

## Appendix A

### MET Short Grit Scale

#### Elementary Items:

1. I often set a goal but later choose to pursue a different one.\* (CoI)
2. Sometimes, when I'm working on a project, I get distracted by a new and different topic.\* (CoI)
3. I have been obsessed with a certain idea or project for a short time but later I lose that interest.\* (CoI)
4. It's hard for me to finish projects that take a long time to complete.\* (CoI)
5. I finish whatever I begin. (PoE)
6. If something is hard to do and I begin to fail at it, I keep trying anyway. (PoE)
7. I am a hard worker. (PoE)
8. I try to do a good job on everything I do. (PoE)

CoI = Items that comprise the Consistency of Interest subscale

PoE = Items that comprise the Perseverance of Effort subscale

\* Items are reverse coded

#### Response scale:

Not like me at all (1)

Not much like me (2)

Some-what like me (3)

Mostly like me (4)

Very much like me (5)

### MET Growth Mindset Scale

#### Elementary & Secondary Items:

1. Your intelligence is something you can't change very much.\*
2. You have a certain amount of intelligence, and you can't really do much to change that.\*
3. You can learn new things, but you can't really change your basic intelligence.\*

\* Items are reverse coded

#### Response Scale:

Disagree A Lot (1)

Disagree (2)

Disagree A Little (3)

Agree a Little (4)  
Agree (5)  
Agree a Lot (6)

MET TRIPOD items used to measure Effort in Class

Elementary & Secondary Items:

1. I have done my best quality work in this class.
2. I have pushed myself hard to understand my lessons in this class.
3. When doing schoolwork in this class, I try to learn as much as I can and I don't worry how long it takes.
4. In this class I stop trying when the work gets hard.
5. In this class I take it easy and do not try very hard to do my best.
6. When homework is assigned for this class, how much do you usually complete?

Response scale for items 1-5:

Totally Untrue (1)  
Mostly Untrue (2)  
Somewhat (3)  
Mostly True (4)  
Totally True. (5)

Response scale for item 6:

Never Assigned (1)  
None of it (2)  
Some of it (3)  
Most of it (4)  
All (5)  
All plus some extra (6)

## Appendix B

### Measures used in the Educational Longitudinal Study analyses

#### Social-Emotional Measures:

All questions were asked using a 1-4 Likert Scale, with “Strongly Disagree”, “Disagree”, “Agree” and “Strongly Agree” assigned values 1 through 4, respectively. For both variables, indices were created by averaging the responses to all sub-questions identified as pertaining to growth mindset and perseverance from the survey. These questions were as follows:

Growth Mindset (in math) (Taken from ELS 2002 Student Questionnaire, Question 88):

- a) Most people can learn to be good at math
- b) You have to be born with the ability to be good at math (reverse coded)

Grit: Perseverance of Effort (Taken from ELS 2002 Student Questionnaire, Question 89):

- a) When studying, I try to work as hard as possible
- b) When studying, I keep working even if the material is difficult
- c) When studying, I try to do my best to acquire the knowledge and skills taught
- d) When studying, I put forth my best effort

#### Achievement Measures:

Input variables, including a composite of math and reading test scores and constructed scores for growth mindset and perseverance, were taken from the original ELS 2002 base year survey. Math and reading assessments were conducted by the ELS group, using materials adapted from previous studies. Math tests included questions on arithmetic, algebra, geometry, statistics, and other advanced material. Reading tests included comprehension questions on passages from literary, science, and social science material. Both tests were predominantly multiple-choice, although the math test did include a few open-ended questions which were scored without partial credit. For both tests, all students took a short “first-stage” test, and then were scored and assigned to a “second-stage” test based on their previous performance. This was done to allow for increased accuracy of the results given the short window of testing time and avoid ceiling and floor effects. Test scores for both reading and math are given in the dataset as standardized Z-scores, which were then averaged and re-standardized to create the “average score” variable used in this analysis. This variable has a mean of zero and a standard deviation of one.

#### Adult Outcome Measures:

Outcome variables were taken from follow-up data collected by the ELS in 2012. Outcome variables were treated to ensure that missing values were dropped in each relevant regression.

Outcomes are further defined below:

- Bachelor's Degree: Coded as 1 if respondent reported receiving a Bachelor's Degree by the 2012 follow-up survey, 0 if they reported receiving any amount of education less than a Bachelor's Degree.
- Employed: Coded as 1 if respondent reported having one or more (at least part-time) jobs, 0 for those who did not work.
- Employment Income: Self-reported annual income from employment.
- Married: Coded as 1 for all married respondents, 0 for all other domestic arrangements.
- Teen Parent: Coded as 1 for respondents who reported first having a child before or at the age of 19, 0 for respondents who reported having a child after age 19. All childless respondents were dropped.
- Registered to Vote: Coded as 1 for respondents who reported being currently registered to vote, 0 if not registered.
- Voted in Presidential Election: Coded at 1 for respondents who reported voting in the 2008 presidential election, 0 if they did not vote.
- Volunteered: Coded as 1 for respondents who reported having performed unpaid volunteer work in the past two years, 0 for those who did not.

## Appendix C

I arrive at estimates for Table 3 Panel B by disattenuating the raw correlation coefficients in Panel A using the Spearman (1904) adjustment. This adjustment is implemented by multiplying an estimated correlation between two random variables,  $x$  and  $y$ , by the inverse of the square root of the product of the reliability of each measure as follows:

$$r_{xy}^* = \frac{\hat{r}_{xy}}{\sqrt{r_{xx}r_{yy}}}$$

I calculate the reliability of the state test score measures by taking the average of the reported test-retest reliabilities in technical manuals for each state across 4<sup>th</sup> and 5<sup>th</sup> grade and then averaging these across districts. I estimate Cronbach's alpha reliabilities for the BAM and SAT9-OE as well as for the four social-emotional measures using data from all 4<sup>th</sup> and 5<sup>th</sup> grade students who participated in the MET project in Year 2. I report these reliabilities in Table C1 below.

Table C1 Estimated Reliabilities of Outcome Measures

State Math	0.924
State ELA	0.893
BAM Math	0.716
SAT9-OE Reading	0.851
Growth Mindset	0.780
Grit: Consistency	0.661
Grit: Perseverance	0.692
Effort in Class	0.561

## Appendix D

Table D1: Model-based Restricted Maximum Likelihood Estimates of Teacher Effects on State Tests, Complex Tasks and Social-Emotional Measures Excluding Gifted Students and English Language Learners (ELLs)

	n	Actual Teacher			Randomly Assigned Teacher (Intent to Treat)	
		(1)	(2)	(3)	(4)	(5)
Panel A: Results Excluding Gifted Students and ELLs						
State Math	3,117	0.194***	0.164***	0.156***	0.147***	0.137***
State ELA	3,120	0.156***	0.149***	0.148***	0.139***	0.137***
BAM Math	2,847	0.158***	0.161***	0.152***	0.137***	0.131***
SAT9-OE Reading	2,861	0.199***	0.197***	0.202***	0.196***	0.195***
Growth Mindset	2,697	0.232***	0.179**	0.177**	0.180**	0.189*
Grit: Consistency	2,633	0.074	0.099	0.069	0.081	0.059
Grit: Perseverance	2,633	0.191***	0.185***	0.191**	0.172**	0.176**
Effort in Class	2,602	0.195***	0.207***	0.246***	0.165**	0.216***
Panel B: Results Including Gifted Students and ELLs from Table 6						
State Math	4,075	0.175***	0.159***	0.150***	0.139***	0.124***
State ELA	4,074	0.142***	0.135***	0.137***	0.123***	0.123***
BAM Math	3,746	0.137***	0.143***	0.126***	0.129***	0.110**
SAT9-OE Reading	3,766	0.168***	0.177***	0.174***	0.176***	0.169***
Growth Mindset	3,551	0.196***	0.146**	0.133*	0.159***	0.154**
Grit: Consistency	3,473	0.082	0.077	0.078	0.080	0.102
Grit: Perseverance	3,473	0.149**	0.149**	0.136*	0.153**	0.138*
Effort in Class	3,435	0.162***	0.157**	0.183***	0.115*	0.149**
Survey-based Controls			Yes	Yes	Yes	Yes
Peer-level Controls		Yes		Yes		Yes
School FE		Yes				
Randomization Block FE			Yes	Yes	Yes	Yes

Notes: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Cells report estimates of the standard deviation of teacher effects from separate regressions. Columns 1 through 3 estimate the effect of 4th and 5th grade students' actual teacher while columns 4 and 5 estimate intent-to-treat effects of the teachers students were randomly assigned to via the MET classroom roster randomization process. All models include controls for students' prior achievement in math and reading, gender, age, race, FRPL, English proficiency status, special education status, and participation in a gifted and talented program. Survey-based controls include self-reported prior grades, the number of books at home, the degree to which English is spoken at home, and the number of computers at home. Peer-level controls are classroom averages of prior achievement as well as all administrative and survey-based measures described above.

Table D2: Correlations of Teacher Effects on State Tests, Complex Tasks, and Socio-Emotional Measures in a Sample Excluding Gifted Students and English Language Learners (ELLs)

	State Math	State ELA	BAM Math	SAT9-OE Reading	Growth Mindset	Grit: Consistency	Grit: Perseverance
Panel A: Results Excluding Gifted Students and ELLs							
State ELA	0.55***						
BAM Math	0.61***	0.36***					
SAT9-OE Reading	0.44***	0.29***	0.52***				
Growth Mindset	0.20**	0.06	0.06	0.15*			
Grit: Consistency	0.25***	0.18*	0.11	0.10	0.20**		
Grit: Perseverance	-0.08	-0.02	0.23**	0.28***	0.05	0.08	
Effort in Class	0.02	0.06	0.24**	0.17*	0.00	0.03	0.62***
Panel B: Results Using Full Analytic Sample from Table 7							
State ELA	0.58***						
BAM Math	0.57***	0.31***					
SAT9-OE Reading	0.38***	0.24***	0.46***				
Growth Mindset	0.21***	0.12	0.10	0.19**			
Grit: Consistency	0.17*	0.21**	0.05	-0.03	0.22***		
Grit: Perseverance	-0.04	0.00	0.10	0.19**	-0.03	0.04	
Effort in Class	0.05	0.09	0.14*	0.10	-0.08	0.06	0.61***

Notes: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001. n=207 in Panel A and n=227 in Panel B. This table reports unadjusted Pearson product-moment correlations using post-hoc predicted BLUE teacher random effect estimates derived from a model using students' actual teachers and controlling for classroom peer characteristics (Column 3 of Table 6).

## Appendix E

Table E1: Correlations of Teacher Effects on State Tests, Complex Tasks, and Social-Emotional Measures from Models Using Randomly Assigned Teachers

	State Math	State ELA	BAM Math	SAT9-OE Reading	Growth Mindset	Grit: Consistency	Grit: Perseverance
Panel A: Results from Models Using Randomly Assigned Teachers							
State ELA	0.55***						
BAM Math	0.49***	0.27***					
SAT9-OE Reading	0.35***	0.18**	0.47***				
Growth Mindset	0.17*	0.12	0.04	0.14*			
Grit: Consistency	0.07	0.10	0.00	-0.05	0.22***		
Grit: Perseverance	-0.07	-0.04	0.07	0.23**	-0.03	0.01	
Effort in Class	-0.03	0.06	0.07	0.12	-0.06	0.01	0.63***
Panel B: Results from Models Using Actual Teachers in Table 7							
State ELA	0.58***						
BAM Math	0.57***	0.31***					
SAT9-OE Reading	0.38***	0.24***	0.46***				
Growth Mindset	0.21***	0.12	0.10	0.19**			
Grit: Consistency	0.17*	0.21**	0.05	-0.03	0.22***		
Grit: Perseverance	-0.04	0.00	0.10	0.19**	-0.03	0.04	
Effort in Class	0.05	0.09	0.14*	0.10	-0.08	0.06	0.61***

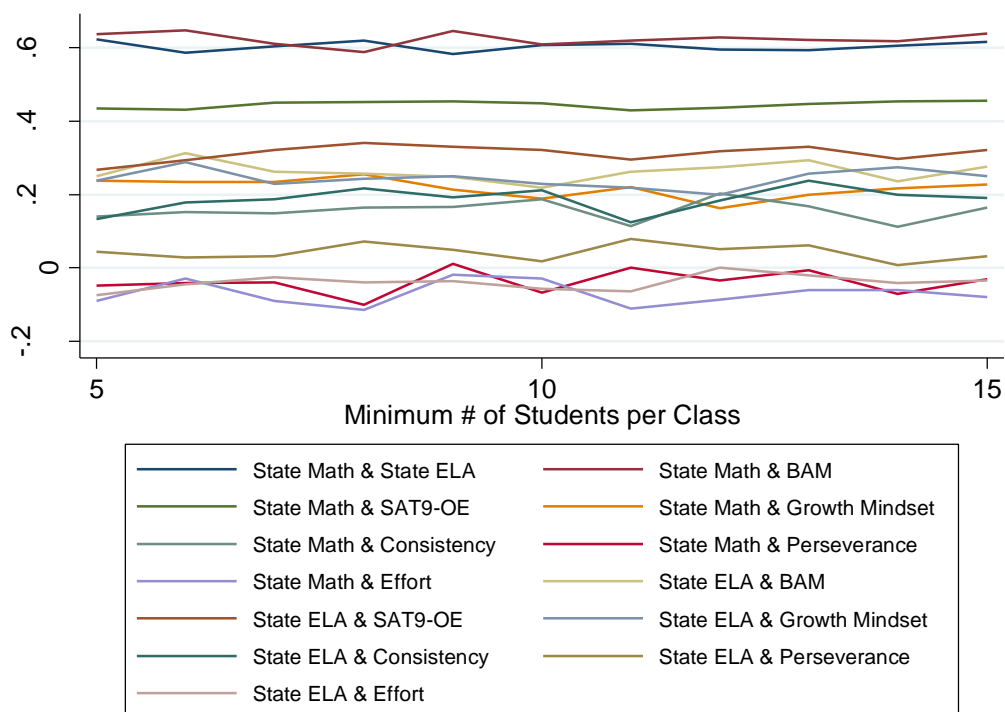
Notes: \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .  $n = 229$  in Panel A and  $n = 227$  in Panel B. This table reports unadjusted Pearson product-moment correlations using post-hoc predicted BLUE teacher random effect estimates derived from a model using students' randomly assigned teachers in Panel A, and actual teachers in Panel B, where both include controls for classroom peer characteristics (Column 5 of Table 6).



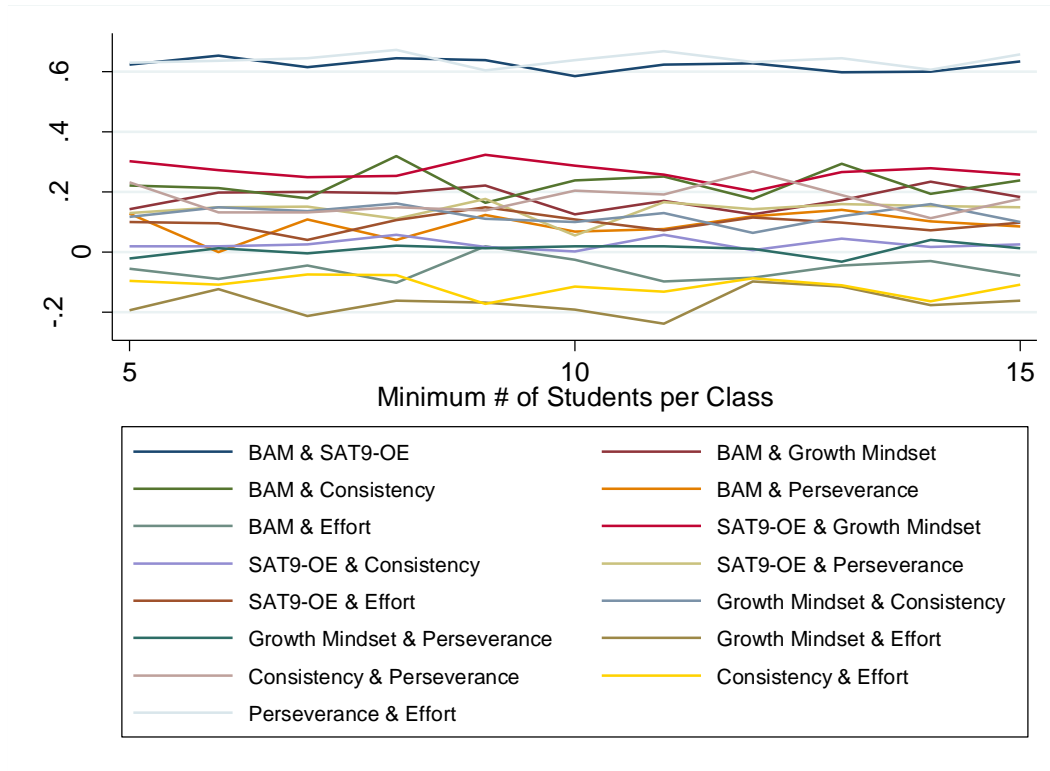
## Appendix F

I examine the sensitivity of my results in Table 7 by re-estimating the teacher effects correlation matrix using a common subsample of teachers that have a minimum of 15 students in their class (between 96 and 104 teachers across pairwise combinations). I then repeatedly drop one student per teacher and re-estimate teacher effects and a corresponding correlation matrix until the minimum class size reaches five students. Figure C1 illustrates the relative stability of the estimated correlations as the sample size increases. These findings suggest the post-hoc predicted BLUE random effect estimates I use when correlating teacher effects sufficiently correct for sampling error among this limited range.

Panel A:



Panel B:



*Figure F1: Trends in pairwise Person product-moment correlations of empirical Bayes teacher effect estimates from Table 7 across class size using successively larger minimum class size requirements. (N=96 to 104 teachers across pairwise combinations). Panel A includes correlations with state tests. Panel B includes correlations with measures of complex cognitive skills and social-emotional competencies.*

Notes: Empirical Bayes estimates are Best Linear Unbiased Estimators of teacher random effects derived from the ML model that uses students' actual teachers and includes peer controls (Column 3 of Table 6).

## Appendix G

I can disattenuate the estimated correlations for both sampling and measurement error using an approach analogous to the Spearman (1904) adjustment described in Appendix C. I estimate the reliability of teacher effects for each of the eight outcomes as follows:

$$r_{\tau_j \tau_j} = \frac{\sigma_{\tau}^2}{\sigma_{\tau}^2 + \sigma_{\varepsilon_j}^2}$$

Table 6 Column 3 provides model-based ML estimates of  $\sigma_{\tau}^2$  for each outcome. I approximate  $\sigma_{\varepsilon_j}^2$  as the average of the squared standard errors of post-hoc predicted BLUE teacher random effects from ML models ( $\overline{SE_{\tau_j}^2}$ ).

Table G1: Estimated Reliabilities of Teacher Effects

State Math	0.539
State ELA	0.592
BAM Math	0.561
SAT9-OE Reading	0.527
Growth Mindset	0.550
Grit: Consistency	0.481
Grit: Perseverance	0.544
Effort in Class	0.516

Notes: Reliabilities are estimated using sample analogues.

Table G2: Disattenuated Correlations among Teacher Effects on State Tests, Complex Tasks and Social-Emotional Measures

	State Math	State ELA	BAM Math	SAT9-OE Reading	Growth Mindset	Grit: Consistency	Grit: Perseverance
Panel A: Disattenuated Correlations							
State ELA	1.00						
BAM Math	1.00	0.62					
SAT9-OE Reading	0.64	0.45	0.79				
Growth Mindset	0.42	0.33	0.22	0.41			
Grit: Consistency	0.35	0.37	0.19	-0.04	0.43		
Grit: Perseverance	-0.11	-0.04	0.18	0.34	-0.04	0.06	
Effort in Class	0.13	0.14	0.26	0.17	-0.09	0.12	1.00
Panel B: Unadjusted Correlations							
State ELA	0.58***						
BAM Math	0.57***	0.31***					
SAT9-OE Reading	0.38***	0.24***	0.46***				
Growth Mindset	0.21***	0.12	0.10	0.19**			
Grit: Consistency	0.17*	0.21**	0.05	-0.03	0.22***		
Grit: Perseverance	-0.04	0.00	0.10	0.19**	-0.03	0.04	
Effort in Class	0.05	0.09	0.14*	0.10	-0.08	0.06	0.61***

Notes: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001. n = 227. Panel A reports disattenuated Pearson product-moment correlations using post-hoc predicted BLUE teacher random effect estimates derived from a model using students' actual teachers and controlling for classroom peer characteristics (Column 3 of Table 6). Correlations are adjusted using the Spearman (1904) correction for attenuation bias based on sample estimates of the reliability of each measure. Disattenuated correlation coefficients are set to 1 when they exceed the possible range.

Table G3. Correlations of Teacher Effects Estimated by the MET Project Using a Covariate Adjustment Model.

	State Math	State ELA	BAM Math	SAT9-OE Reading
Panel A: Same Class of Students				
State ELA	0.47			
BAM Math	0.38	0.28		
SAT9-OE Reading	0.23	0.27	0.35	
Effort in Class	0.18	0.15	0.11	0.10
Panel B: Different Classes of Students Across Years				
State ELA	0.26			
BAM Math	0.17	0.16		
SAT9-OE Reading	0.06	0.06	0.16	
Effort in Class	0.12	0.09	0.00	0.02

Notes: n=236 teachers. Table reports unadjusted Pearson product-moment correlations. Panel A captures the pooled average correlation between teacher effects from the same class using data from 2010 and 2011. Panel B captures the pooled average correlation between teacher effects from classes in different years using estimates derived from combinations where measures in columns are from 2011 and rows are from 2010 and then vice versa. Teacher effects are estimated and provided by the MET Project using a standard covariate adjustment model using students' actual teachers and including all students taught by a teacher.

## Appendix H

Table H1: Model-based Restricted Maximum Likelihood Estimates of Teacher Effects on State Tests, Complex Tasks and Social-Emotional Measures without Prior State Test Scores

	n	Actual Teacher			Randomly Assigned Teacher (Intent to Treat)	
		(1)	(2)	(3)	(4)	(5)
Panel A: Results from Models without Prior State Test Scores						
State Math	4,075	0.178***	0.179***	0.152***	0.157***	0.127***
State ELA	4,074	0.185***	0.189***	0.185***	0.167***	0.162***
BAM Math	3,746	0.173***	0.187***	0.170***	0.166***	0.140***
SAT9-OE Reading	3,766	0.178***	0.186***	0.185***	0.182***	0.177***
Growth Mindset	3,551	0.201***	0.153**	0.141*	0.159**	0.161**
Grit: Consistency	3,473	0.106	0.099	0.105	0.103	0.117*
Grit: Perseverance	3,473	0.155***	0.155**	0.142*	0.154**	0.141*
Effort in Class	3,435	0.169***	0.161**	0.183***	0.119*	.142*
Panel B: Results from Models with Prior State Test Scores from Table 6						
State Math	4,075	0.175***	0.159***	0.150***	0.139***	0.124***
State ELA	4,074	0.142***	0.135***	0.137***	0.123***	0.123***
BAM Math	3,746	0.137***	0.143***	0.126***	0.129***	0.110**
SAT9-OE Reading	3,766	0.168***	0.177***	0.174***	0.176***	0.169***
Growth Mindset	3,551	0.196***	0.146**	0.133*	0.159***	0.154**
Grit: Consistency	3,473	0.082	0.077	0.078	0.080	0.102
Grit: Perseverance	3,473	0.149**	0.149**	0.136*	0.153**	0.138*
Effort in Class	3,435	0.162***	0.157**	0.183***	0.115*	0.149**
Survey-based Controls			Yes	Yes	Yes	Yes
Peer-level Controls		Yes		Yes		Yes
School FE		Yes				
Randomization Block FE			Yes	Yes	Yes	Yes

Notes: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Cells report estimates of the standard deviation of teacher effects from separate regressions. Columns 1 through 3 estimate the effect of 4th and 5th grade students' actual teacher while columns 4 and 5 estimate intent-to-treat effects of the teachers students were randomly assigned to via the MET classroom roster randomization process. Panel A omits prior measures of student achievement on state standardized tests in math and reading while Panel B includes these measures. All models include controls for students' gender, age, race, FRPL, English proficiency status, special education status, and participation in a gifted and talented program. Survey-based controls include self-reported prior grades, the number of books at home, the degree to which English is spoken at home, and the number of computers at home. Peer-level controls are classroom averages of prior achievement as well as all administrative and survey-based measures described above.

Table H2: Correlations of Teacher Effects on State Tests, Complex Tasks, and Socio-Emotional Measures from Models without Prior State Test Scores

	State Math	State ELA	BAM Math	SAT9-OE Reading	Growth Mindset	Grit: Consistency	Grit: Perseverance
Panel A: Results from Models without Prior State Test Scores							
State ELA	0.73***						
BAM Math	0.73***	0.55***					
SAT9-OE Reading	0.47***	0.40***	0.55***				
Growth Mindset	0.27***	0.25***	0.20**	0.26***			
Grit: Consistency	0.37***	0.40***	0.29***	0.14*	0.27***		
Grit: Perseverance	0.07	0.13	0.20**	0.26***	0.02	0.13	
Effort in Class	0.17*	0.21**	0.23**	0.19**	-0.02	0.15*	0.63***
Panel B: Results from Models with Prior State Test Scores from Table 7							
State ELA	0.58***						
BAM Math	0.57***	0.31***					
SAT9-OE Reading	0.38***	0.24***	0.46***				
Growth Mindset	0.21***	0.12	0.10	0.19**			
Grit: Consistency	0.17*	0.21**	0.05	-0.03	0.22***		
Grit: Perseverance	-0.04	0.00	0.10	0.19**	-0.03	0.04	
Effort in Class	0.05	0.09	0.14*	0.10	-0.08	0.06	0.61***

Notes: \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .  $n = 227$ . This table reports unadjusted Pearson product-moment correlations using post-hoc predicted BLUE teacher random effect estimates derived from a model using students' actual teachers and controlling for classroom peer characteristics (Column 3 of Table 6).

## Appendix I

Table II: Unshrunk and Shrunk Average Residual Estimates of Teacher Effects on State Tests, Complex Tasks and Social-Emotional Measures

	n	Actual Teacher			Randomly Assigned Teacher (Intent to Treat)	
		(1)	(2)	(3)	(4)	(5)
Panel A: Unshrunk Average Class Residuals						
State Math	4,075	0.171	0.150	0.131	0.141	0.122
State ELA	4,074	0.153	0.140	0.131	0.135	0.126
BAM Math	3,744	0.168	0.158	0.141	0.148	0.133
SAT9-OE Reading	3,766	0.208	0.197	0.177	0.197	0.178
Growth Mindset	3,551	0.263	0.220	0.194	0.228	0.209
Grit: Consistency	3,473	0.213	0.190	0.175	0.199	0.189
Grit: Perseverance	3,473	0.237	0.222	0.202	0.223	0.203
Effort in Class	3,435	0.250	0.228	0.212	0.208	0.202
Panel B: Shrunk Average Class Residuals						
State Math	4,075	0.112	0.086	0.059	0.045	0.008
State ELA	4,074	0.078	0.062	0.050	0.011	0.023
BAM Math	3,744	0.077	0.065	0.034	0.002	0.060
SAT9-OE Reading	3,766	0.102	0.087	0.055	0.040	0.003
Growth Mindset	3,551	0.150	0.076	0.008	0.018	0.128
Grit: Consistency	3,473	0.032	0.025	0.000	0.000	0.000
Grit: Perseverance	3,473	0.072	0.051	0.001	0.000	0.000
Effort in Class	3,435	0.098	0.066	0.039	0.059	0.111
Survey-based Controls			Yes	Yes	Yes	Yes
Peer-level Controls		Yes		Yes		Yes
School FE		Yes				
Randomization Block FE			Yes	Yes	Yes	Yes

Notes: Cells in panel A report estimates from separate models of the standard deviation of teacher effects estimated by averaging student-level residuals from an OLS model to the teacher level. Cells in panel B report these same estimates from separate models when shrunk towards the grand mean based on the reliability of a teacher's individual estimate. Statistical significance not calculated. All models include controls for students' prior achievement in math and reading, gender, age, race, FRPL, English proficiency status, special education status, and participation in a gifted and talented program. Survey-based controls include self-reported prior grades, the number of books at home, the degree to which English is spoken at home, and the number of computers at home. Peer-level controls are classroom averages of prior achievement as well as all administrative and survey-based measures described above.