Making Teaching Last: Long- and Short-Run Value-Added

Michael Gilraine
New York University

Nolan Pope
University of Maryland

December 3, 2020

Motivation

- Measuring the impact of teachers is a longstanding preoccupation in applied research
  - Natural, given vital role teachers play in education production
- Teacher value-added (VA) estimators have been predominant method to measure teachers' overall quality
  - See Rockoff (2004), Rivkin, Hanushek and Kain (2005), Kane and Staiger (2008), Chetty, Friedman and Rockoff (2014a,b)
- VA estimators measure teacher quality based on the contemporaneous test score growth of their students

Measuring Educational Output

- Ultimately, policymakers and parents are interested in capturing teachers' contributions to long-term outcomes
  - e.g., graduation, college-going, earnings, crime, etc.
- Test scores used as sufficient statistic
  - Test scores have the benefit of being readily-available and contemporaneous
- Policymakers would like to avoid measuring temporary impacts of teachers (e.g. "teaching to the test") and just measure the impact of teachers that persists

Our Paper

1. We develop two new VA estimators that separately capture teachers' contribution to long-term and short-term knowledge
   - i.e., portion of test score gains that is *temporary* and the portion that *persists*
   - To do so, we remove transitory knowledge gains from the standard VA calculation
2. Estimate effect of having a higher quality teacher along both of these measures on long-run outcomes

What We Do

- Separate standard VA into two components:
  - Long-run VA: test score gains that persist to following year
  - Short-run VA: test score gains that do *not* persist
- Our two estimators calculate VA using:
  - Long-run VA: period $t + 1$ test score residuals as outcome
  - Short-run VA: period t *minus* period $t + 1$ test score residuals as outcome
- VA estimators are easy-to-implement using standard VA estimation techniques

Preview of Results I

- Relative contributions of long-run and short-run components to standard VA:
  - Long-run VA: 40 percent
  - Short-run VA: 60 percent
- Students assigned to teachers whose long-run and short-run VA is one unit higher:
  - Long-run VA: Score one SD higher on *next* year's test
  - Short-run VA: Do not score any higher on *next* year's test
- Long-run and short-run VA are negatively correlated
- Long-run VA is positively correlated with non-cognitive VA
  - Whereas short-run VA is not

Preview of Results II

- Students assigned to high long-run VA teachers fare *substantially* better in terms of long-run outcomes
  - Students assigned to high short-run VA teachers do not
  - Long-run outcomes: PSAT scores, SAT taking and scores, Graduation, HS GPA, etc.
- Find no evidence of bias in long- and short-run VA
  - Forecast unbiasedness
  - No sorting on omitted observables
  - No bias in quasi-experiments leveraging teacher turnover
- Policies based on long-run VA greatly increase students' lifelong success (in comparison to standard VA)
  - Benchmark release bottom 5% teacher policy: increase policy effectiveness two-fold

Outline

Relation to Prior Literature

- **Very** large number of papers estimating teacher VA
  - To name a **few**: Rockoff (2004), Rivkin, Hanushek, and Kain (2005), Kane and Staiger (2008), Kane, Rockoff, and Staiger (2008), Jacob and Lefgren (2008), Harris and Sass (2014), Chetty, Friedman, and Rockoff (2014a,b), Feld, Salamanca and Zölitz (forthcoming)
- Large focus on assessing whether teacher VA estimates are biased (whether students are sorted to teachers based on unobservables)
  - Rothstein (2010, 2017); Chetty *et al.* (2017); Kane and Staiger (2008); Kane *et al.* (2013); Bacher-Hicks *et al.* (2014)
- Recent literature focusing on non-test score metrics to capture other dimensions of learning
  - Jackson (2018), Petek and Pope (2018), Rose, Schellenberg, and Shem-Tov (2019)

Standard VA Model

- Test scores are given by:

$$A^*_{ijt} = \beta X_{ijt} + \mu_{jt} + \epsilon_{ijt}\,, \quad i = 1, 2, ..., n_{jt}$$

where:

- i=student, j=teacher, t=year
- A^*_{ijt}: student i's test score in year t
- μ_{jt}: teacher j's contribution
- X_{ijt}: observable determinants of student achievement
    - e.g., lagged test scores, student demographics, etc.
- n_{jt}: class size
- ϵ_{ijt}: student-specific error term

- Remove effect of observable characteristics $A_{ijt} = A^*_{ijt} - \hat{\beta} X_{ijt}$:

$$A_{ijt} = \mu_{jt} + \epsilon_{ijt}$$

Estimation

- Our parameter of interest is teacher's contribution to period t test scores, μ_{jt}:

$$A_{ijt} = \mu_{jt} + \epsilon_{ijt}$$

- **Assumption:** Students are not sorted to teachers based on unobservable determinants of student achievement
  - Commonly-invoked assumption in VA literature
  - See Rothstein (2010, 2017); Chetty *et al.* (2017)

- μ_{jt} can be estimated via fixed effects, although more commonly via VA estimators that:
  - Reduce noise by shrinking less reliable estimates to mean
    - Via parametric (Kane and Staiger 2008) or nonparametric empirical Bayes (Gilraine, Gu and McMillan 2020)
  - Jack-knife by leaving out data from year t in estimation
    - Else regressing year t outcomes on VA introduces same estimation errors on left and right hand side, causing bias

Long- and Short-Run VA Model

- Suppose some types of knowledge persist while others do not
  - Jacob, Lefgren and Sims (2010); Cascio and Staiger (2012)
- Specifically, knowledge consists of two components:
  1. Short-term knowledge that does not persist
     - e.g., rote memorization or "teaching to the test"
  2. Long-term knowledge that persists fully to next period
- Teacher j's contribution to knowledge, μ_{jt}, is then divided into these two components:

$$\mu_{jt} = \mu_{jt}^{L} + \mu_{jt}^{S}$$

where:

  - μ_{jt}^{L}: teacher j's contribution to long-term knowledge
  - μ_{jt}^{S}: teacher j's contribution to short-term knowledge

Model Goal

- Our VA model becomes:

$$A_{ijt} = \mu_{jt}^{L} + \mu_{jt}^{S} + \epsilon_{ijt}$$

- Standard VA estimators thus estimate the sum of teacher j's contribution to long- and short-term knowledge, $\mu_{jt}^{L} + \mu_{jt}^{S}$
- Our **goal** is to separate teacher j's contribution to each component:
    - μ_{jt}^{L}: 'long-run' VA
    - μ_{jt}^{S}: 'short-run' VA

Achievement in Next Period

- Use test scores in $t+1$:

$$A_{ijk,t+1} = \delta^L \mu_{jt}^L + \delta^S \mu_{jt}^S + \mu_{k,t+1} + \epsilon_{ijk,t+1}$$

  - $\mu_{k,t+1}$: contribution of teacher k to student i at time $t+1$
  - δ^L: fade out of long-term knowledge between t and $t+1$
  - δ^S: fade out of short-term knowledge between t and $t+1$

- **Assumption:** Short-term component of knowledge perfectly fades out (i.e., $\delta^S = 0$)
- **Assumption:** Long-term component of knowledge persists fully to next period (i.e., $\delta^L = 1$)

Estimation of Long-Run VA

- Achievement in $t+1$ is now given by:

$$A_{ijk,t+1} = \mu_{jt}^{L} + \mu_{k,t+1} + \epsilon_{ijk,t+1}$$

- **Key Assumption:** Students are not sorted to teachers in $t+1$ based on unobservable determinants of student achievement or teacher assignment in t
    - Analogous to usual VA assumption, but must hold for two periods (rather than one)

- Model becomes:

$$A_{ijk,t+1} = \mu_{jt}^{L} + \tilde{\epsilon}_{ijk,t+1}$$

    - Where $\tilde{\epsilon}_{ijk,t+1} = \mu_{k,t+1} + \epsilon_{ijk,t+1}$

- μ_{jt}^{L} can be estimated using standard VA techniques
    - Can estimate with teacher k fixed effects to account for $\mu_{k,t+1}$

Estimation of Short-Run VA

- Given our equations, we can write *difference* in achievement from t to $t+1$ as:

$$A_{ijt} - A_{ijk,t+1} = \mu^S_{jt} + \breve{\epsilon}_{ijkt} \, .$$

  - Where $\breve{\epsilon}_{ijt} \equiv \epsilon_{ijt} - \tilde{\epsilon}_{ijk,t+1}$

- μ^S_{jt} can be estimated usings standard VA techniques

Summary of Long- and Short-Run VA

Have constructed three different VA measures:

1. **Standard Value-Added:** $A_{ijt} = \mu_{jt} + \epsilon_{ijt}$
   - Teacher j's contribution to test scores in year t
2. **Long-Run Value-Added:** $A_{ijk,t+1} = \mu^{L}_{jt} + \tilde{\epsilon}_{ijk,t+1}$
   - Teacher j's contribution to test scores in year t that *persist* to year $t+1$
3. **Short-Run Value-Added:** $A_{ijt} - A_{ijk,t+1} = \mu^{S}_{jt} + \breve{\epsilon}_{ijkt}$
   - Teacher j's contribution to test scores in year t that *do not persist* to year $t+1$

- All three measures constructed using standard VA techniques
  - Will use jack-knifed parametric empirical Bayes estimates that allow for drift (as in Chetty *et al.* 2014a,b)

Data

Estimate VA using detailed administrative data:

- Large Urban District
  - Cover grades 3-5 from 2003-04 through 2011-12
  - Minor sample restrictions (e.g., must have valid student-teacher match)
  - 1.2 million student-year observations
- Data contain test scores as well as behavioral outcomes we use to construct non-cognitive VA
  - Behavioral outcomes: 'effort' GPA, GPA, suspensions, absences, and grade repetition
- Data also include detailed student demographics
  - e.g., race, gender, FRPM status, EL status, parental education

Summary Statistics

	Full Sample	Value-Added Sample
Cognitive Outcomes:		
Math Score (σ)	0.00	0.06
Reading Score (σ)	0.00	0.05
Non-Cognitive Outcomes:		
Log Days Absent	1.50	1.51
GPA	2.88	2.90
% Suspended	2.26	2.18
% Repeating Grade	0.65	0.49
Demographics:		
% Hispanic	74.3	75.6
% Black	10.1	9.2
% White	8.9	8.6
% Free or Reduced Price Lunch	69.2	70.8
% Parent is High School Dropout	35.6	35.9
% Parent is College Graduate	19.6	19.1
# of Students	649,694	551,814
# of Teachers	15,155	12,964
Observations (student-year)	1,452,367	1,186,858

Summary Statistics

	Full Sample	Value-Added Sample
Cognitive Outcomes:		
Math Score (σ)	0.00	0.06
Reading Score (σ)	0.00	0.05
Non-Cognitive Outcomes:		
Log Days Absent	1.50	1.51
GPA	2.88	2.90
% Suspended	2.26	2.18
% Repeating Grade	0.65	0.49
Demographics:		
% Hispanic	74.3	75.6
% Black	10.1	9.2
% White	8.9	8.6
% Free or Reduced Price Lunch	69.2	70.8
% Parent is High School Dropout	35.6	35.9
% Parent is College Graduate	19.6	19.1
# of Students	649,694	551,814
# of Teachers	15,155	12,964
Observations (student-year)	1,452,367	1,186,858

Summary Statistics

	Full Sample	Value-Added Sample
Cognitive Outcomes:		
Math Score (σ)	0.00	0.06
Reading Score (σ)	0.00	0.05
Non-Cognitive Outcomes:		
Log Days Absent	1.50	1.51
GPA	2.88	2.90
% Suspended	2.26	2.18
% Repeating Grade	0.65	0.49
Demographics:		
% Hispanic	74.3	75.6
% Black	10.1	9.2
% White	8.9	8.6
% Free or Reduced Price Lunch	69.2	70.8
% Parent is High School Dropout	35.6	35.9
% Parent is College Graduate	19.6	19.1
# of Students	649,694	551,814
# of Teachers	15,155	12,964
Observations (student-year)	1,452,367	1,186,858

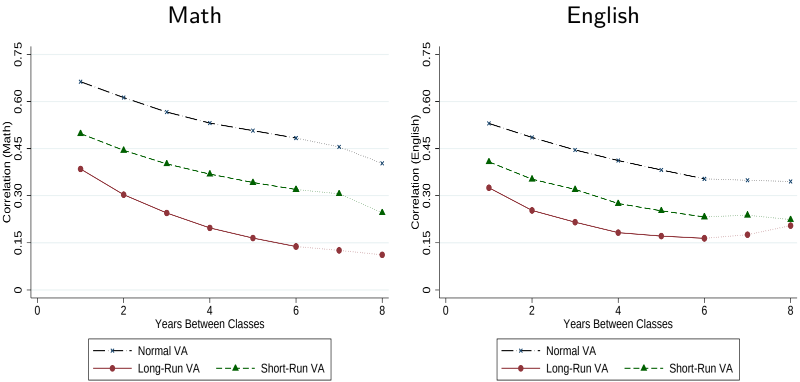
Model Estimates

- Estimate standard VA along with its long- and short-run components
  - Estimates for standard VA near-identical to those in Bacher-Hicks, Kane and Staiger (2014)
- Variance decomposition yields that variation in standard VA comes from:
  - Long-run VA: 40%
  - Short-run VA: 60%

Variance Estimates of Long- and Short-Run VA

	Standard VA	Long-Run VA	Short-Run VA
Estimates of teacher SD (Math)			
Lower bound based on lag 1	0.270	0.168	0.230
Quadratic estimate	0.282	0.194	0.245
Estimates of teacher SD (English)			
Lower bound based on lag 1	0.175	0.115	0.144
Quadratic estimate	0.185	0.135	0.158
Student-Year Observations	1,186,858	1,067,681	1,066,320

Autocorrelation for Math and English Scores



Correlation of Teacher Value-Added Measures

VA Measure	Standard VA	Long-Run VA	Short-Run VA	Non-Cognitive VA
Normal VA	1			
Long-Run VA	0.524	1		
Short-Run VA	0.745	-0.164	1	
Non-Cognitive VA	0.113	0.180	-0.010	1

Correlation of Teacher Value-Added Measures

VA Measure	Standard VA	Long-Run VA	Short-Run VA	Non-Cognitive VA
Normal VA	1			
Long-Run VA	$\boxed{0.524}$	1		
Short-Run VA	$\boxed{0.745}$	-0.164	1	
Non-Cognitive VA	0.113	0.180	-0.010	1

Correlation of Teacher Value-Added Measures

VA Measure	Standard VA	Long-Run VA	Short-Run VA	Non-Cognitive VA
Normal VA	1			
Long-Run VA	0.524	1		
Short-Run VA	0.745	-0.164	1	
Non-Cognitive VA	0.113	0.180	-0.010	1

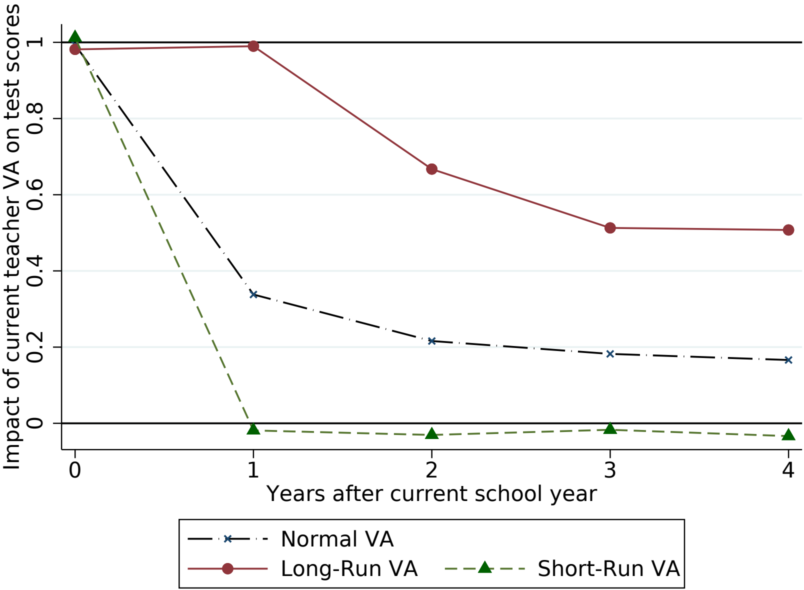
Correlation of Teacher Value-Added Measures

VA Measure	Standard VA	Long-Run VA	Short-Run VA	Non-Cognitive VA
Normal VA	1			
Long-Run VA	0.524	1		
Short-Run VA	0.745	-0.164	1	
Non-Cognitive VA	0.113	0.180	-0.010	1

Short-Term Test Score Impacts

- Regress current and future test scores (residualized) on VA measures
- Under our model, coefficients in year t should be 1 for all VA measures
- Similarly, in year $t+1$ coefficients should equal:
  - Short-run VA: 0
  - Long-run VA: 1
- Test scores for subsequent periods indicate fade out
  - Expect fade out to be lower for long-run VA

Effects of Short- and Long-Run Value-Added on Future Test Scores



Long-Term Effects of VA

- Estimate long-term effects of teacher VA using method in Chetty *et al.* (2014b)
- Identify impact of teacher VA on long-term outcome Y_i^*:
  1. Remove the effect of observable characteristics:

     $$Y_i = Y_i^* - \beta^Y X_{it}$$

     - Coefficients β^Y estimated using variation across students taught by same teacher to account for students sorting to teachers (e.g., $Y_i^* = \alpha_j + \beta^Y X_{it}$)
  2. Estimate (linear) relationship between long-term outcome residuals and teacher VA in school year t:

     $$Y_{it} = a + \kappa_g m_{jt} + \eta_{ijt}$$

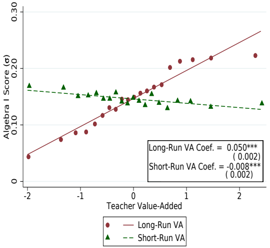
     - Where $m_{jt} \equiv \mu_{jt}/\sigma_\mu$ is teacher j's VA scaled in standard deviation units of teacher VA distribution

Visualizing Long-Term Effects of VA

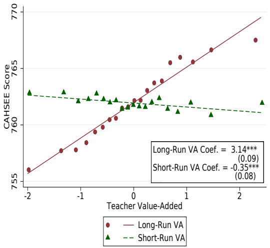
- Will display numerous figures showing impact of long- and short-run VA on various long-term outcomes
- Figures constructed in three steps:
  1. Residualize long-run outcome using variation across students taught by same teacher
  2. Divide scaled VA estimates, m_{jt}, into twenty equal-sized bins and plot mean of long-run outcome residuals in each bin against the corresponding bin mean of m_{jt}
  3. Add back in mean long-run outcome to facilitate interpretation of the scale

Effect of Short- and Long-Run VA on Long-Run Outcomes I

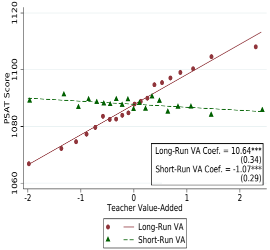
Algebra I



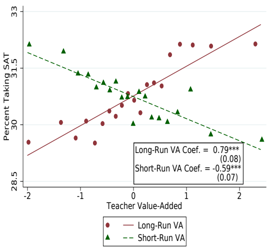
HS Exit Exam



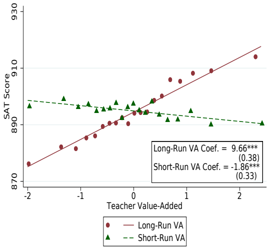
PSAT



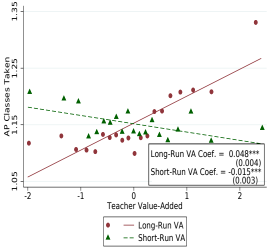
Took SAT



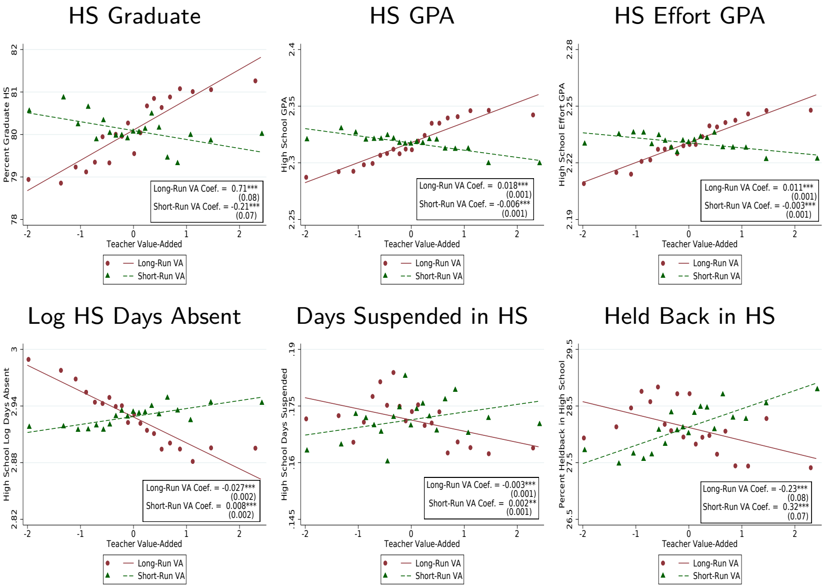
SAT



AP Classes Taken



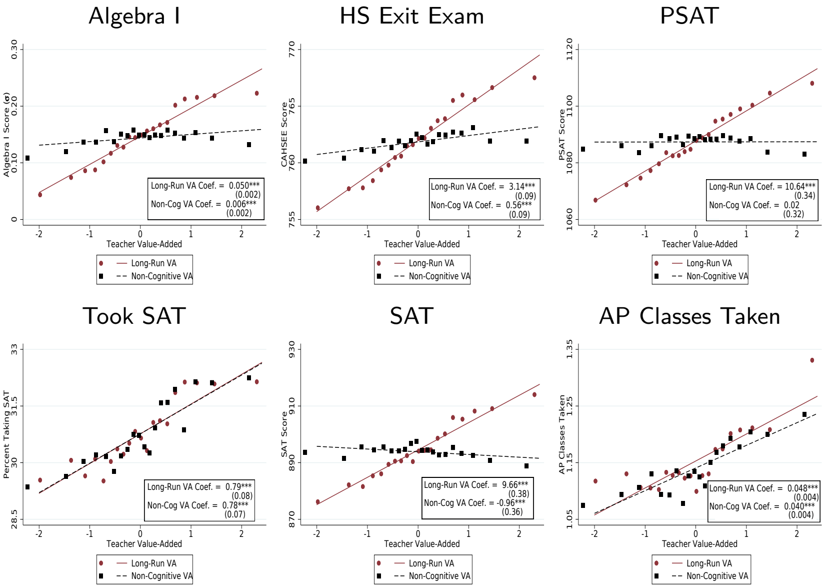
Effect of Short- and Long-Run VA on Long-Run Outcomes II



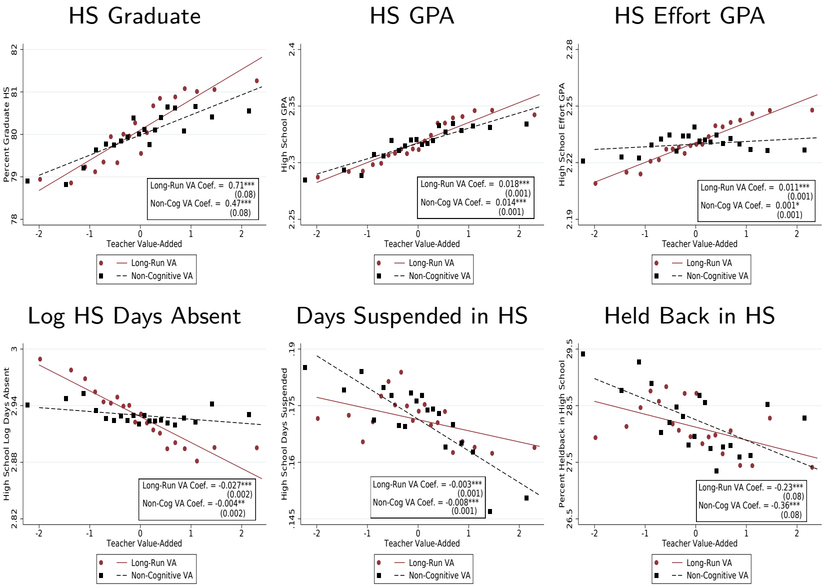
Long-Run VA and Long-Run Outcomes

- Figures indicate that long-run VA dramatically improves long-run outcomes, whereas short-run VA does not
  - Short-run VA often causes *declines* in long-run outcomes
- *Entirety* of VA's impact on long-run outcomes driven by long-run VA
- Prior research highlights that non-cognitive VA independently predicts long-run outcomes
  - Can even be better predictor for behavioral outcomes (e.g., GPA, absences, etc.)
- Can long-run VA outperform non-cognitive VA?

Effect of Non-Cognitive VA and Long-Run VA on Long-Run Outcomes I



Effect of Non-Cognitive VA and Long-Run VA on Long-Run Outcomes II



Long-Run Effects of VA: Multivariate Analysis

- Add all three measures into multivariate model
  - Except standard VA as (effectively) collinear with long- and short-run VA
- Checks whether more than one VA measure independently affects long-term outcomes
- Estimate (linear) relationship between long-term outcome residuals and multivariate teacher VA:

$$Y_{it} = a + \kappa_g^L m_{jt}^L + \kappa_g^S m_{jt}^S + \kappa_g^{\textsf{non-cog}} m_{jt}^{\textsf{non-cog}} + \eta_{ijt}$$

Long-Term Effects of Multivariate VA

Outcome:	Test-Based Outcomes: Exit Exam	Test-Based Outcomes: PSAT Score	Test-Based Outcomes: SAT Score	Behaviour-Based Outcomes: Graduated HS (%)	Behaviour-Based Outcomes: HS GPA	Behaviour-Based Outcomes: Took SAT (%)
Long-run VA (m_{jt}^{L})	3.08*** (0.10)	10.83*** (0.35)	9.75*** (0.40)			
Short-run VA (m_{jt}^{S})						
Non-Cognitive VA ($m_{jt}^{\textit{non-cog}}$)						
Sample Mean	762.4	1087.2	896.2	80.1	2.32	30.6
Sample SD	41.4	153.3	119.2	-	0.78	-

Long-Term Effects of Multivariate VA

	Test-Based Outcomes			Behaviour-Based Outcomes		
Outcome:	Exit Exam	PSAT Score	SAT Score	Graduated HS (%)	HS GPA	Took SAT (%)
Long-run VA (m^{L}_{jt})	3.08*** (0.10)	10.83*** (0.35)	9.75*** (0.40)			
Short-run VA (m^{S}_{jt})	0.17* (0.09)	0.73** (0.30)	-0.10 (0.34)			
Non-Cognitive VA ($m^{\textit{non-cog}}_{jt}$)						
Sample Mean	762.4	1087.2	896.2	80.1	2.32	30.6
Sample SD	41.4	153.3	119.2	-	0.78	-

Long-Term Effects of Multivariate VA

	Test-Based Outcomes			Behaviour-Based Outcomes		
Outcome:	Exit Exam	PSAT Score	SAT Score	Graduated HS (%)	HS GPA	Took SAT (%)
Long-run VA (m_{jt}^{L})	3.08***	10.83***	9.75***			
	(0.10)	(0.35)	(0.40)			
Short-run VA (m_{jt}^{S})	0.17*	0.73**	-0.10			
	(0.09)	(0.30)	(0.34)			
Non-Cognitive VA ($m_{jt}^{\textit{non-cog}}$)	0.10	-1.80***	-2.61***			
	(0.09)	(0.33)	(0.36)			
Sample Mean	762.4	1087.2	896.2	80.1	2.32	30.6
Sample SD	41.4	153.3	119.2	-	0.78	-

Long-Term Effects of Multivariate VA

	Test-Based Outcomes			Behaviour-Based Outcomes		
Outcome:	Exit Exam	PSAT Score	SAT Score	Graduated HS (%)	HS GPA	Took SAT (%)
Long-run *VA* (m^L_{jt})	3.08***	10.83***	9.75***	0.63***	0.015***	0.53***
	(0.10)	(0.35)	(0.40)	(0.09)	(0.001)	(0.08)
Short-run *VA* (m^S_{jt})	0.17*	0.73**	-0.10	-0.08	-0.003**	-0.45***
	(0.09)	(0.30)	(0.34)	(0.07)	(0.001)	(0.07)
Non-Cognitive *VA* ($m^{\text{non-cog}}_{jt}$)	0.10	-1.80***	-2.61***	0.37***	0.011***	0.69***
	(0.09)	(0.33)	(0.36)	(0.08)	(0.001)	(0.08)
Sample Mean	762.4	1087.2	896.2	80.1	2.32	30.6
Sample SD	41.4	153.3	119.2	-	0.78	-

Model Validity

- Key to VA is assumption that students are not sorted to teachers based on unobservables
- Three tests in literature:
  1. "Forecast unbiasedness"
     - i.e., teachers whose estimated VA is one unit higher cause students' test scores to increase by one unit on average
  2. Check for student sorting to high-VA teachers based on (omitted) observables
     - Omitted observable: twice-lagged test scores
  3. Leverage quasi-experimental variation in VA from teacher turnover

Validity Test 1: Forecast Unbiasedness

- "Forecast unbiasedness:" Regressing (residualized) test scores (A_{ijt}) on teacher VA (μ_{jt}) should yield coefficient of 1
- Under our model, should also get coefficient of 1 for long- and short-run VA
- Our fade out assumptions also imply that regressing (residualized) $t+1$ test scores ($A_{ijk,t+1}$) should yield:
    - Long-run VA: coefficient of 1
    - Short-run VA: coefficient of 0

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

Period:	t	$t+1$	$t+2$	$t+3$
Standard VA	1.011	0.349	0.228	0.157
(s.e.)	(0.006)	(0.006)	(0.006)	(0.005)
[95% CI]	[0.999,1.023]	[0.338,0.361]	[0.216,0.239]	[0.147,0.167]
Long-Run VA	0.978	1.021	0.697	0.471
(s.e.)	(0.013)	(0.011)	(0.012)	(0.011)
[95% CI]	[0.952,1.005]	[0.999,1.043]	[0.673,0.722]	[0.450,0.493]
Short-Run VA	1.018	-0.008	-0.022	-0.011
(s.e.)	(0.009)	(0.009)	(0.008)	(0.007)
[95% CI]	[1.000,1.035]	[-0.025,0.008]	[-0.037,-0.006]	[-0.024,0.002]
Observations	1,123,490	1,010,987	833,941	756,557

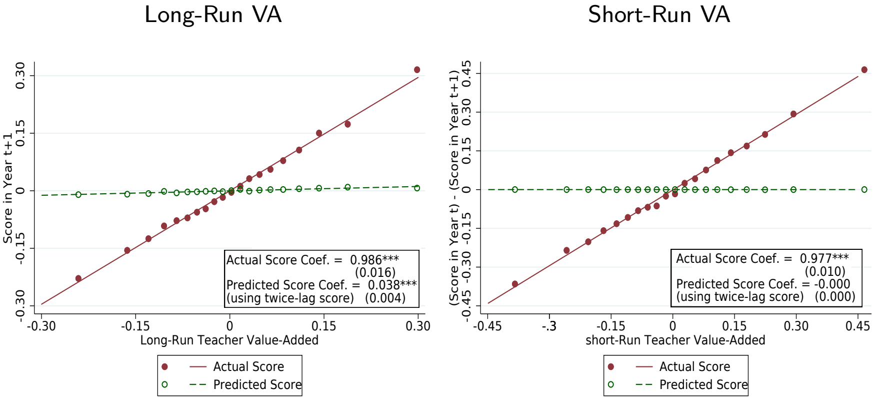
Impacts of Teacher Value-Added Measures on Current and Future Test Scores

Period:	t	$t+1$	$t+2$	$t+3$
Standard VA	1.011	0.349	0.228	0.157
(s.e.)	(0.006)	(0.006)	(0.006)	(0.005)
[95% CI]	[0.999,1.023]	[0.338,0.361]	[0.216,0.239]	[0.147,0.167]
Long-Run VA	0.978	1.021	0.697	0.471
(s.e.)	(0.013)	(0.011)	(0.012)	(0.011)
[95% CI]	[0.952,1.005]	[0.999,1.043]	[0.673,0.722]	[0.450,0.493]
Short-Run VA	1.018	-0.008	-0.022	-0.011
(s.e.)	(0.009)	(0.009)	(0.008)	(0.007)
[95% CI]	[1.000,1.035]	[-0.025,0.008]	[-0.037,-0.006]	[-0.024,0.002]
Observations	1,123,490	1,010,987	833,941	756,557

Validity Test 2: No Sorting on (Omitted) Observables

- Our key assumption is that students are not sorted to teachers based on unobservables
- Untestable, but can check for non-random sorting in observables
- We control for a lot of observables; one not (typically) controlled for is twice-lagged scores
- Plot this visually:
    1. Divide long- and short-run VA into twenty equal sized bins
    2. Plot mean value of residualized test scores (A_{ijt}) in each bin

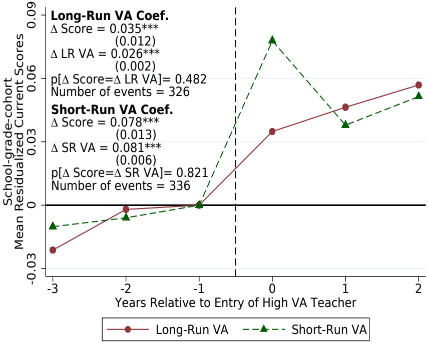
Effects of Short- and Long-Run VA on Actual and Predicted Scores



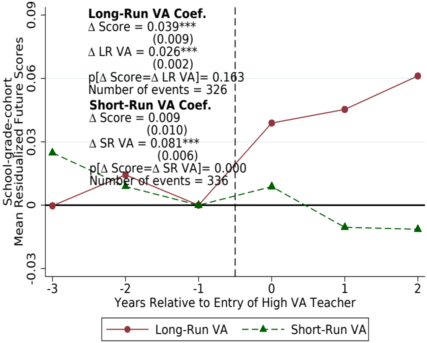
Validity Test 3: Quasi-Experimental Event Studies

Impacts of **High** VA Teacher Entry on Current and Future Scores

Current Test Scores



Future Test Scores (i.e., Scores in $t+1$)

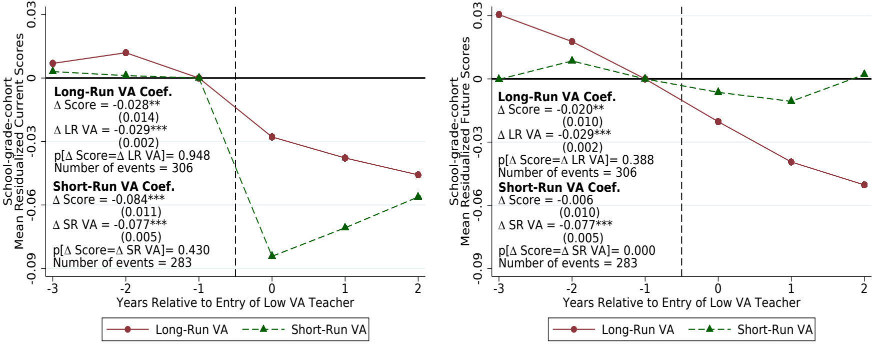


Validity Test 3: Quasi-Experimental Event Studies

Impacts of **Low** VA Teacher Entry on Current and Future Scores

Current Test Scores

Future Test Scores (i.e., Scores in $t+1$)



Validity Test 4: Quasi-Experimental Variation

- **Idea:** Estimate teachers' long-term impacts using quasi-experiment that compares changes in test scores across consecutive cohorts to changes in mean VA
- Formally, regress:

$$\Delta Y_{sgt} = \alpha + \kappa \Delta Q_{sgt} + \Delta \eta_{sgt}$$

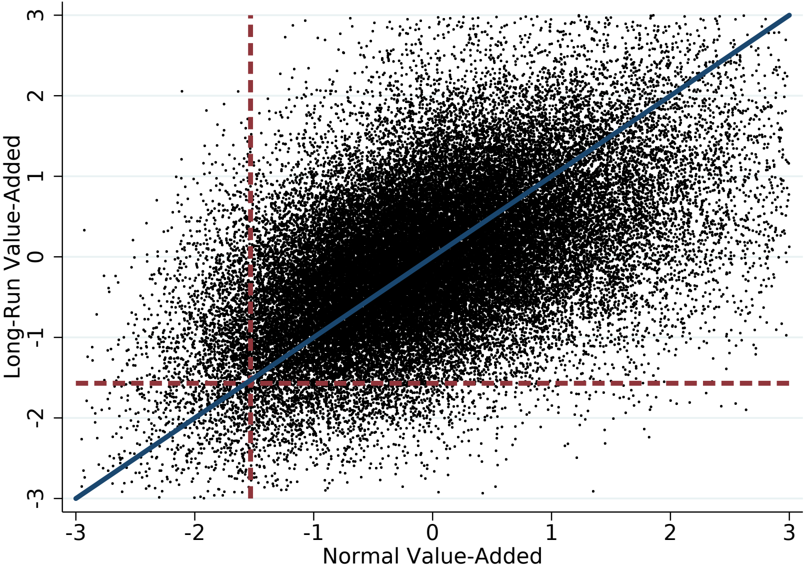
where:

  - s=school, g=grade, t=year
  - ΔY_{sgt}: change in student outcome residuals in school-grade $s-g$
  - ΔQ_{sgt}: change in mean teacher VA in school-grade $s-g$
    - $\Delta Q_{sgt} = Q_{sgt} - Q_{sg,t-1}$, where Q_{sgt} is jack-knifed VA estimate that omits years t and $t-1$
    - Ensures variation in ΔQ_{sgt} driven by staffing changes and not changes in teachers' VA estimates

Policy Implications

- **Benchmark policy:** replace teachers in bottom 5% of VA distribution with mean quality teachers
  - Policy proposed by Hanushek (2009, 2011) and evaluated by Chetty *et al.* (2014b)
- Target teachers based on long-run VA, given its superiority at improving long-run outcomes
  - Presumably should never target short-run VA
- Correlation between standard VA and long-run VA is 0.52, suggesting targeting long-run VA releases different teachers
  - In turn improving policy efficiency

Two-Dimensional Cross Teacher Value-Added Plot



Policy Calculations

- 70% of teachers released under policy using standard VA are not released when long-run VA used
- Calculate policy gains (G^Y) using:

$$G^Y = \Delta m_\sigma \times \kappa^Y$$

  where:
  - Δm_σ: average change in teacher VA caused by releasing bottom 5% teacher
    - If $m_{jt} \sim \mathcal{N}(0, \sigma_\mu)$, then $\Delta m_\sigma = 2.06$
  - κ^Y: estimated increase in long-run outcome Y from one standard deviation increase in VA
- Estimated policy gains using standard VA vs. long-run VA:

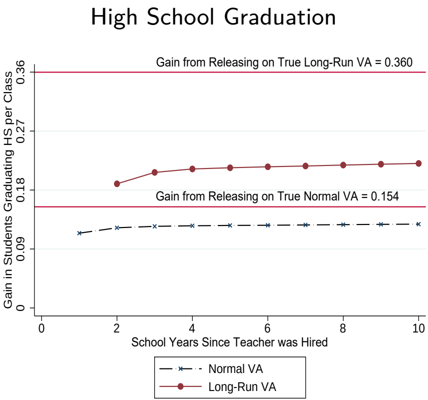
Policy Gains: Standard VA vs. Long-Run VA

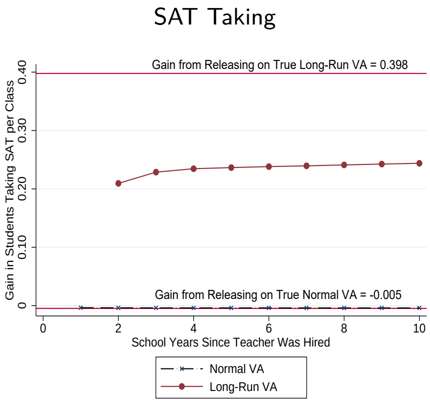
Long-Run Outcome:	Algebra Score	SAT Score	Graduated HS (%)	HS GPA	Took SAT	Log Days Absent
Sample Mean	0.150	894.2	80.1	2.32	29.0	2.92
Average Change in VA of Released Teachers (Δm_{σ}^{PEB})	2.06	2.06	2.06	2.06	2.06	2.06
A. Standard Value-Added						
Benefit ($\hat{\kappa}^{Y}$)	0.026	4.81	0.26	0.006	-0.03	-0.010
Gain of Releasing Bottom 5% (G)	0.054	9.92	0.54	0.012	-0.06	-0.021
B. Long-Run Value-Added						
Benefit ($\hat{\kappa}^{Y}$)	0.048	9.08	0.68	0.017	0.58	-0.026
Gain of Releasing Bottom 5% (G)	0.100	18.73	1.40	0.035	1.20	0.054
Policy Gain Increase from Using Long-Run VA (%)	85.2	88.8	161.5	183.3	∞	160.0

Policy Gains: Standard VA vs. Long-Run VA

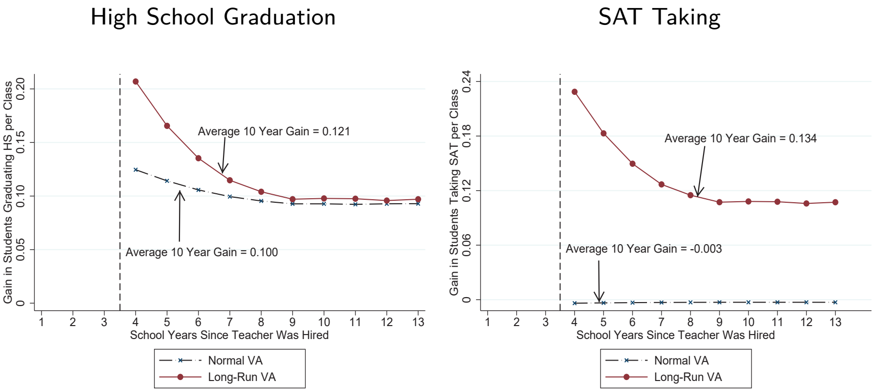
Long-Run Outcome:	Algebra Score	SAT Score	Graduated HS (%)	HS GPA	Took SAT	Log Days Absent
Sample Mean	0.150	894.2	80.1	2.32	29.0	2.92
Average Change in VA of Released Teachers (Δm_{σ}^{PEB})	2.06	2.06	2.06	2.06	2.06	2.06
A. Standard Value-Added						
Benefit ($\hat{\kappa}^{Y}$)	0.026	4.81	0.26	0.006	-0.03	-0.010
Gain of Releasing Bottom 5% (G)	0.054	9.92	0.54	0.012	-0.06	-0.021
B. Long-Run Value-Added						
Benefit ($\hat{\kappa}^{Y}$)	0.048	9.08	0.68	0.017	0.58	-0.026
Gain of Releasing Bottom 5% (G)	0.100	18.73	1.40	0.035	1.20	0.054
Policy Gain Increase from Using Long-Run VA (%)	85.2	88.8	161.5	183.3	∞	160.0

Impacts of Releasing Low Value-Added Teachers One Year After Release





Impacts of Releasing Low Value-Added Teachers Over Time



Conclusion

- We propose new VA estimator that captures teachers' contribution to long-term knowledge
- Find *entirety* of the standard VA's impact on long-run outcomes driven by long-run VA
- Using long-run VA increases policy gains of a benchmark policy two-fold
- Teachers may be twice as valuable in improving later-life outcomes than previously estimated
- Better understanding of teaching practices and policies that incentivize and lead to higher long-run VA and persistent learning gains