How to Measure a Teacher: The Influence of Test and Nontest Value-Added on Long-Run Student Outcomes

Ben Backes, James Cowan, Dan Goldhaber, and Roddy Theobald

Online Appendix

Appendix 1. Additional Tables and Figures

Table A1. Evidence on Value-Added and Long-Run Outcomes

			Impact of increase of one SD of teacher VA	
Study	Setting	Nontest	Test-based VA	Nontest VA
Chetty et al. (2014b)	Gr. 4-8, NYC, 1989-2009	--	College enroll 0.8 pp College quality $299 High Q college 0.7 pp	--
Jackson (2018) *Test-Nontest VA Correlation*:	Gr. 9, NC, 2005-2012 *0.15*	Factor[1]	HS grad: 0.1 % pts Take SAT: 1.2 % pts SAT score: 0.60 Intend 4yr: 0.1 % pts	HS grad: 1.5 % pts Take SAT: 0.1 % pts SAT score: -0.23 Intend 4yr: 1.3 % pts
Liu and Loeb (2021) *Test-Nontest VA Correlation*:	Gr. 7-11, CA district, 2004-2014 *0.12 (math), 0.08 (ELA)*	Unexc. absences	HS grad: 0.1 % pts AP courses: 0.02 AP credits: 0.11	HS grad: 0.7 % pts AP courses: 0.01 AP credits: 0.08
Mulhern and Opper (2022) *Test-Nontest VA Correlation*:	Gr. 5-7, NYC, 2005-2014 *0.04-0.06*	Attendance, grades[3]	HS grad: 0.2 % pts	HS grad: -.6 to -.8 % pts
Petek and Pope (2023) *Test-Nontest VA Correlation*:	Gr. 3-5, Los Angeles USD, 2003-2015 *0.15*	Factor[1]	Dropout: -0.2 % pts Held back: 0.1 % pts Take SAT: -0.2 % pts SAT score: 6.3 points	Dropout: -0.3 % pts Held back: -0.6 % pts Take SAT: 1.0 % pts SAT score: 2.0 points
Gilraine and Pope (2021) *Test-Nontest VA Correlation*:	Gr. 3-5, 1 large district, 2003-2017 *0.21*	Factor[1]	HS grad: 0.12 % pts Take SAT: 0.05 % pts SAT score: 2.9 points	HS grad: 0.83 % pts Take SAT: 0.33 % pts SAT score: 6.59 points
Rose et al., (2022) *Test-Nontest VA Correlation*:	Gr. 4-8, NC, 1996-2013 *0.06*	Factor[1,2]	HS grad: 0.11 pp Arrested: -0.08 pp	HS grad: 0.20 % pts Arrested: -0.36 % pts
This study *Test-Nontest VA Correlation*:	Gr. 7, 8, 10, MA, 2012-2021 *0.10*	Factor[1]	HS grad: 0.0 % pts Take SAT: 0.1 % pts SAT score: 0.01 sd AP tests passed: 0.03 College enroll: 0.2 % pts 4yr college: 0.1 % pts Selective college: 0.4 % pts College quality: $146	HS grad: 0.5 % pts Take SAT: 0.4 % pts SAT score: 0.00 sd AP tests passed: 0.00 College enroll: 1.1 % pts 4yr college: 0.8 % pts Selective college: 0.0 % pts College quality: $223

(1) Factor consists of absences, suspensions, GPA, grade progression originally developed in Jackson (2018).
(2) Rose et al. (2022) do not include GPA in factor.
(3) The measures presented in Mulhern and Opper (2022) are conditional on other test + nontest measures and are thus not directly comparable to the other studies in Table 1. The range of correlations listed in this row is for the correlation of math or ELA value-added and absence value-added over both grade levels (elementary or middle).

Table A2. Summary Statistics

	Short Run Measures			Long Run Outcomes		
	Mean	SD	N	Mean	SD	N
ELA test	-0.018	0.904	3031143	-0.028	0.895	1886138
Math test	-0.008	0.913	3035294	-0.017	0.901	1889099
Nontest index	0.036	0.962	2900555	0.050	0.943	1834642
Retained	0.006	0.080	3113737	0.008	0.087	1942322
Absences	8.683	10.266	3112733	8.625	10.326	1941571
Days suspended	0.247	1.897	3113737	0.253	1.973	1942322
GPA	2.961	0.902	2901226	2.904	0.898	1835118
Next year GPA	2.885	0.930	2827362	2.826	0.914	1830041
AP credits				4.273	8.426	1942322
AP tests taken				1.327	2.079	1942322
AP tests passed				0.887	1.783	1942322
Takes SAT				0.689	0.463	1578297
SAT scores (standard deviations)				0.067	1.001	1087702
Graduate				0.898	0.302	1942322
Dropout				0.034	0.182	1942322
Attends college				0.684	0.465	1942322
Attends 2 year college				0.181	0.385	1942322
Attends 4 year college				0.546	0.498	1942322
Median postsecondary income				39254	17096	1942322
College mobility				0.221	0.173	1942322
Lag math test	-0.010	0.918	2917783	-0.017	0.911	1814965
Lag ELA test	-0.021	0.910	2912736	-0.029	0.901	1810406
Lag retention	0.007	0.080	3038082	0.008	0.088	1900934
Lag absences	7.571	8.485	3027200	7.467	8.469	1893264
Lag days suspended	0.172	1.493	3038083	0.177	1.596	1900934
Lag GPA	3.015	0.879	2655416	2.958	0.875	1708530
Limited English proficient	0.050	0.218	3113737	0.043	0.203	1942322
Male	0.502	0.500	3113737	0.500	0.500	1942322
Free- or reduced-price lunch	0.347	0.476	3113737	0.352	0.478	1942322
Full inclusion special education	0.111	0.314	3113737	0.106	0.308	1942322
Partial inclusion special education	0.021	0.145	3113737	0.023	0.149	1942322
Substantially separate special education	0.006	0.076	3113737	0.006	0.080	1942322
Black student	0.118	0.323	3113737	0.115	0.319	1942322
Asian student	0.084	0.277	3113737	0.079	0.270	1942322
American Indian student	0.029	0.169	3113737	0.028	0.164	1942322
Pacific Islander student	0.010	0.099	3113737	0.010	0.100	1942322
Hispanic student	0.186	0.389	3113737	0.172	0.377	1942322
Takes advanced math	0.265	0.441	3113737	0.271	0.444	1942322
Takes art elective	0.280	0.449	3113737	0.219	0.414	1942322
Takes advanced language	0.105	0.306	3113737	0.143	0.350	1942322
Takes supplemental course	0.096	0.295	3113737	0.090	0.286	1942322

Table A3. Student Outcome Correlations

		Nontest Factor						
	Test	Base	No GPA	Counts	Binary	GPA	Abs	Susp
Test								
Nontest Base	0.58							
Nontest No GPA	0.33	0.84						
Nontest Counts	0.51	0.92	0.89					
Nontest Binary	0.58	0.94	0.76	0.89				
GPA	0.66	0.83	0.43	0.71	0.83			
Ln abs	-0.28	-0.74	-0.72	-0.58	-0.60	-0.39		
Ln susp	-0.22	-0.53	-0.78	-0.76	-0.54	-0.28	0.21	
Retained	-0.10	-0.29	-0.32	-0.25	-0.30	-0.19	0.09	0.09

Notes: Correlations at the student level. For the nontest factor, the "Counts" specification includes binary values of ever suspended and ever absent plus raw counts of days absent plus days suspended, the "Binary" specification uses binary values for ever suspended and chronically absent (at least 10 days absent), and the "No GPA" specification uses a nontest factor that does not include GPA.

Table A4. Cross-Sectional Relationship Between VA Measures and Teacher Experience

	Test				Nontest			
	ELA VA	ELA Actual	Math VA	Math Actual	ELA VA	ELA Actual	Math VA	Math Actual
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2-3 years experience	0.03**	0.02***	0.04***	0.03***	-0.01	-0.00	-0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
4-5 years experience	0.06***	0.02***	0.07***	0.03***	0.00	0.00	-0.00	0.01
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
6-10 years experience	0.07***	0.03***	0.08***	0.03***	0.00	0.01	-0.01	0.00
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
11+ years experience	0.04***	0.02***	0.07***	0.03***	-0.01**	-0.00	-0.03***	-0.01
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Notes: Results from regression of either VA forecast ("VA") or actual student outcome "Actual" by subject and type (test or nontest) on teacher experience and the student-level controls used in Column (6) of Table 3. Omitted group are teachers in their first or second year of teaching.

Table A5. Quasi-Experimental Estimates of Forecast Bias

	Δ Score (1)	Δ Score (2)	Δ Score (3)	Δ Other Subj. score (4)
Panel A. Test scores: grades 7-8 and 10				
Changes in Mean Teacher VA	1.122	0.954	0.911	0.069
Across Cohorts	(0.075)	(0.110)	(0.150)	(0.064)
Sch x Grade x Subj x Year Cells	14290	14290	9137	12870
Panel B. Nontest factor: grades 7-8 and 10				
Changes in Mean Teacher VA	1.193	*See note*	*See note*	0.451
Across Cohorts	(0.178)			(0.146)
Sch x Grade x Subj x Year Cells	13744			12467
Year Fixed Effects	Y			Y
School-year Fixed Effects[1]		Y	Y	
Lagged Score Controls			Y	
Lead and Lag Changes in Teacher VA			Y	
Other-Subject Change in Mean Teacher VA				Y

Notes: replication of Chetty et al. (2014a) Table 4.

Panel B note: including school-year fixed effects is conceptually challenging in the case of nontest value-added due to mean school-grade-year nontest outcomes being identical across subjects for students enrolled in both subjects. This means that with two subjects, we cannot separate the following three impacts within the same school and year: math teacher nontest value-added, ELA teacher nontest value-added, and the school-year fixed effect. True increases in a given subject's value-added holding the other's constant are thus attenuated by the school fixed effect absorbing some of the increase, rendering the coefficient uninformative as a test.

Table A6. Correlations between Various Teacher Quality Measures

		Nontest VA			
	Test VA	Base	No GPA	Counts	Binary
Test VA					
Nontest VA Base	0.10				
Nontest VA No GPA	0.03	0.36			
Nontest VA Counts	0.08	0.84	0.57		
Nontest VA Binary	0.10	0.95	0.30	0.82	
GPA VA	0.08	0.74	0.06	0.55	0.79

Notes: Correlation between different value-added measures for a given teacher in a given year. For the nontest factor, the "Counts" specification includes binary values of ever suspended and ever absent plus raw counts of days absent plus days suspended, the "Binary" specification uses binary values for ever suspended and chronically absent (at least 10 days absent), and the "No GPA" specification uses a nontest factor that does not include GPA.

Table A7. Results by Grade.

	All (7, 8, 10) (1)	7 and 8 (2)	Grade 7 (3)	Grade 8 (4)	Grade 8 (5)
Panel A. AP Tests Passed					
Test VA	0.03***	0.01	0.01	0.01	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Nontest VA	-0.02	-0.02	-0.01	-0.03	0.00
	(0.01)	(0.02)	(0.02)	(0.02)	(0.05)
Panel B. SAT Scores					
Test VA	0.01***	0.01***	0.01	0.01**	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Nontest VA	-0.00	-0.00	0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Panel C. Graduate High School					
Test VA	0.02	0.03	0.01	-0.03	0.01
	(0.08)	(0.10)	(0.16)	(0.14)	(0.10)
Nontest VA	0.54**	0.58**	0.53*	0.34	0.11
	(0.24)	(0.25)	(0.32)	(0.35)	(0.68)
Panel D. College Quality ($)					
Test VA	145.65***	98.04*	27.13	107.38	212.43***
	(39.68)	(52.04)	(81.43)	(66.99)	(60.52)
Nontest VA	222.94**	205.35*	24.95	196.14	466.17
	(108.37)	(114.23)	(154.46)	(152.55)	(352.70)
Panel F. Enroll in College					
Test VA	0.15	0.16	0.19	0.00	0.16
	(0.12)	(0.16)	(0.22)	(0.22)	(0.18)
Nontest VA	1.10***	0.98***	1.22***	0.29	2.51**
	(0.34)	(0.36)	(0.46)	(0.50)	(1.23)
Panel F. Enroll in Four-Year College					
Test VA	0.11	0.18	0.19	0.02	-0.00
	(0.12)	(0.16)	(0.23)	(0.20)	(0.20)
Nontest VA	0.83**	0.79**	0.72	0.32	1.03
	(0.38)	(0.40)	(0.51)	(0.51)	(1.09)
Panel G. Enroll in Selective College					
Test VA	0.40***	0.28**	0.01	0.49***	0.57***
	(0.11)	(0.11)	(0.17)	(0.15)	(0.19)
Nontest VA	0.06	0.02	0.15	-0.06	0.40
	(0.19)	(0.20)	(0.34)	(0.25)	(0.69)

Notes: Coefficients on nontest VA from regressions with student outcomes as the dependent variable with nontest value-added constructed for different samples by student grade. Each coefficient represents the results of a separate regression; Test VA and Nontest VA coefficients obtained in regressions not including the other measure. All regressions include cubic polynomials in prior achievement and the nontest index, student race, gender, and free/reduced-price lunch status, special education status, English learner status, indicators for taking an advanced math course, art elective, foreign language class, supplemental class, and English language learner class, and the means of these variables at the classroom level, along with school-by-track fixed effects. All variables are interacted

with subject and grade level (middle or high school). The sample includes all students in the matched long-run sample described in the text. Standard errors clustered by school in parentheses. * p < 0.10 ** p < 0.05 *** p < 0.01

Table A8. Sensitivity of Results to Alternate Specifications of Nontest Factor.

	Base (1)	Counts (2)	Binary (3)	No GPA (4)	GPA Only (5)
Panel A. AP Tests Passed					
Nontest VA	-0.02	0.01	-0.02	-0.02	-0.01
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Panel B. SAT Scores					
Nontest VA	-0.00	0.00	-0.01	0.01	-0.01*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Panel C. Graduate High School					
Nontest VA	0.54**	0.70**	0.44*	0.58**	0.07
	(0.24)	(0.34)	(0.25)	(0.28)	(0.16)
Panel D. College Quality ($)					
Nontest VA	222.94**	289.82**	196.97*	183.66	75.67
	(108.37)	(141.24)	(113.36)	(131.70)	(82.00)
Panel F. Enroll in College					
Nontest VA	1.10***	1.56***	1.07***	1.39***	0.31
	(0.34)	(0.46)	(0.35)	(0.46)	(0.26)
Panel F. Enroll in Four-Year College					
Nontest VA	0.83**	0.90*	0.81**	0.78*	0.48
	(0.38)	(0.49)	(0.40)	(0.44)	(0.30)
Panel G. Enroll in Selective College					
Nontest VA	0.06	0.13	-0.01	-0.28	0.05
	(0.19)	(0.26)	(0.20)	(0.26)	(0.16)

Notes: Coefficients on nontest VA from regressions with student outcomes as the dependent variable with nontest value-added constructed using alternate nontest composites. The "Counts" specification includes binary values of ever suspended and ever absent plus raw counts of days absent plus days suspended, the "Binary" specification uses binary values for ever suspended and chronically absent (at least 10 days absent), and the "No GPA" specification uses a nontest factor that does not include GPA. All regressions include cubic polynomials in prior achievement and the nontest index, student race, gender, and free/reduced-price lunch status, special education status, English learner status, indicators for taking an advanced math course, art elective, foreign language class, supplemental class, and English language learner class, and the means of these variables at the classroom level, along with school-by-track fixed effects. All variables are interacted with subject and grade level (middle or high school). The sample includes all students in the matched long-run sample described in the text. Standard errors clustered by school in parentheses. * p < 0.10 ** p < 0.05 *** p < 0.01

Table A9. Robustness to Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. AP Tests Passed								
Test VA	0.03***	0.03***	0.03***	0.03				
	(0.01)	(0.01)	(0.01)	(0.02)				
Nontest VA					-0.02	-0.01	-0.01*	-0.04*
					(0.01)	(0.01)	(0.01)	(0.02)
Panel B. SAT Scores (standard deviations)								
Test VA	0.01***	0.01***	0.01***	0.02				
	(0.00)	(0.00)	(0.00)	(0.01)				
Nontest VA					-0.00	0.00	-0.00	0.01
					(0.01)	(0.00)	(0.00)	(0.02)
Panel C. Graduate High School								
Test VA	0.02	0.04	0.02	0.26				
	(0.08)	(0.07)	(0.07)	(0.23)				
Nontest VA					0.54**	0.37**	0.24**	0.06
					(0.24)	(0.17)	(0.12)	(0.42)
Panel D. College Quality ($)								
Test VA	146***	164***	143***	286*				
	(40)	(40)	(35)	(152)				
Nontest VA					223**	195**	122**	173
					(108)	(84)	(49)	(242)
Panel E. Enroll in College								
Test VA	0.15	0.24**	0.10	0.77*				
	(0.12)	(0.12)	(0.10)	(0.42)				
Nontest VA					1.10***	0.96***	0.43***	0.47
					(0.34)	(0.26)	(0.16)	(0.71)
Panel F. Enroll in Four-Year College								
Test VA	0.11	0.15	0.09	1.08**				
	(0.12)	(0.12)	(0.11)	(0.45)				
Nontest VA					0.83**	0.82***	0.45**	0.11
					(0.38)	(0.29)	(0.18)	(0.65)
Panel G. Enroll in Selective College								
Test VA	0.40***	0.42***	0.42***	0.31				
	(0.11)	(0.10)	(0.10)	(0.26)				
Nontest VA					0.06	-0.00	0.00	0.50
					(0.19)	(0.15)	(0.11)	(0.42)
N	1581369	1379647	1581369	8489	1569318	1374526	1569318	8489
Baseline	Y				Y			
Other-subj VA		Y				Y		
Resid on Tch			Y				Y	
Switching Design				Y				Y

Notes: Columns 1 and 5 re-produce the baseline models in Tables 5 and 6. Columns 2 and 6 add controls for the test and nontest value-added of student's teacher in other subject. Columns 3 and 7 use teacher fixed effects rather than school fixed effects in the first stage residualization process to estimate the relationship between outcome (test or

nontest) and controls. Columns 4 and 8 use grade-school-year aggregates of value-added and outcomes in the quasi-experimental test developed in Chetty et al. (2014b).

Appendix 2. Construction of Tracks

We follow Jackson (2014, 2018) and construct academic tracks using the 10 most enrolled classes in each grade level. In each case, students in each track have the same enrollment status in each of the 10 classes and the academic level of the math and ELA classes (basic, general, advanced, or postsecondary). In **Table B.1**, we show the distribution of the number of teachers in each track for both the full sample of matched students in 7th, 8th, and 10th grades as well as the restricted sample with long-run outcomes. Tracks do tend to be smaller than the school-grade cells as a whole, although most tracks do have multiple teachers. The modal number of teachers within school-grades is 5 (4 for the long-run sample); it is 4 (3 for the long-run sample) within tracks.

We show summary statistics for each of these courses in **Tables B.2** through **B.4**. The italicized courses are the 10 most popular in each grade. As shown in the tables, popular courses better differentiate students' academic ability in high school than in middle school. This is primarily because there are fewer courses in each subject in the middle school course categorization. To better use information on tracking embedded in class assignments, we construct five additional covariates used in both the value-added models and the regression analyses. The courses used to construct these indicators (among the courses with at least 5% of students enrolled) are indicated in bold in Tables B.2 through B.4. We construct an indicator for advanced math courses if students take Pre-Algebra or Algebra in 7th grade; Algebra in 8th grade; or Algebra II in 10th grade. We define advanced foreign languages if students take any foreign language in 7th or 8th grade; or if students take a third year foreign language class in 10th grade. We construct an arts elective for students in 7th or 8th grade who take an art course other than the grade-specific Art or Music course. As we show in Tables B.2 and B.3, these are mostly band, chorus, and drama courses. We additionally construct an indicator for supplemental courses for students who take either a Tutorial class or a Supplemental course. The Supplemental courses are usually offered in math. Finally, we construct an indicator for students who take an English as a Second Language (ESL) class. Not all students classified as English language learners take an ESL class, so this indicator is distinct from the limited English proficient indicator. As can be seen in Tables B.2 and B.3, some of the arts, foreign language, supplemental, and ESL classes – although not among the top 10 most enrolled – are strongly predictive of student outcomes.

Table B.1. Distribution of the Number of Teachers per Track

Number of Teachers	VA: School-Grade	VA: School-Track	LR: School-Grade	LR: School-Track
1	41582	132502	47606	124155
2	99646	260533	127542	228731
3	179756	342344	179671	288014
4	215080	363909	191448	244810
5	275270	334234	178299	219297
6	248733	298912	160457	175350
7	246530	237419	117284	140951
8	176665	200050	100784	111038
9	183389	194989	81585	90815
10	175829	154790	67230	78657
11	128890	100409	50696	55503
12	129854	93835	57657	42433
13	97499	76039	51485	38259
14	88755	66839	52181	41517
15+	826259	256933	563694	148089

Notes: Counts of teachers per school-grade or school-track cells for the value-added (2012-2019, [VA]) and long-run [LR] samples.

Table B.2. Summary Statistics by Course Enrollment (Grade 7)

Course	N	LEP	Prior ELA Score	Prior Math Score	Prior Retention	Prior Absences	Prior Days Suspended	Prior GPA	Special Education
French	50529	0.01	0.43	0.40	0.00	6.26	0.04	3.57	0.07
General Band	56214	0.03	0.30	0.35	0.00	5.69	0.05	3.49	0.09
Spanish	159304	0.01	0.27	0.26	0.00	6.48	0.06	3.47	0.08
Drama (grade 7)	48237	0.04	0.26	0.26	0.00	6.45	0.06	3.46	0.15
Foreign Language (grade 7)	91557	0.02	0.25	0.25	0.00	6.49	0.06	3.36	0.09
Family and Consumer Science—Comprehensive	33008	0.02	0.26	0.24	0.00	6.27	0.04	3.48	0.14
Pre-Algebra	145368	0.06	0.19	0.23	0.00	6.90	0.09	3.29	0.13
World Geography	60920	0.01	0.21	0.20	0.00	6.33	0.04	3.47	0.16
Pre-Engineering Technology	49371	0.03	0.17	0.17	0.00	6.80	0.06	3.36	0.15
Chorus	85350	0.03	0.22	0.13	0.00	6.75	0.05	3.43	0.13
Engineering and Technology—Other	35484	0.02	0.10	0.12	0.00	6.85	0.07	3.42	0.15
Engineering Technology	60196	0.02	0.11	0.11	0.00	6.94	0.08	3.31	0.17
Computer Applications	42167	0.02	0.02	0.01	0.00	7.04	0.07	3.30	0.16
Health Education	351108	0.06	0.01	0.01	0.00	7.18	0.12	3.26	0.16
Introduction to Computers	52026	0.03	0.04	0.01	0.00	7.04	0.08	3.23	0.16
Health and Fitness	39924	0.04	-0.02	0.00	0.00	7.78	0.13	3.34	0.15
Art (grade 7)	492507	0.06	0.00	0.00	0.00	7.28	0.12	3.26	0.16
Music (grade 7)	296091	0.06	-0.03	-0.02	0.00	7.26	0.12	3.24	0.15
Physical Education (grade 7)	652874	0.06	-0.03	-0.03	0.00	7.32	0.14	3.21	0.17
Writing (grade 7)	38072	0.03	-0.06	-0.03	0.00	7.18	0.14	3.22	0.16
Computer and Information Technology	77996	0.05	-0.02	-0.04	0.00	7.48	0.14	3.28	0.15
Language Arts (grade 7)	624103	0.05	-0.05	-0.05	0.00	7.53	0.15	3.17	0.17
Social Studies (grade 7)	501381	0.06	-0.07	-0.06	0.01	7.63	0.16	3.13	0.17
Science (grade 7)	591268	0.06	-0.07	-0.07	0.00	7.57	0.15	3.16	0.17
Technological Literacy	63938	0.06	-0.08	-0.09	0.00	7.44	0.15	3.19	0.16
Computer Literacy	67280	0.08	-0.11	-0.09	0.00	7.41	0.14	3.10	0.16
Study Skills	86406	0.04	-0.12	-0.13	0.00	7.63	0.12	3.08	0.30
World History—Overview	47541	0.12	-0.12	-0.14	0.01	7.18	0.17	3.26	0.18
Mathematics (grade 7)	507623	0.07	-0.16	-0.17	0.00	7.74	0.17	3.12	0.18
Exploratory	34426	0.07	-0.15	-0.19	0.00	7.92	0.22	3.06	0.16
Reading (grade 7)	72460	0.06	-0.29	-0.29	0.00	7.67	0.14	3.08	0.25

Grade 7	35996	0.06	-0.40	-0.42	0.00	8.16	0.21	3.02	0.47
Tutorial	73386	0.10	-0.49	-0.49	0.01	8.30	0.31	2.92	0.29
Mathematics—Supplemental	33873	0.11	-0.50	-0.54	0.01	9.13	0.33	2.78	0.22
English as a Second Language	33139	0.94	-1.40	-1.20	0.01	7.96	0.25	2.50	0.14

Notes: Summary statistics for students enrolled in courses in grade 7 (2012-2019) with enrollments of at least 5% of the total enrollment. Courses in bold are included in the course type indicators used as covariates in the regression analyses. Courses indicated in italics are used to construct academic tracks.

Table B.3. Summary Statistics by Course Enrollment (Grade 8)

Course	N	LEP	Prior ELA Score	Prior Math Score	Prior Retention	Prior Absences	Prior Days Suspended	Prior GPA	Special Education
French	45887	0.01	0.48	0.46	0.00	6.53	0.05	3.48	0.05
General Band	48437	0.03	0.32	0.40	0.00	5.79	0.06	3.44	0.08
Algebra I	193580	0.05	0.31	0.39	0.00	6.98	0.10	3.36	0.10
U.S. History—Comprehensive	33902	0.02	0.30	0.31	0.00	7.03	0.07	3.37	0.15
Foreign Language (grade 8)	87693	0.01	0.26	0.29	0.00	6.78	0.08	3.29	0.08
Spanish	178715	0.01	0.27	0.27	0.00	6.75	0.09	3.38	0.07
Drama (grade 8)	37190	0.04	0.21	0.23	0.00	7.22	0.12	3.37	0.16
Family and Consumer Science—Comprehensive	34621	0.02	0.16	0.18	0.00	6.93	0.08	3.34	0.13
Pre-Engineering Technology	52838	0.03	0.10	0.14	0.00	7.35	0.12	3.25	0.15
Chorus	77240	0.03	0.23	0.14	0.00	7.32	0.09	3.35	0.13
Introduction to Computers	40820	0.03	0.06	0.13	0.00	7.02	0.11	3.20	0.14
Engineering and Technology—Other	36815	0.02	0.09	0.12	0.00	7.45	0.12	3.30	0.15
Engineering Technology	60303	0.02	0.10	0.12	0.00	7.55	0.13	3.20	0.16
Writing (grade 8)	38593	0.03	0.04	0.07	0.00	7.77	0.21	3.20	0.15
Health Education	336912	0.05	0.04	0.05	0.00	7.58	0.16	3.18	0.16
Health and Fitness	37198	0.04	0.01	0.05	0.00	8.18	0.16	3.18	0.14
Computer and Information Technology	71086	0.05	-0.03	0.01	0.00	7.74	0.17	3.19	0.15
Art (grade 8)	471194	0.05	-0.02	0.00	0.00	7.87	0.18	3.15	0.16
Physical Education (grade 8)	648621	0.06	-0.03	-0.01	0.00	7.80	0.20	3.12	0.16
World History—Overview	84768	0.06	-0.03	-0.02	0.00	7.41	0.19	3.21	0.16
Language Arts (grade 8)	616852	0.05	-0.05	-0.05	0.00	8.11	0.21	3.08	0.16
Social Studies (grade 8)	476767	0.05	-0.06	-0.05	0.00	8.18	0.21	3.06	0.16
Music (grade 8)	257261	0.06	-0.07	-0.05	0.00	7.92	0.18	3.12	0.15

Science (grade 8)	591897	0.06	-0.08	-0.07	0.00	8.18	0.22	3.07	0.17
Technological Literacy	68029	0.07	-0.07	-0.08	0.00	7.87	0.21	3.12	0.16
Computer Literacy	65203	0.08	-0.14	-0.13	0.00	8.06	0.20	3.00	0.16
Exploratory	35381	0.06	-0.11	-0.17	0.00	8.66	0.28	3.03	0.15
Study Skills	78544	0.05	-0.26	-0.25	0.00	8.55	0.24	2.98	0.33
Mathematics (grade 8)	364597	0.08	-0.27	-0.29	0.00	8.70	0.29	2.94	0.20
Pre-Algebra	75596	0.04	-0.25	-0.32	0.00	8.71	0.18	2.94	0.20
Tutorial	68992	0.10	-0.48	-0.48	0.00	8.92	0.41	2.84	0.29
Mathematics—Supplemental	37397	0.09	-0.43	-0.48	0.01	9.56	0.39	2.78	0.21
Reading (grade 8)	46992	0.07	-0.49	-0.49	0.01	8.86	0.29	2.81	0.32
English as a Second Language	32411	0.93	-1.43	-1.21	0.01	8.64	0.33	2.46	0.13

Notes: Summary statistics for students enrolled in courses in grade 8 (2012-2019) with enrollments of at least 5% of the total enrollment. Courses in bold are included in the course type indicators used as covariates in the regression analyses. Courses indicated in italics are used to construct academic tracks.

Table B.4. Summary Statistics by Course Enrollment (Grade 10)

Course	N	LEP	Prior ELA Score	Prior Math Score	Prior Retention	Prior Absences	Prior Days Suspended	Prior GPA	Special Education
French III	40224	0.00	0.66	0.64	0.00	5.20	0.02	3.32	0.02
Algebra II	170280	0.01	0.49	0.58	0.01	5.96	0.10	3.27	0.04
Spanish III	140810	0.00	0.46	0.47	0.00	5.52	0.05	3.22	0.03
Chemistry	262342	0.02	0.22	0.25	0.01	6.76	0.13	3.01	0.08
Modern World History	54697	0.03	0.21	0.24	0.01	7.25	0.21	2.91	0.16
Health and Fitness	73106	0.03	0.09	0.10	0.01	7.51	0.19	2.87	0.16
Integrated Math—multi-year equivalent	42822	0.07	0.03	0.05	0.02	8.30	0.21	2.80	0.20
Physical Education/Health/Drivers' Education	75195	0.03	0.01	0.01	0.01	6.93	0.18	2.89	0.16
Health Education	139773	0.04	-0.02	0.00	0.01	8.10	0.28	2.82	0.14
English/Language Arts II (10th grade)	567927	0.03	-0.05	-0.04	0.01	7.96	0.26	2.77	0.15
Spanish II	129778	0.02	-0.09	-0.07	0.01	7.74	0.21	2.74	0.09
Visual Arts—Comprehensive	43093	0.06	-0.04	-0.09	0.02	8.93	0.34	2.74	0.15
Early U.S. History	216734	0.04	-0.09	-0.09	0.01	8.12	0.29	2.71	0.15
Physical Education	314700	0.06	-0.13	-0.10	0.02	8.52	0.34	2.69	0.15
Modern U.S. History	116438	0.04	-0.15	-0.14	0.02	8.15	0.31	2.68	0.15

U.S. History—Comprehensive	141176	0.08	-0.19	-0.17	0.02	9.09	0.30	2.67	0.15
Biology	259602	0.07	-0.25	-0.25	0.02	9.15	0.37	2.57	0.18
Geometry	393656	0.06	-0.26	-0.27	0.01	8.70	0.32	2.60	0.15
Spanish I	50259	0.04	-0.50	-0.51	0.02	11.17	0.57	2.24	0.25
Tutorial	36870	0.07	-0.66	-0.71	0.03	12.02	0.68	2.23	0.61
Study Skills	59947	0.05	-0.69	-0.72	0.03	12.24	0.54	2.24	0.64
Algebra I	32889	0.24	-0.70	-0.81	0.06	15.25	0.93	1.90	0.25
English as a Second Language	49243	0.96	-1.64	-1.35	0.07	9.73	0.35	2.23	0.07

Notes: Summary statistics for students enrolled in courses in grade 10 (2012-2019) with enrollments of at least 5% of the total enrollment. Courses in bold are included in the course type indicators used as covariates in the regression analyses. Courses indicated in italics are used to construct academic tracks.

Appendix 3. Construction of College Quality Measure

We use a college quality measure that takes the median earnings measure from the Chetty et al. (2017) Mobility Report Card (MRC) for students who enroll in a college during their expected grade 13 school year. Although this does not perfectly match the sample construction in MRC, it is quite close. They define students as attending a college using the school with the longest period of attendance during their age 19 - 22 school years; they use the first school attended in case of ties. Because they rely on enrollment (and not college completion), their measures should be appropriate for assigning to students we observe to enroll (but not necessarily graduate) college. MRC match student records to personal income tax data in 2014 for the 1980-1982 birth cohorts, when students were aged 32-34. In the publicly available data, they estimate the median individual earnings for each college and use a masking procedure to blur the results by aggregating data across similar colleges; nonetheless, they find that these estimates are quite accurate (mean absolute estimation error: $266).

We combine the MRC data with two additional categories using the American Community Survey (ACS): (1) students who did not complete high school, and (2) students who completed high school but did not attend college. The latter group is included as a category in the MRC, although it includes data from all states (and not just Massachusetts). Because cell sizes in individual years for these groups can be small, we pool data from the 2015 5-year sample of the ACS. We limit the sample to individuals born in Massachusetts, who are aged 33-35 in the current year (so that reported earnings are from the age 32-34 years). For group (1), we estimate median earnings for people whose highest level of education is 12^{th} grade or less (without a high school diploma) or a GED. We include GED recipients in the non-high school sample given that they tend to earn similarly to high school dropouts and we do not observe GED status in the administrative data (Cameron & Heckman, 2003). For group (2), we estimate median earnings for those with a regular high school diploma. These measures are likely to understate earnings for our samples of high school non-completers and non-college attenders given that we measure these outcomes as of the age 19 school year (and not at ages 32-34); nonetheless, the length of our panel precludes following students' educational attainment over longer time horizons. To that extent, we may slightly overstate the effect of nontest VA on the earnings measure since it is most predictive of effects on the college/non-college margin. However, we note that among cohorts we observe through grade 16, only 8% of students not enrolled in college in grade 13 were enrolled in college in grade 16.

We display summary statistics by college tier attended in Table C1. Unsurprisingly, students who attended a highly selective college tend to have taken the SAT, scored quite highly on the SAT, and had high standardized test scores as course grades.

Table C1. Summary Statistics by College Tier.

		SAT					
	Median $	Score	Take	Test	GPA	N	Percent
Highly Selective	$63,550	1.14	93%	0.89	3.65	54407	12%
Selective Four-year	$48,379	0.08	95%	0.22	3.21	177494	38%
Non-selective Four-year	$45,751	-0.24	88%	-0.04	2.94	15680	3%
Two-year or for profit	$30,926	-0.78	62%	-0.46	2.52	67498	14%
Grad HS, No College	$26,392	-0.47	38%	-0.49	2.42	101252	21%
Did not Graduate HS	$12,480	-0.83	10%	-0.74	1.95	56647	12%

Note: Summary statistics by level of college (as defined in Chetty et al., 2017). Median earning column represents median earnings of graduates from college attended for the first four rows of college attenders, while median earnings for the no college groups are obtained from ACS as described in text.

When we consider these tiers as separate outcomes in our models (Table C2 below), we see that the relationship between nontest VA and college quality is mostly driven by impacts both on "does not graduate HS" and "HS graduate but no college", while the relationship between test VA and college quality is driven by a shift on the margin between "selective four-year" and "highly selective four-year". In other words, students assigned to a higher VA teacher are more likely to attend a highly selective four-year university and less likely to attend a selective four-year university. Given the substantial economic returns to attending a highly-selective university (Chetty et al., 2023), this is sufficient to drive the "college quality" impacts documented in the main text.

Table C2. Impacts of Teacher Quality on Mutually Exclusive Long-Run Educational Attainment Measures

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Does not Graduate HS						
Test VA	-0.02	-0.02			-0.00	-0.00
	(0.07)	(0.07)			(0.07)	(0.08)
Nontest VA			-0.53**	-0.54**	-0.53**	-0.55**
			(0.22)	(0.23)	(0.22)	(0.24)
Panel B. HS Graduate but no College						
Test VA	-0.19	-0.12			-0.18	-0.12
	(0.13)	(0.12)			(0.13)	(0.12)
Nontest VA			-0.68**	-0.58*	-0.65**	-0.56*
			(0.33)	(0.33)	(0.33)	(0.33)
Panel C. Two-year or For Profit College						
Test VA	0.16	0.04			0.15	0.03
	(0.10)	(0.10)			(0.10)	(0.10)
Nontest VA			0.32	0.32	0.30	0.32
			(0.30)	(0.29)	(0.30)	(0.29)
Panel D. Non-Selective Four-Year						
Test VA	-0.04	-0.04			-0.04	-0.04
	(0.05)	(0.05)			(0.05)	(0.05)
Nontest VA			-0.18	-0.15	-0.17	-0.14
			(0.12)	(0.13)	(0.12)	(0.13)
Panel E. Selective Four-Year						
Test VA	-0.49**	-0.26*			-0.51**	-0.27*
	(0.23)	(0.16)			(0.23)	(0.16)
Nontest VA			1.01**	0.90**	1.09**	0.94**
			(0.47)	(0.41)	(0.48)	(0.41)
Panel F. Highly Selective Four-Year						
Test VA	0.59***	0.40***			0.58***	0.40***
	(0.14)	(0.10)			(0.14)	(0.11)
Nontest VA			0.04	0.05	-0.04	-0.01
			(0.22)	(0.19)	(0.23)	(0.19)
School-Grade-Year FE	Y		Y		Y	
School-Track-Year FE		Y		Y		Y

Notes: Coefficients from regressions with mutually exclusive and exhaustive binary student enrollment/attainment outcome as the dependent variable. All regressions include cubic polynomials in prior achievement and the nontest index, student race, gender, and free/reduced-price lunch status, special education status, English learner status, indicators for taking an advanced math course, art elective, foreign language class, supplemental class, and English language learner class, and the means of these variables at the classroom level. All variables are interacted with subject and grade level (middle or high school). The sample includes all students in the matched long-run sample described in the text. Standard errors clustered by school in parentheses. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

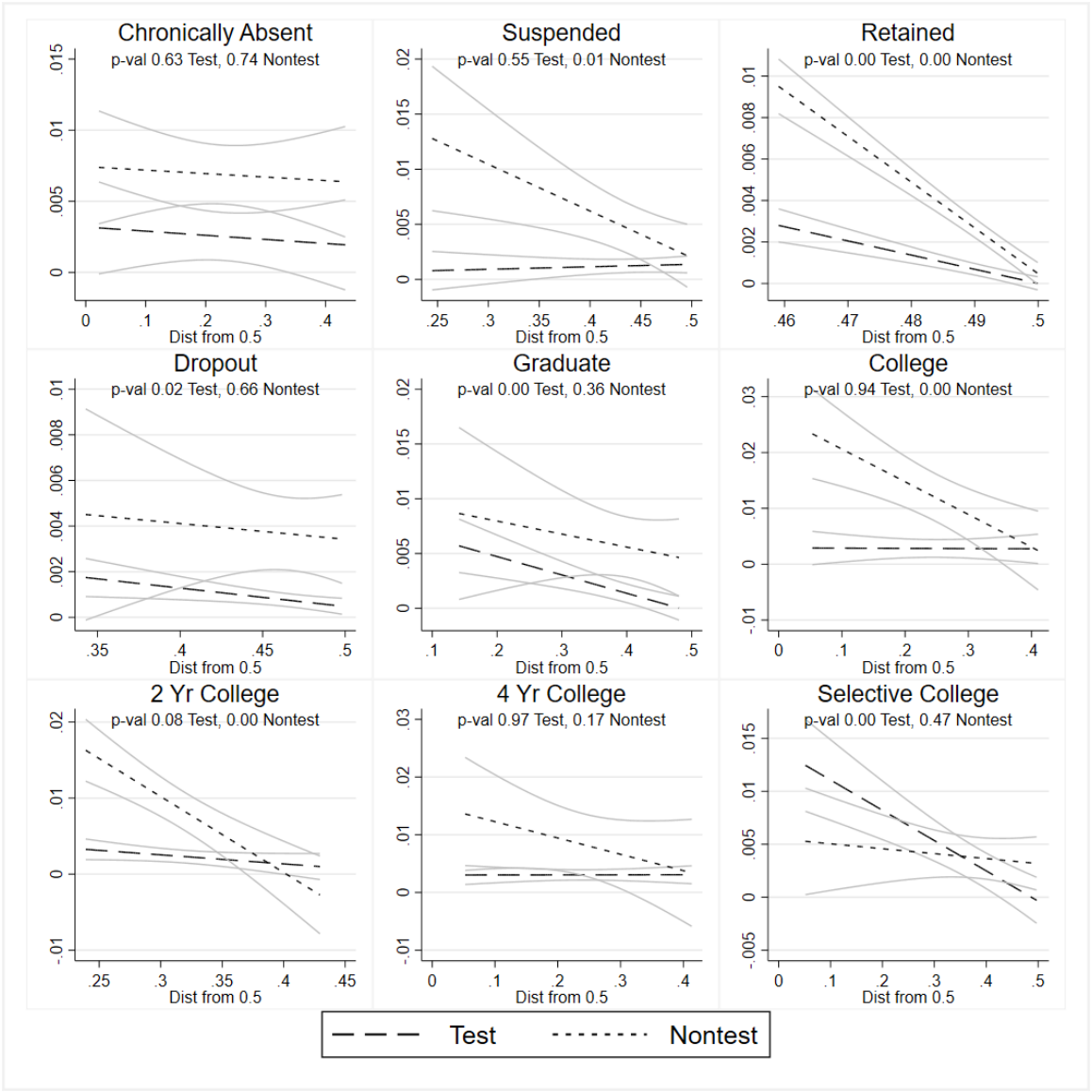
Appendix 4. Tests for Mechanical Heterogeneity

The results in Figure 3 suggest that the impact of nontest value-added on short-run and some long-run outcomes is larger for students at the bottom of the distribution. As pointed out by Jackson et al. (2022), this pattern could be driven by either (a) teachers truly having larger impacts on the soft skills of these students or (b) the students who are marginal for some of the underlying components of the nontest factor used to estimate nontest value-added (e.g., absences and suspensions) being the same as those who are marginal for binary long-run outcomes. Either of these explanations would tend to amplify the average effect of nontest value-added. However, in the latter case, the patterns observed in Figure 3 may not be reflective of true differences in teacher effects on unmeasured student skills across the skill distribution.

Jackson et al. (2022) propose a test measuring the relationship between the strength of the value-added effect in a given decile and how close students in that decile are to being marginal for an outcome (i.e., the average distance between the decile-specific mean and 0.5). We perform this exercise below and find evidence of mechanical heterogeneity for suspensions and grade retention for nontest value-added; i.e., the largest impacts on these outcomes through nontest value-added are for students closest to a 0.5 base rate. In addition, we find similar evidence for college-going and four-year college-going for nontest value-added and for selective college-going for test value-added. Thus, many of the relationships observed in Figure 3 are likely mechanical. Put another way, the differential between the test-based and nontest value-added findings for student outcomes is not necessarily evidence of heterogeneity of teacher effects on the underlying skills of their students. For example, it is not necessarily the case that teacher value-added predicting selective college-going for students at the top of the distribution means that teacher skills captured by test value-added are most impactful for high-achieving students. This is what one might conclude based on a study where selective college-going were the only long-run outcome available. However, it just happens to be the case that this outcome satisfies the conditions of having a long-run outcome sensitive to changes in the short-run outcome and the short-run outcome being sensitive to changes in teacher quality.

In the lower portion of the distribution, a similar argument holds for the impact of nontest value-added on college attendance. This illustrates a benefit using a continuous long-run outcome measure (college quality): it captures each of these marginal changes at different points in the student distribution to provide a more complete measure of teacher effects. While the estimated value-added effects on this measure are consistent with teachers who raise the nontest skills of their students having the largest impacts for students at the bottom of the distribution, estimates are very noisy.

Appendix Figure D1. Mechanical Heterogeneity.



Notes: each panel contains the results from 10 separate regressions, one from each decile of incoming educational advantage (the average of prior test scores and the nontest factor), which includes out-of-sample test value-added and nontest value-added as regressors. The y-axis represents estimated impact of a given value-added type on that decile, and the x-axis the distance between the mean of the outcome in that decile and 0.5. Regressions are from the specification in Equation (2), which includes student demographic information, prior test and nontest outcomes, class-level averages of each, and school-track-year fixed effects. The p-values shown in each plot are tests for whether the fitted slope is equal to zero.