}
-1.5 \\
-1 \\
-0.5 \\
0 \\
0.5 \\
1 \\
1.5
\end{array}
\end{align*} \]

\[ \text{Earnings at Age 28 ($)} \]

\[ \text{Coef.} = \$350 \]

\[ (92) \]
## Impacts of Teacher Value-Added on Earnings

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Earnings at Age 28 ($)</th>
<th>Earnings at Age 28 ($)</th>
<th>Earnings at Age 28 ($)</th>
<th>Working at Age 28 (%)</th>
<th>Total Income at Age 28 ($)</th>
<th>Wage growth Ages 22-28 ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher VA</td>
<td>349.84 (91.92)</td>
<td>285.55 (87.64)</td>
<td>308.98 (110.17)</td>
<td>0.38 (0.16)</td>
<td>353.83 (88.62)</td>
<td>286.20 (81.86)</td>
</tr>
<tr>
<td>Mean of Dep. Var.</td>
<td>21,256</td>
<td>21,256</td>
<td>21,468</td>
<td>68.09</td>
<td>22,108</td>
<td>11,454</td>
</tr>
<tr>
<td>Baseline Controls</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Parent Chars. Controls</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lagged Score Controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Observations</td>
<td>650,965</td>
<td>650,965</td>
<td>510,309</td>
<td>650,965</td>
<td>650,965</td>
<td>650,943</td>
</tr>
</tbody>
</table>
Women with Teenage Births vs. Teacher Value-Added

Coef. = -0.61% (0.06)
Neighborhood Quality at Age 28 vs. Teacher Value-Added

Percent College Graduates in ZIP at Age 28

Normalized Teacher Value Added ($m_{jt}$)

Coef. = 0.25% (0.04)
Retirement Savings at Age 28 vs. Teacher Value-Added

Normalized Teacher Value Added ($\hat{m}_{jt}$) with a Coefficient of 0.55% (0.16).
## Heterogeneity in Impacts of 1 SD of Teacher VA by Demographic Group

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>College Quality at Age 20 ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Girls (1)</td>
</tr>
<tr>
<td>Value-Added</td>
<td>290.65 (23.61)</td>
</tr>
<tr>
<td>Mean College Quality</td>
<td>27,584</td>
</tr>
<tr>
<td>Impact as % of Mean</td>
<td>1.05%</td>
</tr>
</tbody>
</table>
### Heterogeneity in Impacts of 1 SD of Teacher VA by Subject

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>College Quality at Age 20 ($)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Elementary School</td>
<td></td>
<td>Middle School</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Math Teacher</td>
<td>207.81</td>
<td>106.34</td>
<td>265.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value-Added</td>
<td>(21.77)</td>
<td>(28.50)</td>
<td>(43.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Teacher</td>
<td>258.16</td>
<td>189.24</td>
<td></td>
<td>521.61</td>
<td></td>
</tr>
<tr>
<td>Value-Added</td>
<td>(25.42)</td>
<td>(33.07)</td>
<td></td>
<td>(63.67)</td>
<td></td>
</tr>
<tr>
<td>Control for Average VA in Other Subject</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
Teacher Impacts by Grade

- Reduced-form impacts of having better teachers in each grade include tracking to better teachers in future grades.

- We can net-out the impact of tracking from the reduced-form coefficients by estimating tracking process.
  - Estimate impact of current teacher VA on VA of future teachers.
  - Subtract out impacts of future teachers.
Policy Proposal 1: Deselection of Low VA Teachers

What are the gains from replacing teachers with VA in bottom 5% with teachers of median quality (Hanushek 2009)?
Policy Calculations

- Use estimates to evaluate gains from improving teacher quality
- Measure impact of teacher VA on present value of lifetime earnings

Assumptions

- Ignore general equilibrium effects and non-monetary gains [Oreopoulos and Salvanes 2011, Heckman 2000]
- Constant percentage impact on earnings over life
- Life-cycle earnings follows cross-sectional life-cycle path in 2010
- 2% wage growth with 5% discount rate back to age 12
  - Undiscounted lifetime earnings gains are roughly 5 times larger
Consider replacing teachers in the bottom 5% of VA distribution with teachers of average quality (Hanushek 2009)

Select on true VA $\rightarrow$ NPV gain for a class of average size: $407,000$

In practice, gains are reduced by two factors

- Estimation error in VA
- Drift in VA over time
Deselecting Teachers on the Basis of Value-Added

Teacher Effect on Test Scores

- Population
- Observed Below 5th Percentile After 3 Years
Earnings Impact in First Year After Deselection Based on Estimated VA

Present Value Gain from Deselection on True VA = $406,988

Graph showing lifetime earnings gain per class in $1000s vs. number of years used to estimate VA.
Deselection Based on Estimated VA After 3 Years: Earnings Impacts in Subsequent Years

Average 10 Year Gain = $184,234
Average 10 Year Gain = $184,234

Average 10 Year Gain = $246,744

Earnings Impact Over Time

Lifetime Earnings Gain Per Class ($1000s)

School Years Since Teacher Was Hired

Deselected on Estimated VA in Year 4

Deselected on True VA in Year 4

Average 10 Year Gain = $184,234

Average 10 Year Gain = $246,744
Rothstein (2013) estimates that deselecting bottom 5% of teachers based on VA would require a salary increase of $700 for all teachers.

Avg. gain from deselection policy is $184,000 \times 5\% = $9,250

Gain 10 times as large as cost $\Rightarrow$ VA could be a useful policy tool

Key concern: gains may be eroded when VA is actually used

- Using VA in high-stakes evaluation could lead to teaching to the test or cheating [Jacob 2005, Neal and Schanzenbach 2010, Barlevy and Neal 2012]

Broader policy lesson: improving teacher quality, whether through VA or other metrics, likely to have very large returns
Policy Proposal 2: Retention of High VA Teachers

What are the gains from increasing retention of high value-added teachers by paying salary bonuses?
Retaining a teacher whose VA is at the 95\textsuperscript{th} percentile (based on 3 years of data) for an extra year would generate PV earnings gain of $266K.

Clotfelter et al. (2008) analyze impacts of bonus payments to teachers.

$1,800 bonus would raise teacher retention by 1.5 percentage points \rightarrow earnings gain of $3,200.

Net return relatively small because most of the bonus payments go to teachers who would not have left anyway.

Have to pay bonuses to 60 teachers to retain 1 teacher on average.
Conclusion

- Further work needed to assess value-added as a policy tool

- Using VA measures in high-stakes evaluation could induce negative behavioral responses such as teaching to the test or cheating

- Errors in personnel decisions must be weighed against mean benefits

- Results highlight large potential returns from developing policies to improve teacher quality

- From a purely financial perspective, parents should be willing to pay about $7,000/year to get a 1 SD higher VA teacher for their child
Appendix Figures
Rankings of Colleges Based on Earnings at Ages 23 and 27 vs. Age 32
Correlation of College Rankings Based on Earnings at Age 32
With Rankings Based on Earnings at Earlier Ages

Correlation with Ranking Based on Earnings at Age 32

Age at Which Earnings are Measured
Correlation of Individual Earnings with Earnings 12 Years Later, by Age
College Attendance

Percent Attending College at Age 20

Table:

<table>
<thead>
<tr>
<th>Test Score (SD)</th>
<th>No Controls</th>
<th>With Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Graph:

- Line for No Controls
- Line for With Controls

Test Score (SD) vs. Percent Attending College at Age 20
Teenage Birth

Percent of Females with Teenage Birth

- No Controls
- With Controls

Test Score (SD)

Percentage of Jacob and Levitt (2003) Outliers

Percentile of Teacher Value-Added

Percentage of Jacob and Levitt (2003) Outliers