Multi-Dimensional Teacher Effects

Elizabeth Santorella *

October 30, 2017

Abstract

I estimate the covariance structure of teacher effects on several outcomes: present and future test scores, present and future attendance, and high school graduation. Studying the covariance matrix of teacher effects reveals the magnitude of teacher effects on each outcome and the relationship between teacher effects on different outcomes while sidestepping the need to estimate individual teacher effects.

Teachers have substantial effects on high school graduation, and on test scores and attendance four years in the future. Students of a teacher who is one standard deviation above average at improving graduation rates are 5 percentage points more likely to graduate high school. Although teacher quality is an important determinant of graduation rates and of future test scores and attendance, long-term effects cannot be predicted well by short-term effects. For example, even if teacher effects on contemporaneous outcomes were perfectly measured, they would only explain about 3% of the variance in teacher effects on high school graduation. My results also suggest that teacher effects on attendance could be an important supplement to score-based measures of teacher value-added. Teachers who improve attendance tend to improve high school graduation rates. However, teacher effects on attendance are only weakly correlated with effects on test scores. I also use Principal Components Analysis to show that the correlation matrix of teacher effects can be well-represented by three easily interpretable components.

Previous work has established that teachers vary widely in their effects on student test scores. Many large school districts now incorporate test score-based value-added measures in teacher evaluation. However, measures of teacher quality based on short-run test scores may be highly incomplete, because they neglect teacher effects both on non-test score measures and on long-term outcomes. Education influences students' "non-cognitive skills" or "character skills"; these skills, important in life and the labor market, are not well-captured by test scores (Carneiro *et al.*, 2007). If teachers influence their students in ways that are not reflected in test scores, score-based value-added measures miss important components of teacher quality. Furthermore, if teacher effects on short-run outcomes are poor proxies for teacher effects on long-run outcomes, value-added measures based on short-term measurements will be highly incomplete. In this paper, I estimate the variance of teacher effects on students' test scores and attendance, both contemporaneously and up to four years in the future, and on high school graduation. I estimate the covariance structure of teacher effects on various outcomes and, using these covariances, investigate how correlated

*Many thanks to Gary Chamberlain, Christopher Jencks, Raj Chetty, Larry Katz, Isaac Opper, and seminar participants at Harvard University.

teacher effects are in different domains, such as math test scores and attendance; how quickly the effects of teachers who improve a short-term outcome fade out; and how well short-term teacher effects predict long term teacher effects.

Using administrative data from the New York City public elementary and middle schools, I find that teachers' causal effects on attendance — their "attendance value-added" — are highly variable; a teacher whose attendance value-added is one standard deviation above average improves her students' attendance by about 0.07 standard deviations, or 1.5 days. Much like teacher effects on test scoes, these effects fade out quickly: students of a teacher who improves test scores or attendance have test scores or attendance four years later that are barely better than would be predicted by demographic characteristics.

However, teacher quality is an important determinant of future test scores and of future attendance. This is not a paradox: the teachers who improve test scores in the long term are not particularly likely to be the teachers who improve test scores in the short term. The correlation between a teacher's effect on test scores and her effect on attendance is only about 0.12 to 0.15 [1], implying that teacher quality metrics that use only test scores miss much of a teacher's impact on her students. I also find that a teacher who is one standard deviation above average in improving her students' high school graduation rates increases graduation rates by 5 percentage points and graduation with an Advanced Regents designation by 6 percentage points.

I build on the teacher value-added literature. Teacher value-added methods estimate both the variance of teachers' causal effects on their students' outcomes and an individual causal effect, or "value-added", for each teacher. Value-added models typically use variance decompositions and "moment-matching" to estimate what portion of variance in test scores is due to teachers. These models must avoid giving teachers credit for receiving more able students rather than for their causal effects on test scores. They typically achieve this by controlling for a rich set of student covariates, such as demographic factors and previous test scores; thus, these models estimate the value a teacher *adds* to a student above what that student would achieve with an average teacher. This literature typically finds that teachers vary largely in their effects on students: teachers account for about 1% of the variance in test scores. In other words, a teacher who is one standard deviation above average in her effectiveness at increasing test scores ("score VA") increases her students' test scores by an average of 0.1 standard deviations. I extend this methodology by incorporating a variety of different outcomes, including leads of outcomes, and estimating not only the variance of teacher effects but the covariance of effects on different measures.

Value-added measures based on short-term test scores have become a common component of teacher evaluations, but there are reasons to suspect that these measures are incomplete. Chetty *et al.* (2014b) shows that students of teachers who improve test scores are more likely to go to college, have higher incomes, and live in better neighborhoods as adults, but these effects appear to be too large to be explainable by test score gains. And studies that do not specifically involve teachers find that quantity and quality of education improve skills that are not captured by test scores, but the mechanisms for this are unclear.

Rewarding teachers based on test scores is unpopular with many parents, partially due to concerns that test scores reflect only part of teachers' beneficial effects on their students. If

[1] By contrast, the correlation between teacher effects on math test scores and on English Language Arts test scores is 0.66. See Table 10.

so, policymakers face a multitasking problem in the spirit of Holmstrom and Milgrom (1991): we want teachers to make their students motivated persistent, and informed, but we can only design contracts on the basis of observable factors. Teachers who are incentivized to increase test scores may behave in counterproductive ways. This concern has empirical merit: Teachers or administrators under low to moderate incentives to improve test scores cheat or manipulate scores (Jacob and Levitt (2003), Dee *et al.* (2016), Loughran and Comiskey (1999)), spend less time on non-tested subjects (Jacob, 2005), spend much more time on test preparation (Klein *et al.* (2000)), and move students into special education so that those students will not be counted school progress indicators (Figlio and Getzler (2002), Jacob (2005)).

One area that could be given short shrift by an increasing focus on test scores is non-cognitive, or character, skills. Education is important in transmitting these skills, such as the drive and persistence to attend school and work hard, and while the mechanisms are poorly understood, teachers may be an important factor. Recent papers show that teachers influence their students' non-cognitive abilities in the short term. Gershenson (2016) studies third through fifth graders in North Carolina and finds that teachers have "arguably causal, statistically significant effects on student absences that persist over time," and that "teachers who improve test scores do not necessarily improve student attendance." Similarly, Jackson (2016) studies ninth graders in North Carolina and finds that teachers have medium-term effects on student absences, suspensions, grades, and on-time grade progression.

Recent policy changes hint at a shift away from a focus on test scores and a drive towards quantitative evaluations that incorporate other metrics. The federal Every Student Succeeds Act (ESSA) mandates that each state measure school quality based on a metric of its own devising that includes test scores but also includes at least one substantially different outcome. Many states have chosen to measure and reward attendance.

In this paper, I demonstrate that teachers influence medium-term outcomes, that immediate effects on test scores are a poor proxy for medium-term effects, and that teacher effects on attendance are persistent and uncorrelated with other aspects of teacher value-added.

My estimates are only credible insofar as identification restrictions are satisfied. Teachers must be sorted to students only on observables: Conditional on covariates, no teacher should be systematically assigned students who have unobservable characteristics that cause high or low performance. The rich, longitudinal nature of my data makes it possible to control for a variety of student and classroom characteristics and lagged values of outcomes, making the sorting on observables requirement plausible. For example, it is possible that high-SES students or students who have been improving relative to their peers are, on average, assigned to better teachers. But since I observe and control for ethnicity, free lunch status, lagged values of test scores and attendance, and many interactions – among other variables recommended by the teacher value-added literature – this sorting would be predictable from the control variables and would not violate the sorting on observables restriction. I use a pre-trend test as an empirical check of this restriction. Although other authors have found that controlling for lagged scores is sufficient to find unbiased measures of parameter estimates, I measure a small but statistically significant pre-trend Chetty *et al.* (2014a) in which teacher effects on contemporaneous outcomes "predict" past achievement.

This paper also provides descriptive evidence on patterns of absenteeism. The patterns documented are consistent with poor and minority students often missing school voluntarily or for reasons other than illness. Students in New York City are absent extremely often, and chronic absenteeism — missing more than 10% of a school year — is common. Students are far more likely to be absent in later grades, and there are large ethnic gaps in school attendance.

This paper proceeds as follows. In Section 1, I recap the literature on teacher value-added and the influence of education on non-cognitive skills. In Section 2, I develop a model in which student outcomes like test scores or attendance are a function of teacher effects, covariates, and random shocks. In Section 3, I describe the data and provide descriptive evidence on the pervasiveness of poor attendance in the New York City public schools and correlates of poor attendance. Section 4 describes the estimation procedure and the conditions under which parameters of interest are identified. Section 5 contains results on the distribution of teacher effects on student attendance and test scores; the persistence of these effects; the usefulness of contemporaneous effects for predicting teacher effects on future student achievement; and principal components analysis of the correlation matrix of teacher effects. Section 6 demonstrates that results are robust to using a different estimator – maximum likelihood rather than moment-matching – and to different choices of covariates. Section 7 concludes.

1 Literature Review

Skills that aren't well captured by traditional educational metrics, often termed "non-cognitive skills" or "character skills", are correlated with many outcomes, including earnings, educational attainment, health, and crime.

However, there is little direct evidence on the degree to which education — and teachers in particular — affects character skills, despite the growing literature on teacher effects on test scores, and the use of test score-based value-added measures in large school systems such as New York City, Los Angeles, Chicago, and Washington, DC. Jackson (2016) and Gershenson (2016) examine the effects of teachers on outcomes other than test scores. Gershenson studies third through fifth graders in North Carolina and finds that teachers have "arguably causal, statistically significant effects on student absences that persist over time," and that "teachers who improve test scores do not necessarily improve student attendance." Similarly, Jackson (2016) studies ninth graders in North Carolina and finds that teachers have medium-term effects on student absences, suspensions, grades, and on-time grade progression, and that teacher effects on test scores have modest correlations with teacher effects on behavioral variables. This paper is similar in spirit and methodology, but tracks teacher effects over a longer time period.

There are reasons to suspect that teachers may affect their students in the long term in ways that are not captured in test scores. First, increases in the quality or quantity of education are correlated with measures of socioeconomic success, even after conditioning on test scores. For example, Heckman and Rubinstein (2001) shows that conditional on AFQT scores, GED recipients earn less than high school graduates who do not attend college. (Chetty *et al.*, 2014b) show that teachers who improve test scores also cause their students to have higher incomes, attend college, and live in better neighborhoods. The mechanisms for this are unclear, since teacher effects on test scores fade out dramatically after several years (Chetty *et al.* (2014b), Chetty *et al.* (2011)); test scores effects alone do not seem sufficient to explain the magnitude of teacher effects on long-term outcomes. Evidence from Project STAR suggests that kindergarten "class quality has significant impacts on non-cognitive measures in fourth and eighth grade such as effort, initiative, and lack of disruptive behavior," and that high-quality kindergarten classes improve test scores in the short run, but it is not clear whether these effects are due to teacher quality, peer effects, or some other factor. In summary, teachers may impact their students' persistence and motivation, and these effects may be more meaningful or persistent than teacher effects on test scores.

Ultimately, we would like to understand whether teachers affect their students' behavioral outcomes, whether teachers who improve behavioral outcomes have meaningful and persistent effects on their students, and whether teachers recruitment, training, or compensation should be influenced by quantitative measures of teacher effects on student behavior. In this project, I demonstrate that teachers have moderately persistent effects on student behavior, and that these effects are not highly correlated with teacher effects on test scores.

2 Model

I follow Chamberlain (2013) in developing a model that defines teacher effects as best linear predictor coefficients. In order to preserve independence across teachers, I ensure that the same student does not appear in the data more than once by treating each grade separately and by dropping students who repeat a grade. (After estimating the covariance of teacher effects for each grade, I pool across grades with a frequency-weighted mean.) Therefore, observations are at the student level, so I index by student i and use $j(i)$ to refer to the teacher of student i and $c(i)$ to refer to teacher-year cohorts. That is, if $j(i) = j(i')$ but $c(i) \neq c(i')$, then students i and i' are taught by the same teacher but in different years. Subscripts index observations, and superscripts index outcomes. There are $H = 31$ outcomes: Math test scores, reading test scores, attendance z-scores, four years of leads for each of those variables, and indicators forfour-year high school graduation, four-year high school graduation with a Regents diploman, and four-year high school graduation with Advanced Regents diploma. Outcomes are $y_i \in \mathbb{R}^H$ and $x_i \in \mathbb{R}^{H \times K}$. Teacher j's effect on outcomes, $\mu_j \in \mathbb{R}^H$, is defined as a best linear predictor coefficient. For each outcome h,

$$E^* \left[y_i^h \middle| x_i, \mu_{j(i)}^h \right] = x_i^T \beta^h + \mu_{j(i)}^h \tag{1}$$

We can also define errors ν_i:

$$y_i^h \equiv x_i^T \beta^h + \mu_{j(i)}^h + \nu_i$$

In order to interpret parameter estimates causally, we need sorting on observables. This means, first, that variation in teacher effects that cannot be captured in covariates must be orthogonal to ν: $\mu_{j(i)} \perp \nu_i \,|x_i$. This restriction would be violated if, say, better teachers tend to teach high-ability students, and ability is not well-captured by x. Another, more subtle restriction is that unobservable shocks to outcomes need to be independent of the teacher's identity, conditional on covariates. This restriction is necessary so that, when estimating variances, we don't mistake the tendency for some teachers to consistently receive better or worse students for the presence of teachers who consistently teach well or poorly. Imagine that all teachers are identical — $\mu_j = 0 \quad \forall j$ — but some teachers are consistently assigned students with high or low values of ν. Some teachers will consistently have students who over- or under-perform what would be expected from their covariates, making it appear that teachers vary in quality when they do not.

I further assume that errors across different classrooms are orthogonal: $\mathbb{E}\left[\nu_i \nu_{i'} \,|c(i) \neq c(i')\,\right] = 0$. The parameter of interest is the covariance matrix of teacher effects,

$$\begin{aligned} \text{Var}\left(\mu_j\right) &= \text{Var}\left(y_i - x_i^T \beta\right) \\ &\equiv \Sigma_\mu \in \mathbb{R}^{H \times H} \end{aligned} \tag{2}$$

3 Data, Setting, and Descriptive Statistics

The data includes almost all New York City public school students in grades preschool through 12 in the 2001-02 to 2015-16 school years. I observe rich individual-level data and can track students across years. For each student, I can observe several outcomes of interest: Test scores, attendance, and what type of diploma the student received. New York's tiered high school diploma system is explained below.

While my data includes 9.4 million student-year observations that include demographic and attendance information, the effective size of the data is somewhat smaller. I can only link teachers to students starting in the 2005-06 school year, although I use observations from earlier years to construct lagged variables and for pre-trend tests. I avoid dropping observations due to missing data in independent variables. Instead, I impute the missing field as the average for that student's grade and year and also use an indicator variable for missingness. This happens most often when lagged variables are missing because the student recently moved into the district. However, I do require one lagged test score to be present.

I observe demographic information on each student. My data contains each student's ethnicity and date of birth, which are filled in by parents when the student enters the school system. Other information is recorded by the school administration: grad level, number of days absent, number of days present, and whether the student has a "team teaching" arrangement, as is common in special education classrooms. I also observe the student's registrar data, which explains whether the student is still enrolled, whether the teacher has graduated and with what type of diploma, whether the student has a disability requiring an Individualized Education Program (IEP), and whether the student has dropped out.

New York State has a tiered system of high school diplomas. High school students must take standardized examinations known as Regents exams, and those who pass exams in global history, U.S. history, ELA, math, and science graduate with a "Regents diploma." Until the 2011-2012 school year, students who met their high school's graduation but did not meet the requirements for a Regents diploma earned a less prestigious "local diploma"; now, students cannot earn a local diploma unless they have a disability. There also exist diplomas that are harder to attain than a Regents diploma. [2] Students who pass additional

[2]When the following results refer to a Regents Diploma, this refers to a Regents Diploma that is *not* with an Advanced Designation, so it is ambiguous whether increasing the number of Regents Diplomas is positive.

Table 1: *Student summary statistics, for students who can be matched to teachers.*

	Mean	St. Dev	Min	Max	Missing
Grade	5.98	1.42	4	8	0%
Year	2009.21	2.08	2006	2013	0%
Disabled	0.16	0.37	0	1	0%
Female	0.49	0.50	0	1	0%
English Language Learner	0.12	0.32	0	1	0%
Free Lunch	0.83	0.37	0	1	0%
Days absent	11.43	11.95	0	182	0%
Days present	169.83	12.96	2	186	0%
Days Absent Lag (Z-Score)	0.06	0.89	-14.66	1.34	2.91%
Math Score (Z-Score)	0.07	0.97	-6.35	3.89	0%
Math score lag (Z-Score)	0.06	0.97	-10	3.96	5.34%
ELA Score (Z-Score)	0.05	0.98	-11.1	7.76	0%
ELA Score Lag (Z-Score)	0.05	0.97	-11.1	7.76	8.08%
4-Year Graduation	0.68	0.47	0	1	66.47%
4-Year Graduation, Regents Diploma	0.45	0.50	0	1	66.46%
4-Year Graduation, Advanced Regents Diploma	0.20	0.40	0	1	66.45%

exams earn a Regents Diploma with Advanced Designation, and students who satisfy the requirements of the Advanced Designation and attain high scores can attain a "Regents with Advanced Designation with Honors."

In addition to the high school Regents exams, students take standardized math and English Language Arts (ELA) tests every year in third through eighth grade. Starting in spring 2013, these tests have been based on Common Core standards.

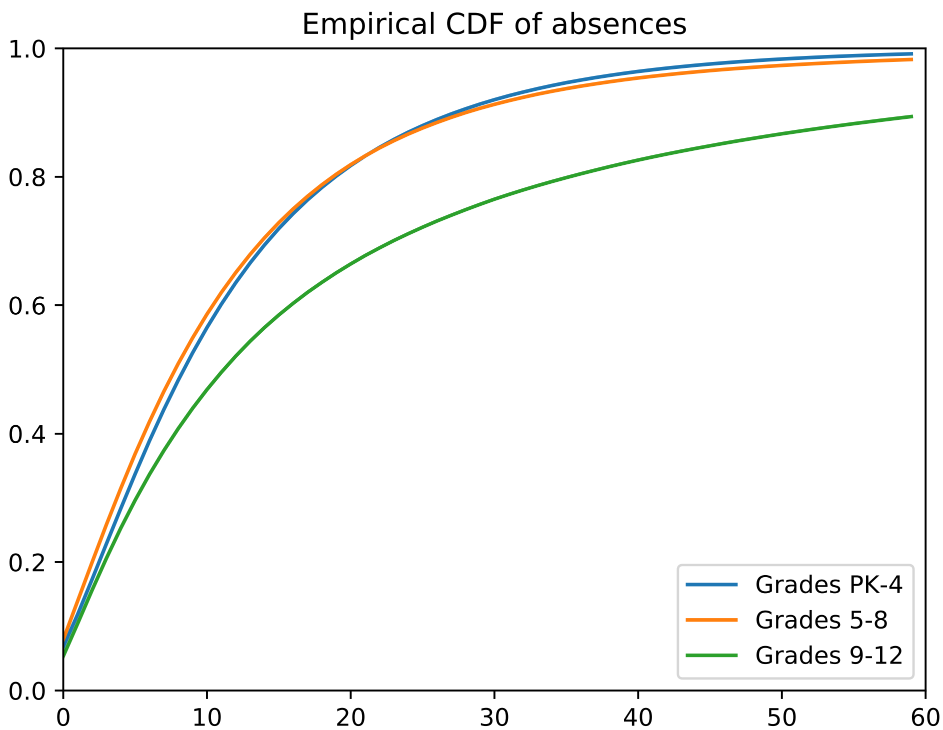
Math and ELA scores are scored on a scale that varies every year. Therefore, I normalize test scores to have a mean of zero and variance of 1 within each grade and year. Normalization obscures the fact that New York City's test scores rose dramatically over this period, both in terms of the percentage of students scoring at the proficient level and in comparison to the rest of New York State. Four-year graduation rates rose from 45.5% for students starting ninth grade in 2001 to 70.5% for students starting ninth grade in 2011.

3.1 Descriptive Statistics

Table 1 contains summary statistics for the sample that is used to estimate value-added, containing students who can be matched to teachers. (Other descriptive statistics in this section include high school students, who cannot be matched to teachers.) The district is relatively poor, with 77% of students qualifying for free or reduced-price lunch. A plurality of students are Hispanic. Although about 43% of students, as of 2013, speak a language other than English at home, only 13% of students are classified as English Language Learners (ELLs). English Language Learners are students who either take a class in English as a New Language or participate in bilingual education.

Teachers also have a unique anonymized identifier that is consistent across years.

Figure 1: *Empirical CDF of absences in elementary schools (grades PK through 4), middle schools (grades 5 through 8), and high schools (grades 9-12).*



3.1.1 Attendance

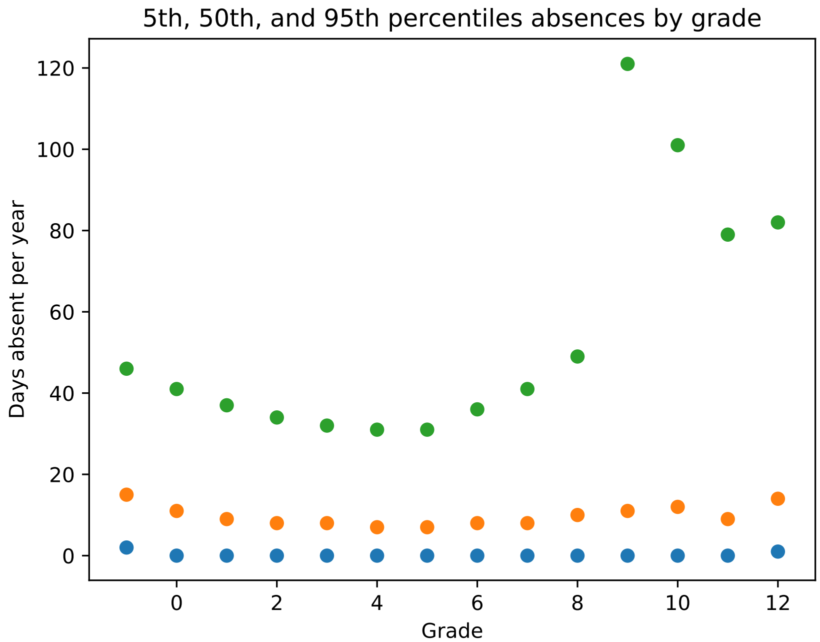
Student absence is frequent in the New York City schools. Descriptive, non-causal evidence suggests that attendance matters for student achievement. Poor attendance is associated with lower test scores and a lower likelihood of graduating high school, and there are large socioeconomic gaps in attendance. Although this data cannot explain why students miss school so often, the data is consistent with the hypothesis that students are absent far more often than necessitated by illness.

The average New York City public school student is absent 16 days in an approximately 180-day school year, or 9%. By comparison, the average student nationally is absent on about 7% of days.

Attendance deteriorates dramatically across grades, as shown in Figure 1. In elementary, middle, and high school, about 8% of students are never absent. However, high school students are absent far often than elementary or middle school students, especially in the right tail. Figure 2 shows this by plotting the fifth, twenty-fifth, fiftieth, seventy-fifth, and ninetieth percentiles of absences within each grade.

Figure 3 illustrates the large ethnic gaps in school attendance. In any grade, Hispanic, Black, and Native American students are absent almost twice as often as Asian students,

Figure 2: *Percentiles of attendance in each grade.*



with non-Hispanic White and multi-racial students in the middle.

Tables 2, 3, and 4 show coefficients from regressing ELA scores, math scores, and attendance on student demographic characteristics, with and without lagged values of outcomes. All outcomes have been z-scored within grade-year, so coefficients are comparable in magnitude. A student with an attendance z-score of 1 is absent one standard deviation *less* than other students in her grade and year. Although coefficients on regressions with ELA scores (Table 2) and wtih math scores (Table 3) as the dependent variable have similar coefficients, they are not similar to the coefficients from predicting attendance (Table 4). While all coefficients are highly statistically significant, the only other coefficients that are large in magnitude are coefficients on lagged values of outcomes. Despite the large ethnic gaps in school attendance, attendance is hard to predict (within grade-years), with an R^2 of only 0.06 in a regression that includes ethnicity, indicators for common home languages, year dummies, and a student's status as disabled, ELL, receiving free lunch, female, or in special education. R^2 rises to 0.179 in a regression that includes lagged values of attendance and test scores. Test scores, however, are much more predictable, generating R^2 values of 0.30 to 0.49 across specifications.

Figures 4 and 5 show the association between attendance and test scores by plotting the fifth, fiftieth, and ninety-fifth percentiles of test scores within each percentile bin of attendance. Across most of the distribution of attendance, students who are absent more often score moderately worse: the median student in the lowest percentile of absences scores around the sixty-fifth percentile on math test, and the median student at the eightieth percentile of absences scores around the fortieth percentile on math tests. But above the

Figure 3: *Average number of days absent by ethnicity.*

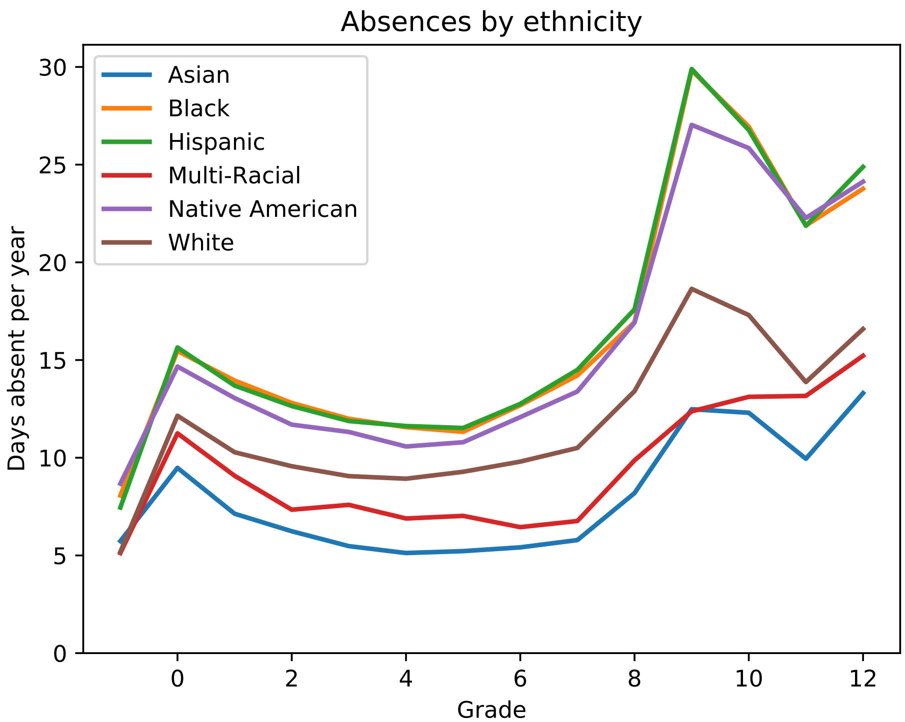


Table 2: *Predictors of English Language Arts scores. For categorical variables – ethnicity, home language, and year – an F-statistc and its p-value are shown. Other variables have standard errors in parentheses. *** indicates significance at the 0.1% level.*

	0	1	2
Constant	0.797***	0.513***	0.477***
	(0.003)	(0.002)	(0.002)
English Language Learner	-0.868***	-0.553***	-0.554***
	(0.002)	(0.001)	(0.001)
Female	0.147***	0.125***	0.123***
	(0.001)	(0.001)	(0.001)
Free Lunch	-0.306***	-0.194***	-0.179***
	(0.001)	(0.001)	(0.001)
IEP	-0.829***	-0.497***	-0.475***
	(0.001)	(0.001)	(0.001)
Attendance (Z Score)			0.13***
			(0.001)
Missing Attendance			-0.312***
			(0.006)
Lag ELA Score		0.379***	0.379***
		(0.001)	(0.001)
Lag Math Score		0.198***	0.196***
		(0.001)	(0.001)
Lagged Attendance		0.018***	-0.068***
		(0.001)	(0.001)
Lagged Values Missing		-0.106***	-0.105***
		(0.001)	(0.001)
Ethnicity	29869.5	13195.2	12461.8
	(0.0)	(0.0)	(0.0)
Home Language	1694.84	689.174	595.075
	(0.0)	(0.0)	(0.0)
Year	546.815	173.296	151.73
	(0.0)	(0.0)	(0.0)
N	3.9 M	3.9 M	3.9 M
R-squared	0.305656	0.446884	0.456547

Table 3: *Predictors of math scores. For categorical variables – ethnicity, home language, and year – an F-statistc and its p-value are shown. Other variables have standard errors in parentheses. *** indicates significance at the 0.1% level.*

	0	1	2
Constant	0.87***	0.557***	0.505***
	(0.003)	(0.002)	(0.002)
English Language Learner	-0.655***	-0.37***	-0.382***
	(0.001)	(0.001)	(0.001)
Female	-0.02***	-0.014***	-0.018***
	(0.001)	(0.001)	(0.001)
Free Lunch	-0.266***	-0.151***	-0.131***
	(0.001)	(0.001)	(0.001)
IEP	-0.782***	-0.434***	-0.404***
	(0.001)	(0.001)	(0.001)
Attendance (Z Score)			0.191***
			(0.0)
Missing Attendance			-0.756***
			(0.005)
Lag ELA Score		0.11***	0.109***
		(0.001)	(0.001)
Lag Math Score		0.49***	0.486***
		(0.001)	(0.001)
Lagged Attendance		0.066***	-0.06***
		(0.001)	(0.001)
Lagged Values Missing		-0.113***	-0.102***
		(0.001)	(0.001)
Ethnicity	41380	20031.1	18964.4
	(0.0)	(0.0)	(0.0)
Home Language	6136.21	3397.84	2853.06
	(0.0)	(0.0)	(0.0)
Year	130.131	664.138	505.403
	(0.0)	(0.0)	(0.0)
N	4.0M	4.0M	4.0M
R-squared	0.297017	0.463534	0.486825

Table 4: *Predictors of attendance. For categorical variables – ethnicity, home language, and year – an F-statistc and its p-value are shown. Other variables have standard errors in parentheses. *** indicates significance at the 0.1% level.*

	0	1
Constant	0.462***	0.413***
	(0.002)	(0.002)
English Language Learner	-0.046***	0.011***
	(0.001)	(0.001)
Female	0.027***	0.021***
	(0.001)	(0.001)
Free Lunch	-0.125***	-0.093***
	(0.001)	(0.001)
IEP	-0.306***	-0.234***
	(0.001)	(0.001)
Lag ELA Score		-0.003***
		(0.001)
Lag Math Score		-0.01***
		(0.001)
Lagged Attendance		0.645***
		(0.001)
Lagged Values Missing		-0.095***
		(0.001)
Ethnicity	18691.9	13103.3
	(0.0)	(0.0)
Home Language	11216.6	6512.18
	(0.0)	(0.0)
Year	26.3365	261.221
	(0.0)	(0.0)
N	9.4M	9.4M
R-squared	0.064287	0.178718

Table 5: *Predictors of attendance. For categorical variables – ethnicity, home language, and year – an F-statistc and its p-value are shown. Other variables have standard errors in parentheses. *** indicates significance at the 0.1% level.*

	0	1
Constant	0.513*	0.355*
	(0.31)	(0.284)
English Language Learner	-0.156***	-0.032***
	(0.002)	(0.002)
Female	0.07***	0.058***
	(0.001)	(0.001)
Free Lunch	-0.097***	-0.034***
	(0.001)	(0.001)
IEP	-0.206***	-0.032***
	(0.001)	(0.001)
Attendance		0.143***
		(0.001)
Chronically Absent		0.003*
		(0.002)
ELA score		0.059***
		(0.001)
Math score		0.107***
		(0.001)
Ethnicity	2441.58	343.839
	(0.0)	(0.0)
Home Language	328.433	55.3449
	(0.0)	(0.0)
Year	213.836	198.975
	(0.0)	(0.0)
N	823370	823370
R-squared	0.112862	0.255901

Figure 4: *Dots plot the fifth, fiftieth, and ninety-fifth percentiles of math test scores, within each percentile of attendance. The solid line is a quadratic trend.*

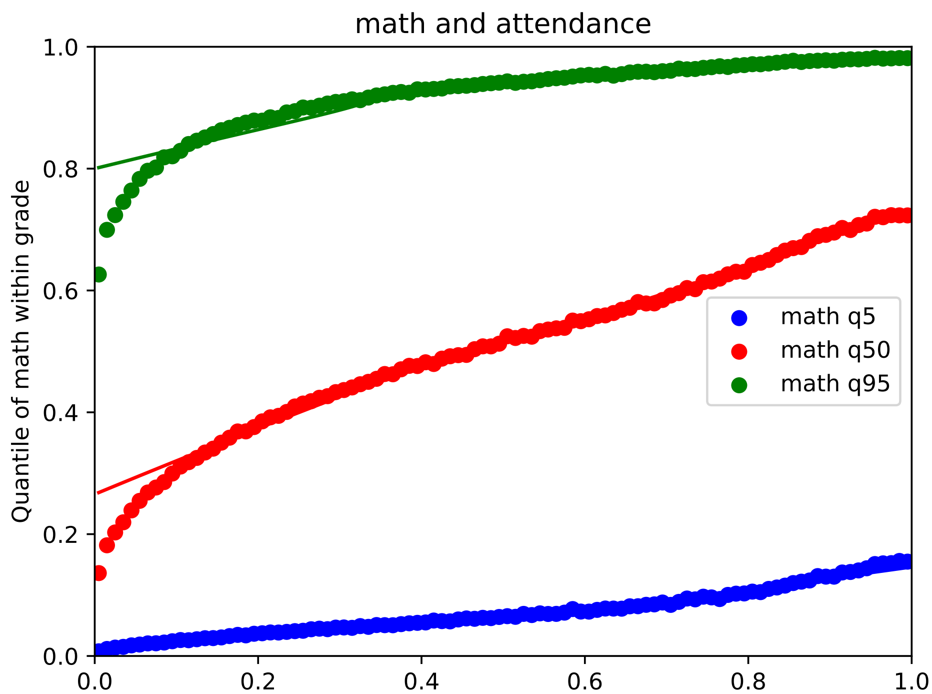
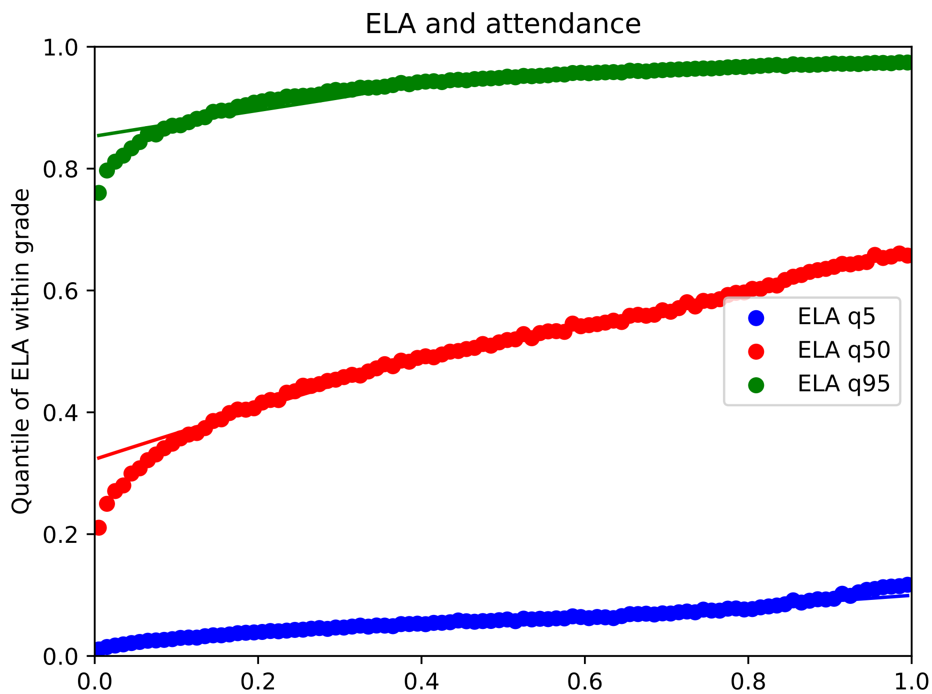


Figure 5: *Dots plot the fifth, fiftieth, and ninety-fifth percentiles of English Language Arts test scores, within each percentile of attendance. The solid line is a quadratic trend.*



eightieth percentile, math scores deteriorate rapidly, with the median student in the worst percentile of attendance scoring below the fifteenth percentile on math tests. Reading scores are less strongly related to attendance.

4 Estimation

I estimate the covariance of teacher effects, Σ_μ, using a moment-matching procedure similar to that of Kane and Staiger (2008). In order to preserve independence across teachers, I prevent the same student from appearing in the data more than once by running estimation separately for each grade and by dropping students for repeating a grade. The first step in estimation is residualizing outcomes by estimating the β of Equation 1 by regressing within-teacher variation in outcomes on within-teacher variation in covariates:

$$\hat{\beta}^h = \arg\min_b \sum_i \left(y_i^h - \bar{y}_{j(i)}^h - \left(x_i - \bar{x}_{j(i)} \right)^T b \right)^2$$

Key to estimating Σ_μ is that error in Equation 1 are independent across different classrooms, so the average product of residuals in classrooms taught by the same teacher is the covariance of teacher effects. Let $C(j)$ be the set of all classrooms taught by teacher j, and let $\bar{y}_c$ and $\bar{x}_c$ be average outcomes and covariates in classroom c. When errors are independent across classrooms, then

$$\mathbb{E}\left[\left(\bar{y}_c - \bar{x}_c^T \beta \right) \left(\bar{y}_{c'} - \bar{x}_{c'}^T \beta \right)^T \middle| c, c' \in C(j), c \neq c' \right] = \Sigma_\mu.$$

Using a "moment-matching" procedure as in Kane and Staiger (2008) and Chamberlain (2013), $\hat{\Sigma}_\mu$ is the average variance of residualized outcomes in different classrooms taught by the same teacher:

$$\hat{\Sigma}_\mu = \frac{2}{\sum_j |C(j)| \left(|C(j)| - 1 \right)} \sum_j \sum_{c, c' \in C(j): c \neq c'} \left(\bar{y}_c - \bar{x}_c^T \hat{\beta} \right) \left(\bar{y}_{c'} - \bar{x}_{c'}^T \hat{\beta} \right)^T$$

4.1 Inference

I estimate the posterior distribution of Σ_μ using the Bayesian Bootstrap (Rubin, 1981). In the n^{th} Bayesian Bootstrap draw, reweight the data with teacher-level weights $\omega^n \in \mathcal{R}^{\text{N teachers}}$ drawn $\omega^n \sim \text{Dirichlet}\left(1, 1, \ldots, 1\right)$. First, estimate

$$\hat{\beta}^{h,n} = \arg\min_b \sum_i \omega_{j(i)}^n \left(y_i - \bar{y}_{j(i)} - \left(x_i - \bar{x}_{j(i)} \right)^T b \right)^2$$

Then the n^{th} draw of Σ_μ is

$$\hat{\Sigma}_\mu^n = \frac{2}{\sum_j \omega_j^n |C(j)| \left(|C(j)| - 1 \right)} \sum_j \omega_j^n \sum_{c, c' \in C(j): c \neq c'} \left(\bar{y}_c - \bar{x}_c^T \hat{\beta} \right) \left(\bar{y}_{c'} - \bar{x}_{c'}^T \hat{\beta} \right)^T$$

4.2 Identification

In order to understand how teachers contribute to variation in student outcomes, we must ensure that teachers receive credit or blame only for changes in outcomes they *cause*, and not for changes that would have happened with an average teacher. A threat to our ability to identify the variance of teacher value-added is systematic sorting of students to teachers.

This "sorting on observables" restriction, discussed formally in Section 2, requires that random shocks be independent of a teacher's identity, conditional on covariates. I include a rich set of controls to make the sorting on observables restriction plausible. I estimate the model separately for each grade level, so coefficients can change from year to year to allow for, for instance, the persistence of test scores to vary with grade. I control for age, gender, limited English proficiency status, free or reduced-price lunch eligibility, twice-lagged absences, twice-lagged test scores, disability status, teacher-grade level averages of all of these variables, indicators for ethnicity, indicators for year, and indicators for the ten most common home languages. Rather than drop students with missing data, I set missing variables to zero and control for indicators for missingness. I do, however, drop students missing lagged attendance or test scores.

My controls are similar to those in Chetty *et al.* (2014a). There are three differences between their controls and mine: They control for suspensions, which I do not observe; they estimate their model on all grades and interact many variables with grade dummies; and they control for cubics in previous test scores and classroom means of previous test scores. To ensure that my results are robust to using their controls, I generate point estimates of the diagonal of Σ_μ while additionally controlling for cubics in previous test scores, previous attendance, and classroom means of previous test scores and previous attendance. These results are in Section 6; controlling for cubics does not substantially change results.

Although the sorting on observables restriction cannot be directly tested, I follow Chetty *et al.* (2014a) in using pre-trend tests to test whether students show unusual improvement or declines in outcomes in the years before being assigned to a higher value-added teacher. Pre-trend coefficients, discussed below, correspond to negative years in Figures 7 and 14, and in Tables 8 and 17. Visually, pre-trends appear small, but they are statistically significant. These results suggest that controlling for more lags could (but might not) mitigate sorting on observables. In Section 6, I estimate the diagonal of Σ_μ while controlling for all available lags in addition to the baseline controls; additional lags to not appear to affect the results at all.

5 Results

I conducted analysis separately for math teachers and English teachers. Results for math and English teachers currently look very similar, because the sample includes many fourth and fifth grade teachers who teach both math and English. Figures for math teachers are in this section and figures for English teachers are in Appendix A.

The square root of the diagonal of Σ_μ is the standard deviation of teacher effects on each outcome. For example, the square root of the diagonal element of Σ_μ corresponding to four-year high school graduation rates is 0.049, indicating that students of a teacher who is one standard deviation above average at improving graduation rates are 4.9 percentage

Figure 6: *The top plot shows the variance of math teachers' effects on English Language Arts scores, in the same year that the student has this teacher and 1, 2, 3, and 4 years after. The second and third plots show the variances of math teachers' effects on math scores and attendance. Error bars plot a 95% credible interval based on 1000 Bayesian Bootstrap iterations.*

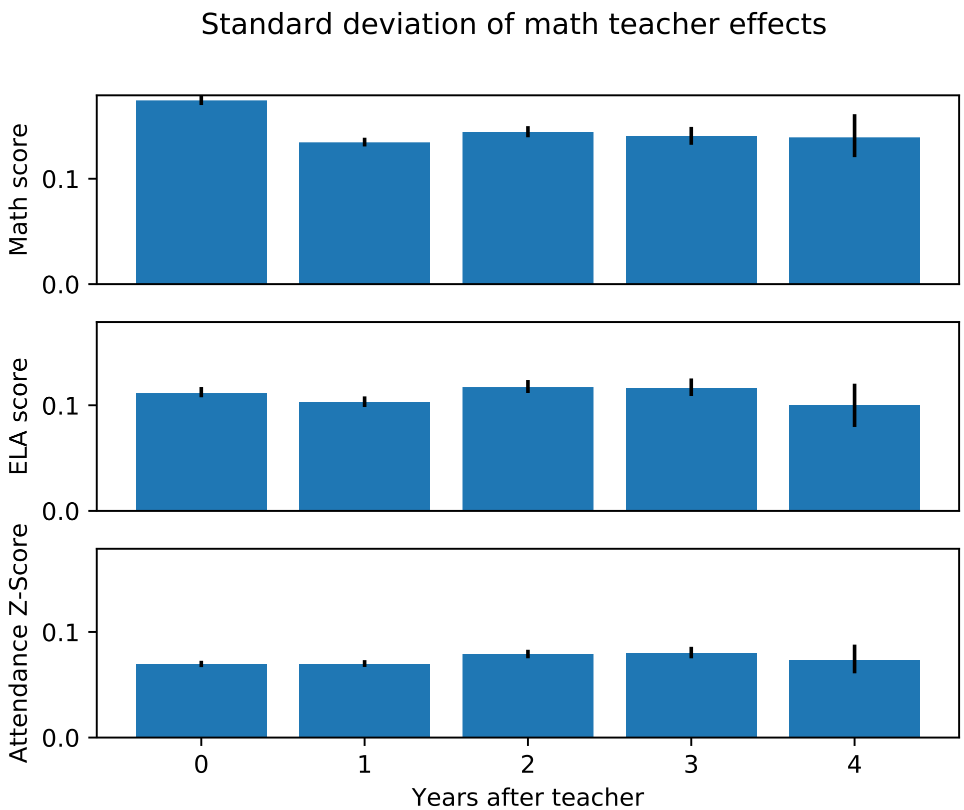


Table 6: *The standard deviation of math teacher effects on test scores and attendance, one to four years out. This table displays the same information as Figure 6. 95% credible interval from 1000 Bayesian Bootstrap iterations in parentheses.*

	Year 0	1	2	3	4
ELA score	0.112	0.103	0.117	0.117	0.100
	(0.108, 0.117)	(0.099, 0.108)	(0.112, 0.124)	(0.109, 0.125)	(0.080, 0.121)
Math score	0.174	0.134	0.144	0.141	0.139
	(0.170, 0.179)	(0.131, 0.139)	(0.139, 0.150)	(0.132, 0.149)	(0.121, 0.161)
Attendance Z-Score	0.069	0.070	0.079	0.080	0.073
	(0.067, 0.073)	(0.067, 0.073)	(0.075, 0.083)	(0.075, 0.086)	(0.061, 0.088)

Table 7: *The standard deviation of math teacher effects on four-year high school graduation. 95% credible interval in parentheses.*

	Standard Deviation of Teacher Effect
Graduated, 4-year	0.049
	(0.047, 0.060)
Regents Diploma, 4-year	0.070
	(0.068, 0.079)
Advanced Regents Diploma, 4-year	0.061
	(0.058, 0.065)

points more likely to graduate high school. The variance of teacher effects on high school graduation is 0.0024 (0.049^2), accounting for 1% of the variance in high school graduation rates. Figure 6 and Table 7 show the effects of math teachers on ELA scores, math scores, and attendance. The standard deviation of math teacher effects on contemporaneous math scores is about 0.174, roughly in line with previous results, indicating that math teachers account for 3% of variance in same-year math test scores (0.174^2). The standard deviation of math teacher effects is slightly smaller in all succeeding years, around 0.14. By contrast, teacher effects on English Language Arts test scores are smaller – Appendix Table 16 shows the analog of 6 for ELA teachers – but are more consistent across years. One explanation for these different patterns in math and ELA scores is that the subject matter on math tests overlaps less across grades than the subject matter on ELA tests. For example, grades 6 through 8 share the same Common Core ELA standards, which cover topics like "Determine the central ideas or information of a primary or secondary source," while the math topics change every year and cover narrower topics like "rational and irrational numbers."

For attendance, the standard deviation of math teacher effects is about 0.07 in each year. Attendance, like test scores, has been Z-scored within each grade and year, so this implies that having a teacher who is one standard deviation above average at increasing attendance increases students' test scores by 0.07 of a standard deviation relative to their peers.

My findings on test scores are consistent with previous literature and but do not resolve the "fade-out mystery" of teacher effects on test scores: if students of high score-VA teachers have only slightly improved test scores four years later, how is it that teachers have substantial effects on later student outcomes? I find that teachers who improve test scores are *not* much better than average at improving attendance, so the fade-out puzzle cannot be resolved by high score-VA teachers improving attendance.

We can also ask, given a teacher's effect on outcome h', what is her expected effect on outcome h? Other work has addressed this question by regressing outcomes on estimated teacher effects and controls, but it can also be answered using only Σ_μ, if we update Equation 2 to assume homoskedasticity:

$$\mathbb{E}\left[\mu_{j(i)}\mu_{j(i)}^T | x_i\right] = \Sigma_\mu. \tag{3}$$

Figure 7: *The best linear predictor coefficient of a teacher's effect on a future outcome given her effect on a present outcome. Error bars plot a 95% credible interval based on 1000 Bayesian Bootstrap iterations.*

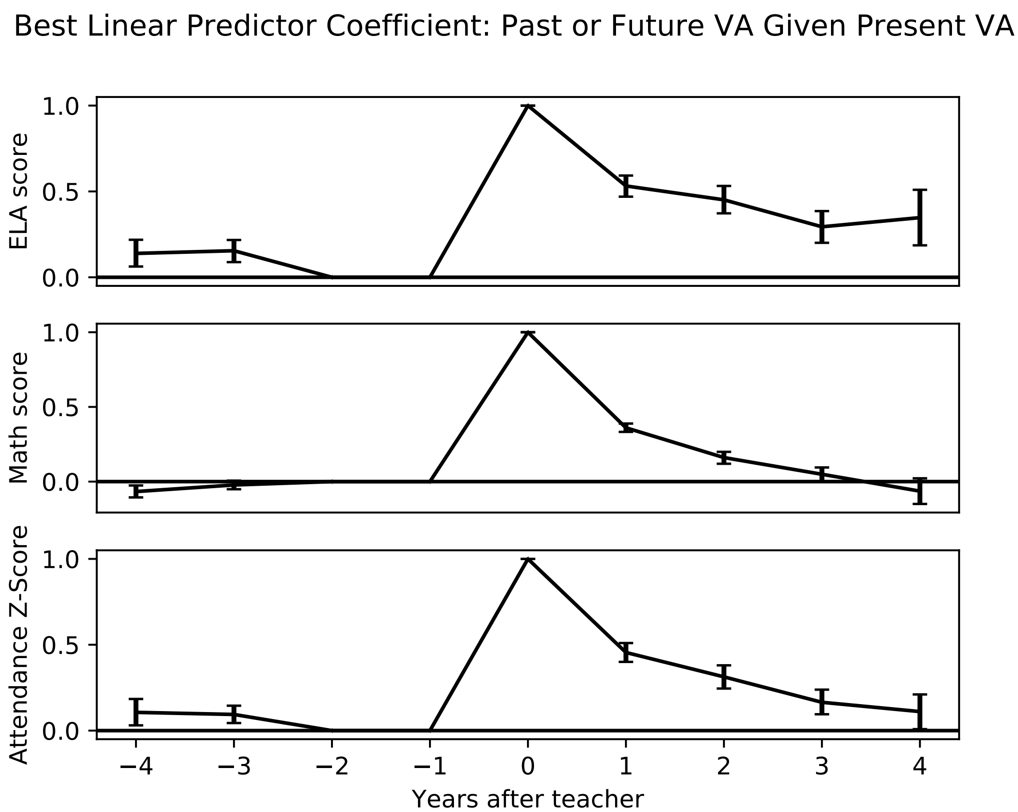


Table 8: *Best linear predictor coefficients. 95% credible interval based on 1000 Bayesian Bootstrap iterations in parentheses.*

	Year				
	-4	-3	-2	-1	0
ELA score	0.139	0.155	0.000	0.000	1.000
	(0.063, 0.219)	(0.088, 0.217)	(0.000, 0.000)	(0.000, 0.000)	(1.000, 1.000)
Math score	-0.067	-0.023	0.000	0.000	1.000
	(-0.106, -0.026)	(-0.051, 0.006)	(0.000, 0.000)	(0.000, 0.000)	(1.000, 1.000)
Attendance Z-Score	0.106	0.094	0.000	0.000	1.000
	(0.031, 0.184)	(0.044, 0.145)	(0.000, 0.000)	(0.000, 0.000)	(1.000, 1.000)
	Year				
	1	2	3	4	Graduated (4-year)
ELA score	0.532	0.451	0.294	0.347	0.035
	(0.470, 0.593)	(0.372, 0.532)	(0.201, 0.386)	(0.186, 0.510)	(-0.060, 0.108)
Math score	0.360	0.161	0.049	-0.064	-0.003
	(0.333, 0.389)	(0.119, 0.199)	(-0.001, 0.095)	(-0.150, 0.022)	(-0.038, 0.024)
Attendance Z-Score	0.455	0.313	0.165	0.111	0.105
	(0.400, 0.510)	(0.245, 0.380)	(0.095, 0.239)	(0.009, 0.211)	(-0.001, 0.205)

Combining Equations 1 and 3,

$$\begin{aligned}
E^*\left[y_i^h | \mu_{j(i)}^{h'}, x_i\right] &= E^*\left[\mu_{j(i)}^H | \mu_{j(i)}^{h'}, x_i\right] + x_i^T \beta^h \\
&= \frac{Cov\left(\mu_j\right)_{h,h'}}{\text{Var}\left(\mu_j^{h'}\right)} \mu_j^{h'} + x_i^T \beta^h \\
&= \frac{\Sigma_\mu^{h,h'}}{\Sigma_\mu^{h',h'}} \mu_j^{h'} + x_i^T \beta^h \\
&\equiv \gamma^{h,h'} \mu_j^{h'} + x_i^T \beta^h
\end{aligned}$$

Figure 7 plots $\gamma^{h,h'}$ where $h =$ contemporaneous English Language Arts scores, math scores, and attendance and h' represents leads and lags of those outcomes (again using only math teachers). Since controls include lagged test scores and attendance, the coefficients on previous-year value-added is zero. Figure 7 and Table 8 show that math teachers who improve English Language Arts scores by x are expected to improve their students' scores in the next year by than half of x, and their scores in the year after by very little. This is a similar result to that found in Chetty *et al.* (2014a). Combined with Figure 6, we see that although teachers do vary in their effects on their students' test scores four years in the future, very little of that variation is captured by teacher effects on same-year test scores. That is, there are teachers who are much better or worse than average at boosting long-term test scores, but they not especially likely to be the teachers who raise short-term test scores. By contrast, there is much less fade-out for effects on attendance. In the Appendix, Figure

14 and Table 17 repeat the same figure and table for English teachers.

5.1 Are current teacher effects a good proxy for future teacher effects?

How well can teacher effects on a long-term outcome be captured by teacher effects on short-term outcomes? We can answer this question by constructing an R^2-like statistic, "goodness of proxy," that reflects how much of the variation in some long-term teacher effect h is captured by short-term teacher effects on outcomes in set Q. If we assume that μ_j has a multivariate normal distribution, then the expectation of one component of μ_j given other components is linear and can be expressed as a function of the covariances in Σ_μ. Say we are interested in predicting value-added at component h, μ_j^h, and know a vector-valued value-added $\mu_j^{\mathbf{Q}}$ for outcomes in set Q. Then the expectation of μ_j^h given $\mu_j^{\mathbf{Q}}$ is

$$\begin{aligned}\mathbb{E}\left[\mu_j^h|\mu_j^{\mathbf{Q}}\right] &= E^*\left[\mu_j^h|\mu_j^{\mathbf{Q}}\right] \\ &= \left(\mu_j^{\mathbf{Q}}\right)^T \operatorname{Var}\left(\mu^{\mathbf{Q}}\right)^{-1} \operatorname{Cov}\left(\mu^{\mathbf{Q}}, \mu^p\right)\end{aligned}$$

Equation 4 defines a "goodness of proxy" statistic that is zero when μ_j^Q does not help predict μ_j^p and equals one when it is perfectly predictable. This measure is also computable using only the components of Σ_μ.

$$\begin{aligned}\text{goodness of proxy}_{h,Q} &\equiv 1 - \frac{\mathbb{E}\left[\operatorname{Var}\left(\mu^h|\mu^{\mathbf{Q}}\right)\right]}{\operatorname{Var}\left(\mu^h\right)} \\ &= \frac{\operatorname{Var}\left(E\left[\mu^h|\mu^{\mathbf{Q}}\right]\right)}{\operatorname{Var}\left(\mu^h\right)} \\ &= \frac{\operatorname{Cov}\left(\mu^{\mathbf{Q}}, \mu^p\right)^T \operatorname{Var}\left(\mu^{\mathbf{Q}}\right)^{-1} \operatorname{Cov}\left(\mu^{\mathbf{Q}}, \mu^p\right)}{\operatorname{Var}\left(\mu^h\right)} \end{aligned} \tag{4}$$

Table 9 and Appendix Table 18 show the goodness of proxy for outcomes graduation and four-year-lead test scores and attendance, using same-year test scores and attendance as a proxy. These tables show that teacher effects on present test scores do not predict teacher effects on current test scores well, but that attendance effects are much more predictable. This is consistent with Tables 7 and 16, which show that teachers vary substantially in their effects on test scores four years in the future, and with Tables 8 and 17, which show that teachers who improve test scores do not have persistent effects, while teachers who improve attendance do.

5.2 Bivariate Correlations and Principal Components Analysis

Can Σ_μ be represented well by several factors? What are they? I answer this question by translating Σ_μ from a covariance matrix to a correlation matrix $\tilde{\Sigma}_\mu$, and applying principal components analysis. I use only the component of the matrix that corresponds to non-negative years, since these are the years in which we expect teachers to have real effects.

Table 9: *Goodness of proxy for graduation and four-year-lead test scores and attendance, using same-year test scores and attendance. 95% credible set based on 1000 Bayesian Bootstrap iterations in parentheses.*

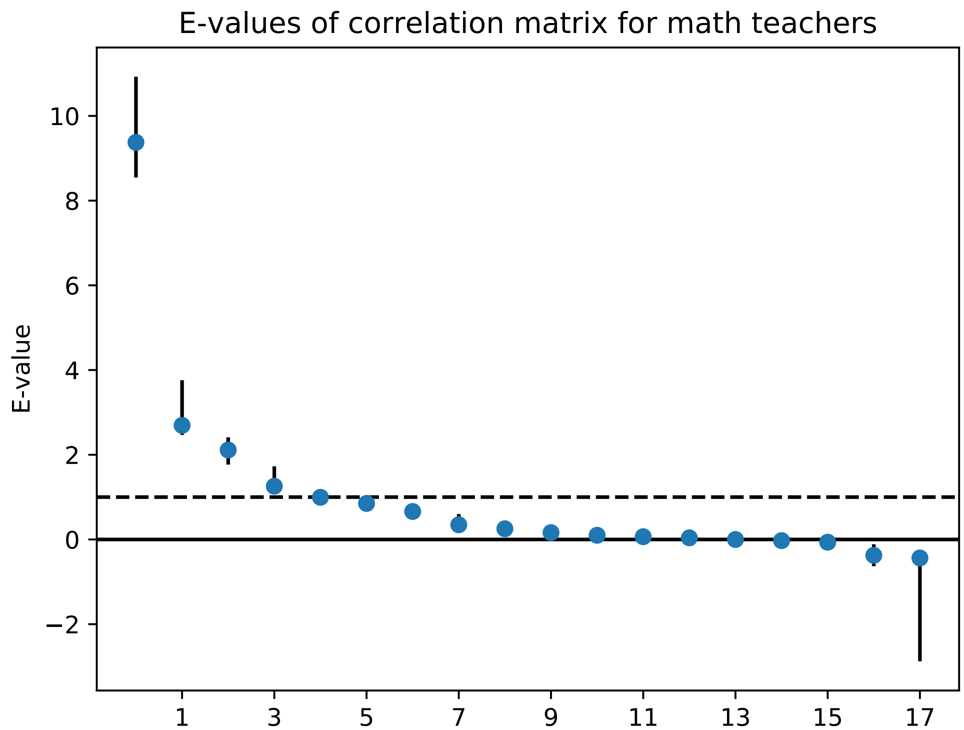
	Goodness of Proxy
ELA score (4 years later)	0.294
	(0.140, 0.525)
Math score (4 years later)	0.043
	(0.013, 0.125)
Attendance Z-Score (4 years later)	0.037
	(0.010, 0.093)
Graduated, 4-year	0.033
	(0.006, 0.118)
Regents Diploma, 4-year	0.015
	(0.002, 0.077)
Advanced Regents Diploma, 4-year	0.062
	(0.031, 0.101)

Table 10: *Correlations between teachers' value-added on different outcomes.*

	ELA score	Math score	Attendance	4-year Grad	4-yr Reg. Dip.
ELA score	1.00				
	(1.00, 1.00)				
Math score	0.66	1.00			
	(0.63, 0.68)	(1.00, 1.00)			
Attendance	0.15	0.12	1.00		
	(0.10, 0.20)	(0.08, 0.17)	(1.00, 1.00)		
4-year Grad	0.08	-0.01	0.15	1.00	
	(-0.12, 0.22)	(-0.12, 0.08)	(-0.00, 0.27)	(1.00, 1.00)	
4-yr Reg. Dip.	-0.11	-0.05	0.01	0.52	1.00
	(-0.27, 0.00)	(-0.15, 0.02)	(-0.09, 0.11)	(0.46, 0.63)	(1.00, 1.00)
4-yr Adv. Reg. Dip.	0.22	0.07	0.08	0.15	-0.74
	(0.15, 0.28)	(0.02, 0.12)	(0.01, 0.14)	(0.04, 0.21)	(-0.77, -0.66)

Before applying principal components analysis, we can gain some intuition for what the factors might look like from Table 10, which shows the correlation structure of teacher effects on different present-year outcomes. Although teacher effects on math and reading test scores are moderately highly correlated, with a correlation of 0.22, teacher effects on test scores and on attendance are much lower.

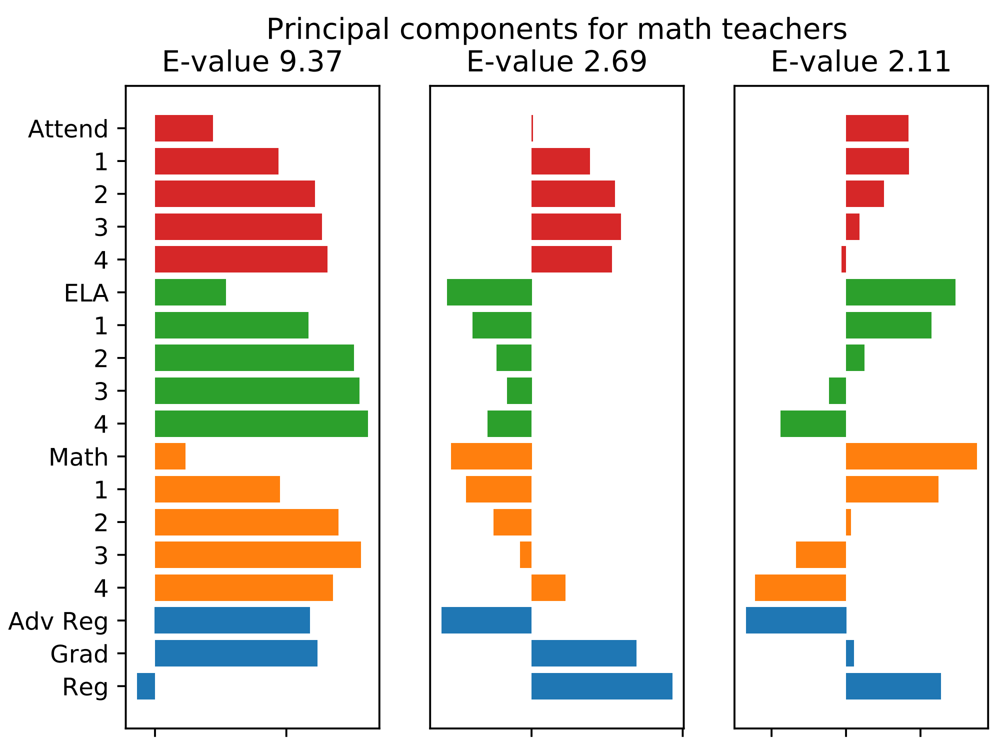
Figure 8: *Eigenvalues of $\tilde{\Sigma}_\mu$ for math teachers. Error bars show a 95% credible interval based on 1000 Bayesian Bootstrap iterations.*



There are several rules of thumb for choosing the number of components. One is to count the number of eigenvalues that are greater than one. This heuristic suggests two components for both math and English teachers. Another heuristic suggests plotting the eigenvalues, as in Figures 8 and 15, and look for a point where the eigenvalues start to level off. This heuristic suggests one to three components for math teachers and three for ELA teachers.

Figure 9 shows the first three principal components for math teachers. The first principal component is by far the most important. It is associated with positive effects on all but one outcome. (The one opposite-signed element, graduationg with a Regents Diploma, is not necessarily a good outcome, because it precludes graduating with an Advanced Regents Diploma.) The second principal component reflects positive effects on graduation with a less-rigorous Regents diploma and on attendance, but negative effects on test scores. The third component is associated with having opposite-signed effects on short-term and

Figure 9: *First three principal components of $\tilde{\Sigma}_{\mu}$ for math teachers.*



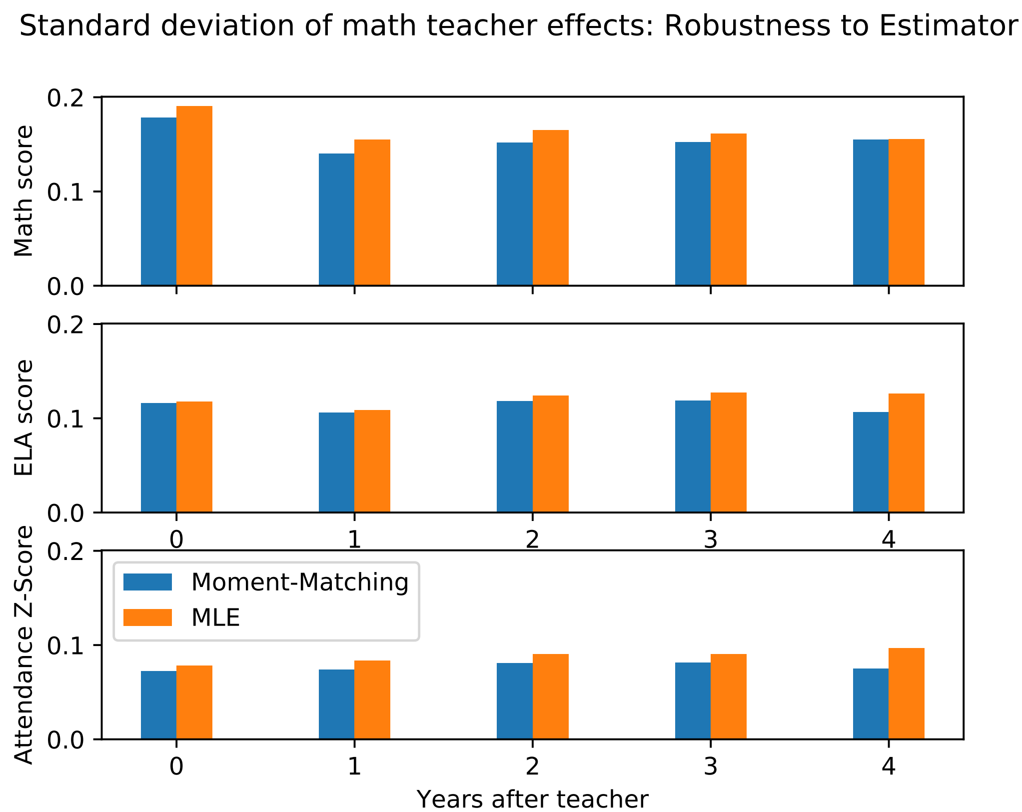
long-term outcomes.

6 Robustness Checks

Various methods are available to estimate the diagonal of Σ_μ, which reveals the variance of teacher effects on each outcome, as surveyed in Chapter 2 of this dissertation. Of course, infinitely many sets of control variables are possible. This section estimates the square root of the diagonal of Σ_μ for two different estimators and three sets of controls for math teachers; results from English teachers are in the appendix.

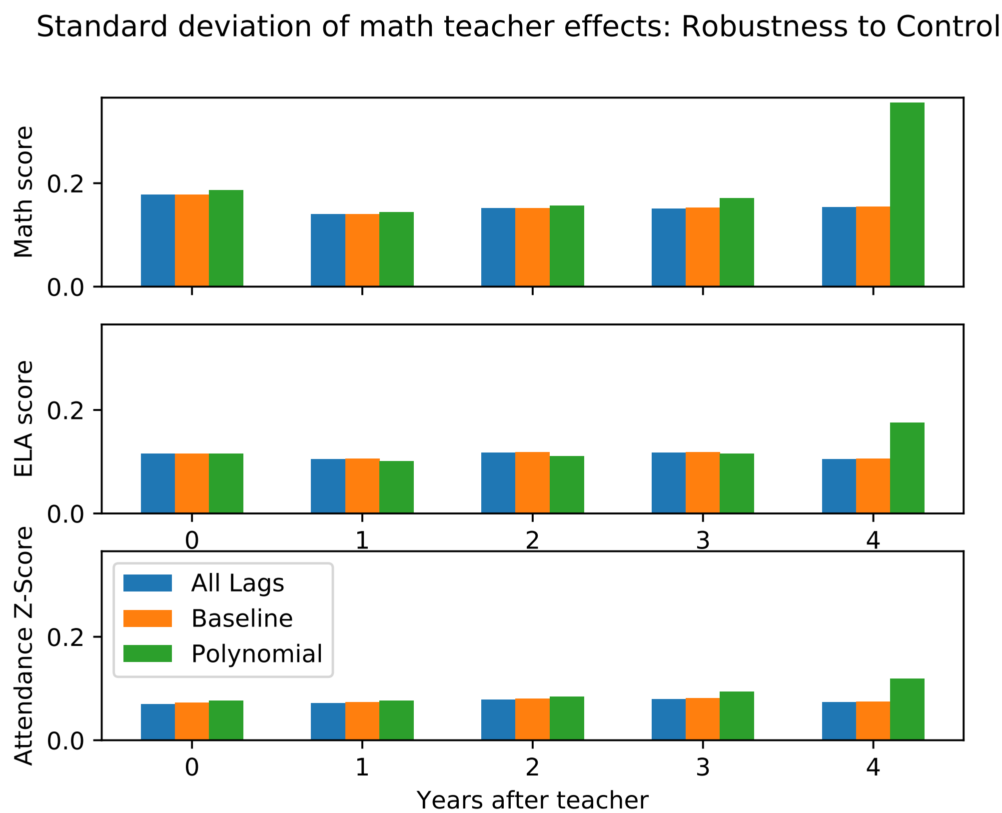
The method used in this paper is a "moment-matching" estimator that residualizes test scores using within-teacher variation and compares mean residuals in different classrooms taught by the same teacher. It is identical to the "modified-KS" estimator in Chapter 2. An alternative estimation method is maximum likelihood. Figure 10 and Appendix Figure 17.

Figure 10: *Standard deviation of math teacher effects on present and future outcomes, for both the "moment-matching" estimator used above and maximum likelihood.*



In the "baseline" controls used to generate the results in Section 5, I control for age, gender, limited English proficiency status, free or reduced-price lunch eligibility, twice-lagged absences, twice-lagged test scores, disability status, teacher-grade level averages of all of these variables, indicators for ethnicity, indicators for year, and indicators for the ten most common home languages. Rather than drop students with missing data, I set missing variables to zero and control for indicators for missingness. I do, however, drop students missing lagged attendance or test scores. Figure 11 and Appendix Figure 18 additionally show results from two other sets of control variables. The "Polynomial" controls control for all baseline variables, cubics in lagged test scores and attendance, and cubics in teacher-year level means of lagged test scores and attendance. The "Polynomial" controls are very similar to the controls used in Chetty *et al.* (2014a). The "All Lags" controls control for up to four years of lags, with missing lags imputed to zero and an indicator for missingness.

Figure 11: *Standard deviation of math teacher effects on present and future outcomes, for the baseline controls used above, baseline controls plus all available lags, and for the controls from Chetty et al. (2014a), which include third-degree polynomials.*



Tables 11, 12, 13, and 14 (and Appendix Tables 19, 20, 21, and 22) display all of the results found in the figures, as well as results from maximum likelihood estimation with the non-baseline sets of controls.

Table 11: *Standard deviation of math teacher effects on ELA scores: Robustness to choice of controls and estimator.*

Controls	Estimator	Year 0	1	2	3	4
Baseline	Moment-Matching	0.116	0.106	0.118	0.119	0.106
Baseline	MLE	0.118	0.109	0.124	0.127	0.126
Polynomial	Moment-Matching	0.116	0.101	0.111	0.116	0.175
Polynomial	MLE	0.120	0.113	0.123	0.137	0.197
All Lags	Moment-Matching	0.115	0.105	0.117	0.118	0.105
All Lags	MLE	0.130	0.110	0.124	0.128	0.126

Table 12: *Standard deviation of math teacher effects on math scores: Robustness to choice of controls and estimator.*

Controls	Estimator	Year 0	1	2	3	4
Baseline	Moment-Matching	0.178	0.140	0.152	0.152	0.155
Baseline	MLE	0.191	0.155	0.165	0.162	0.156
Polynomial	Moment-Matching	0.186	0.144	0.157	0.171	0.355
Polynomial	MLE	0.200	0.163	0.177	0.180	0.390
All Lags	Moment-Matching	0.178	0.140	0.151	0.151	0.154
All Lags	MLE	0.194	0.155	0.165	0.162	0.159

Table 13: *Standard deviation of math teacher effects on attendance: Robustness to choice of controls and estimator.*

Controls	Estimator	Year 0	1	2	3	4
Baseline	Moment-Matching	0.072	0.074	0.081	0.081	0.075
Baseline	MLE	0.078	0.084	0.090	0.091	0.097
Polynomial	Moment-Matching	0.077	0.077	0.085	0.094	0.120
Polynomial	MLE	0.087	0.086	0.092	0.098	0.127
All Lags	Moment-Matching	0.070	0.072	0.079	0.080	0.074
All Lags	MLE	0.083	0.082	0.089	0.090	0.101

Table 14: *Standard deviation of math teacher effects on graduation: Robustness to choice of controls and estimator.*

Controls	Estimator	Graduated	Regents Diploma	Advanced Regents Diploma
Baseline	Moment-Matching	0.049	0.074	0.065
Baseline	MLE	0.051	0.076	0.065
Polynomial	Moment-Matching	0.059	0.072	0.076
Polynomial	MLE	0.059	0.072	0.062
All Lags	Moment-Matching	0.047	0.073	0.065
All Lags	MLE	0.052	0.078	0.065

7 Conclusion

To-dos:

- Different framing of this section: There are two possible policy changes, evaluating teachers on more outcomes than test scores, and evaluating teachers on future outcomes. This section is left over from an older version.

Absenteeism is pervasive in urban school districts, and while teachers explain only a small portion of the variance in student attendance, they can be effective in reducing absenteeism. The teachers who reduce absences continue to affect their students for at least four years into the future. Holding teacher behavior constant, incorporating attendance into existing value-added measures would make these measures more fair and informative, since teachers who are effective in improving test scores are not especially likely to improve attendance. Since school districts routinely collect attendance data and since many districts have adopted quantitative value-added measures, it would be very feasible for these districts to incorporate attendance and other outcomes into value-added scoring.

However, there may be risks to compensating, evaluating, or firing teachers based on attendance, analogous to the risks in incentivizing high test scores. Although there is a wide literature on incentivizing *schools* for higher test scores or meeting proficiency standards, the effects of quantitative incentives for *teachers* are far from clear. For example, Fryer (2013) discusses a randomized trial in which New York City public schools were eligible for more funding if they reached test score targets. Test scores did not improve in treatment schools, but this may have been because most schools chose group incentives. On the other hand, Lavy (2002) studied group incentives on several performance measures, including test scores, in Israel and found that they increased both contemporaneous test scores and a variety of outcome measures in the following year. Although incentives for higher test scores are intended to increase teacher effort, teachers often seem to respond in less desirable ways, such as increasing time spent on test prep Glewwe *et al.* (2010). And even in the presence of school-level performance incentives that only weakly affect individual teachers, teachers may “teach to the test" (Klein *et al.* (2000)) or cheat (Jacob and Levitt (2003), Jacob (2005), Loughran and Comiskey (1999)).

Although I show that high attendance value-added teachers have historically improved their students' achievement, this does not imply that teachers should be compensated, evaluated, or fired based on attendance-VA measures. Similarly, this paper has little to say about whether score-VA should be a component of teacher evaluation. Policymakers face a multitasking problem in the spirit of Holmstrom and Milgrom (1991): we want teachers to make their students motivated, persistent, and informed, but we can only design contracts on the basis of observable factors. The risk of perverse incentives that comes with test score-based teacher evaluation measures may make attendance-based value added more appealing as an alternative, but it should also caution us that incentivizing teachers for student behavior may lead to unintended outcomes. For example, teachers could encourage students to come to school even when sick, or make class more fun at the expense of being edifying by, for example, showing movies. In addition, in the long run, teachers may require significantly higher pay to compensate them for the stress and uncertainty that a merit-based pay and retention system could generate.

While I have shown that teachers impact attendance in an environment that does not reward teachers for student attendance, the welfare implications of rewarding teachers for good student attendance are not clear. This is an avenue for further research and could be most clearly studied through an experiment.

Many questions remain unanswered in this area. Using this data, it is possible to explore heterogeneity in teacher effects; for example, do teachers have larger effects on absences for students who are absent more often? It would also be helpful to track these students farther into the future to explore the effects of teachers who reduce absences on high school graduation rates, college attendance and completion rates, and income.

References

Carneiro, P., Crawford, C., Goodman, A. and Centre for the Economics of Education (Great Britain) (2007). *The impact of early cognitive and non-cognitive skills on later outcomes*. London: Centre for the Economics of Education, London School of Economics, oCLC: 183819510.

Chamberlain, G. E. (2013). Predictive effects of teachers and schools on test scores, college attendance, and earnings. *Proceedings of the National Academy of Sciences*, **110** (43), 17176–17182.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W. and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project Star. *Quarterly Journal of Economics*, **126** (4), 1593–660.

—, — and Rockoff, J. E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, **104** (9), 2593–2632.

—, — and — (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, **104** (9), 2633–2679.

Dee, T. S., Dobbie, W., Jacob, B. A. and Rockoff, J. (2016). *The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations*. Tech. rep., National Bureau of Economic Research.

Figlio, D. and Getzler, L. (2002). Accountability, Ability, and Disability: Gaming the System? *NBER Working Paper*.

Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from New York City public schools. *Journal of Labor Economics*, **31** (2), 373–407.

Gershenson, S. (2016). Linking Teacher Quality, Student Attendance, and Student Achievement. *Education Finance and Policy*, **11** (2), 125–149.

Glewwe, P., Ilias, N. and Kremer, M. (2010). Teacher Incentives. *American Economic Journal: Applied Economics*, **2** (3), 205–227.

Heckman, J. J. and Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *The American Economic Review*, **91** (2), 145–149.

Holmstrom, B. and Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization*, **7** (Special Issue: Papers from the Conference on the New Science of Organization), 24–52.

Jackson, C. K. (2016). *What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes*. Tech. rep., National Bureau of Economic Research.

Jacob, B. A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, **89** (5-6), 761–796.

— and Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, **118** (3), 843–877.

Kane, T. and Staiger, D. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. *NBER Working Paper*.

Klein, S., Hamilton, L., McCaffrey, D. and Stecher, B. (2000). What Do Test Scores in Texas Tell Us?

Lavy, V. (2002). Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement. *Journal of Political Economy*, **110** (6), 1286–1317.

Loughran, R. and Comiskey, T. (1999). *Cheating the Children: Educator Misconduct on Standardized Tests*. Tech. rep., City of New York: The Special Commissioner of Investigation for the New York City School District.

Rubin, D. B. (1981). The Bayesian Bootstrap. *The Annals of Statistics*, **9** (1), 130–134.

A Appendix: Figures and Tables for ELA Teachers

Figure 12: *Histogram of student absence frequency.*

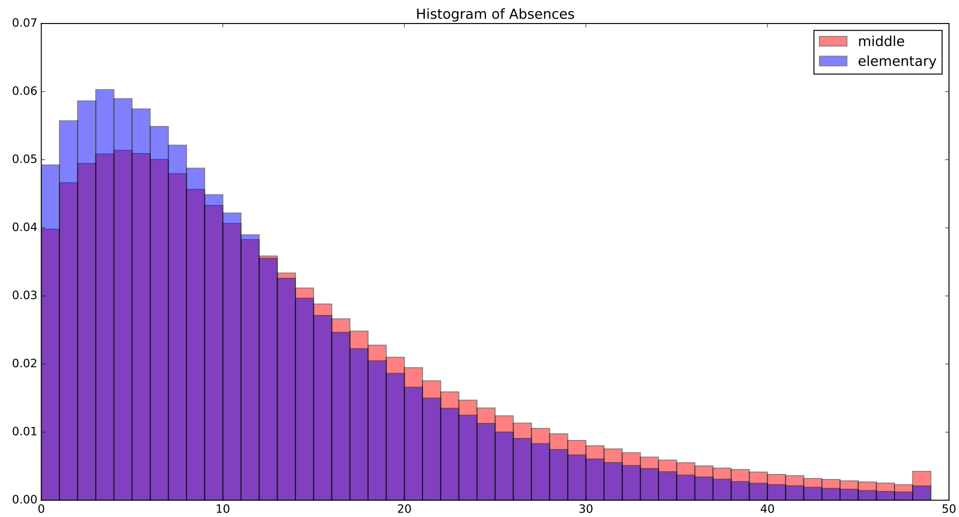


Table 15: *The standard deviation of English teacher effects on test scores and attendance, one to four years out. This table displays the same information as 13, but in standard deviation units. 95% credible interval from 1000 Bayesian Bootstrap iterations in parentheses.*

	Year				
	0	1	2	3	4
ELA score	0.116	0.104	0.119	0.117	0.099
	(0.112, 0.121)	(0.099, 0.110)	(0.113, 0.125)	(0.110, 0.125)	(0.080, 0.121)
Math score	0.165	0.135	0.144	0.140	0.139
	(0.160, 0.170)	(0.131, 0.139)	(0.140, 0.150)	(0.133, 0.149)	(0.119, 0.163)
Attendance Z-Score	0.071	0.072	0.078	0.079	0.080
	(0.069, 0.075)	(0.069, 0.075)	(0.075, 0.083)	(0.075, 0.085)	(0.068, 0.095)

Table 16: *The standard deviation of English teacher effects on four-year high school graduation. This table displays the same information as 13, but in standard deviation units. 95% credible interval in parentheses.*

	Standard Deviation of Teacher Effect
Graduated, 4-year	0.048
	(0.046, 0.060)
Regents Diploma, 4-year	0.068
	(0.066, 0.077)
Advanced Regents Diploma, 4-year	0.061
	(0.059, 0.066)

Figure 13: *The top plot shows the variance of English teachers' effects on English Language Arts scores, in the same year that the student has this teacher and 1, 2, 3, and 4 years after. The second and third plots show the variances of English teachers' effects on math scores and attendance. Error bars plot a 95% credible interval based on 1000 Bayesian Bootstrap iterations.*

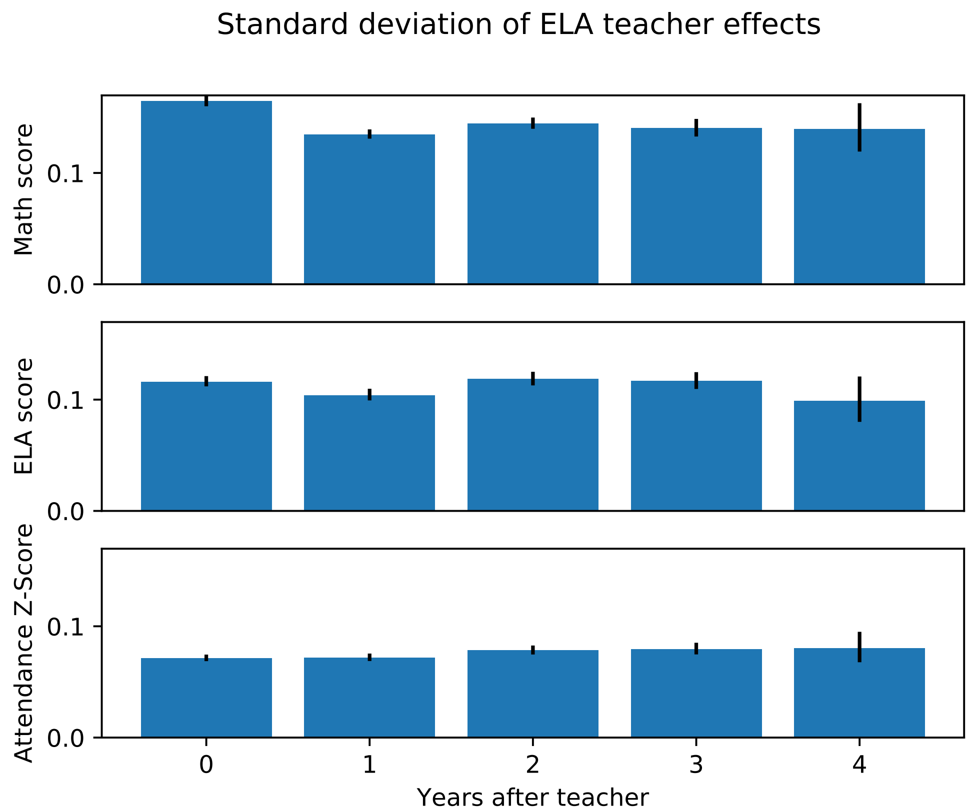


Figure 14: *The best linear predictor coefficient of a teacher's effect on a future outcome given her effect on a present outcome. Error bars plot a 95% credible interval based on 1000 Bayesian Bootstrap iterations.*

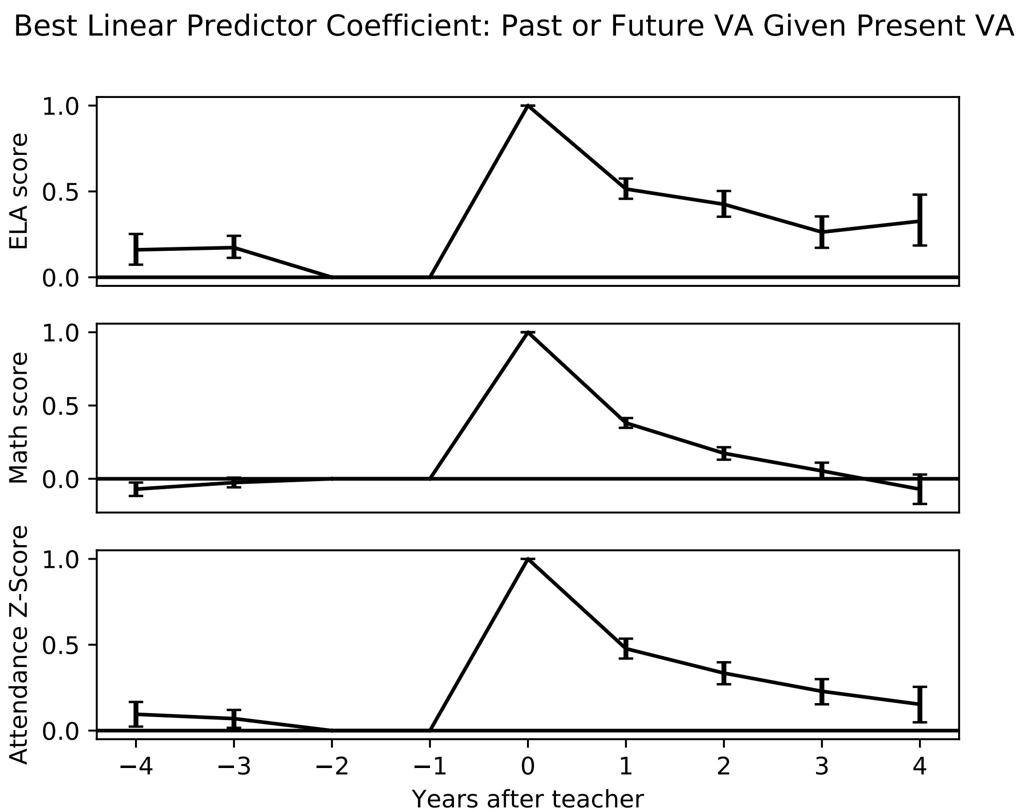


Table 17: *Best linear predictor coefficients. 95% credible interval based on 1000 Bayesian Bootstrap iterations in parentheses.*

	Year				
	-4	-3	-2	-1	0
ELA score	0.160	0.173	0.000	0.000	1.000
	(0.073, 0.252)	(0.113, 0.242)	(0.000, 0.000)	(0.000, 0.000)	(1.000, 1.000)
Math score	-0.071	-0.026	0.000	0.000	1.000
	(-0.117, -0.026)	(-0.058, 0.008)	(0.000, 0.000)	(0.000, 0.000)	(1.000, 1.000)
Attendance Z-Score	0.095	0.070	0.000	0.000	1.000
	(0.023, 0.167)	(0.016, 0.120)	(0.000, 0.000)	(0.000, 0.000)	(1.000, 1.000)
	Year				
	1	2	3	4	Graduated (4-year)
ELA score	0.514	0.425	0.263	0.327	0.028
	(0.457, 0.575)	(0.353, 0.503)	(0.172, 0.355)	(0.185, 0.481)	(-0.057, 0.097)
Math score	0.381	0.174	0.054	-0.071	0.002
	(0.348, 0.414)	(0.130, 0.215)	(0.002, 0.110)	(-0.172, 0.029)	(-0.034, 0.032)
Attendance Z-Score	0.477	0.335	0.229	0.153	0.118
	(0.420, 0.535)	(0.270, 0.398)	(0.153, 0.300)	(0.048, 0.255)	(0.022, 0.214)

Table 18: *Goodness of proxy for graduation and four-year-lead test scores and attendance, using same-year test scores and attendance. 95% credible set based on 1000 Bayesian Bootstrap iterations in parentheses.*

	Goodness of Proxy
ELA score (4 years later)	0.328
	(0.168, 0.593)
Math score (4 years later)	0.053
	(0.018, 0.136)
Attendance Z-Score (4 years later)	0.024
	(0.007, 0.061)
Graduated, 4-year	0.036
	(0.007, 0.111)
Regents Diploma, 4-year	0.025
	(0.005, 0.095)
Advanced Regents Diploma, 4-year	0.070
	(0.039, 0.112)

Figure 15: *Eigenvalues of $\tilde{\Sigma}_\mu$ for English teachers. Error bars show a 95% credible interval based on 1000 Bayesian Bootstrap iterations.*

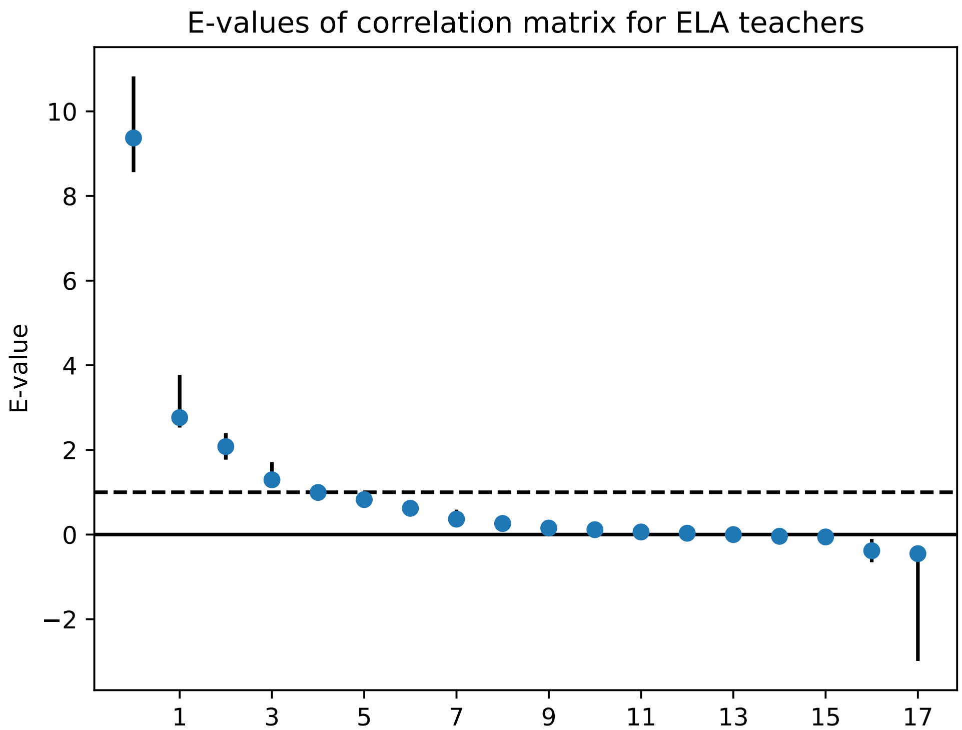


Table 19: *Standard deviation of ELA teacher effects on ELA scores: Robustness to choice of controls and estimator.*

		Year				
Controls	Estimator	0	1	2	3	4
Baseline	Moment-Matching	0.117	0.106	0.118	0.119	0.106
Baseline	MLE	0.118	0.109	0.124	0.127	0.126
Polynomial	Moment-Matching	0.117	0.101	0.111	0.116	0.175
Polynomial	MLE	0.121	0.113	0.123	0.137	0.197
All Lags	Moment-Matching	0.116	0.105	0.117	0.118	0.105
All Lags	MLE	0.130	0.110	0.124	0.128	0.126

Figure 16: *First three principal components of $\tilde{\Sigma}_\mu$ for ELA teachers.*

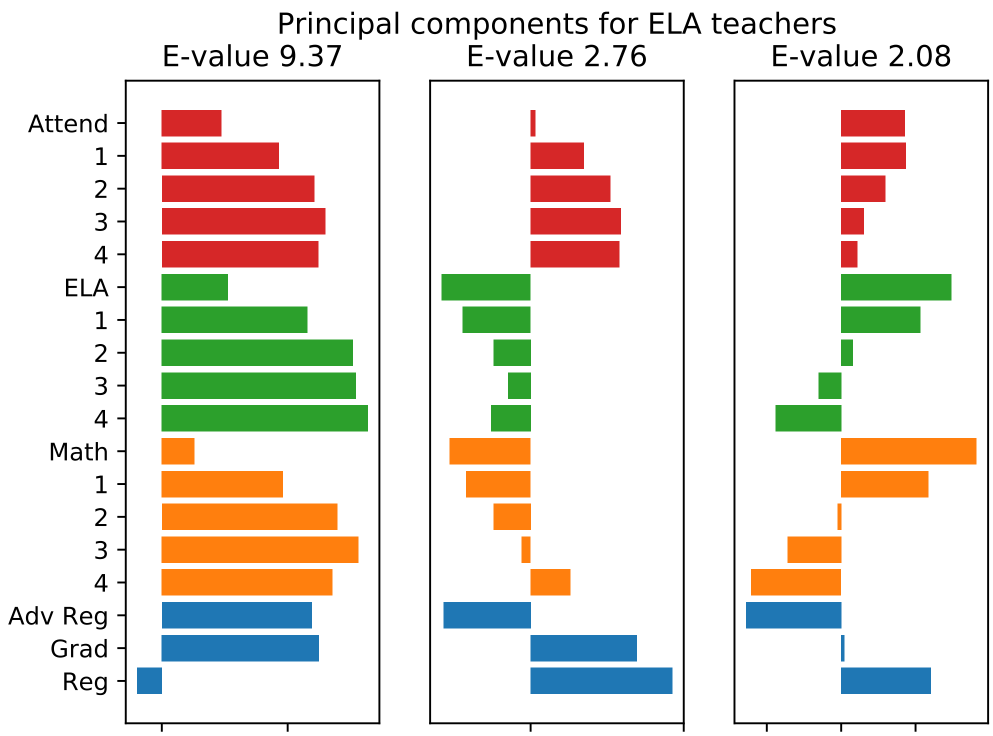


Table 20: *Standard deviation of English teacher effects on math scores: Robustness to choice of controls and estimator.*

		Year				
Controls	Estimator	0	1	2	3	4
Baseline	Moment-Matching	0.179	0.140	0.152	0.152	0.155
Baseline	MLE	0.191	0.155	0.165	0.162	0.156
Polynomial	Moment-Matching	0.187	0.144	0.157	0.171	0.355
Polynomial	MLE	0.200	0.163	0.177	0.180	0.390
All Lags	Moment-Matching	0.179	0.140	0.151	0.151	0.154
All Lags	MLE	0.194	0.155	0.165	0.162	0.159

Figure 17: *Standard deviation of English teacher effects on present and future outcomes, for both the "moment-matching" estimator used above and maximum likelihood.*

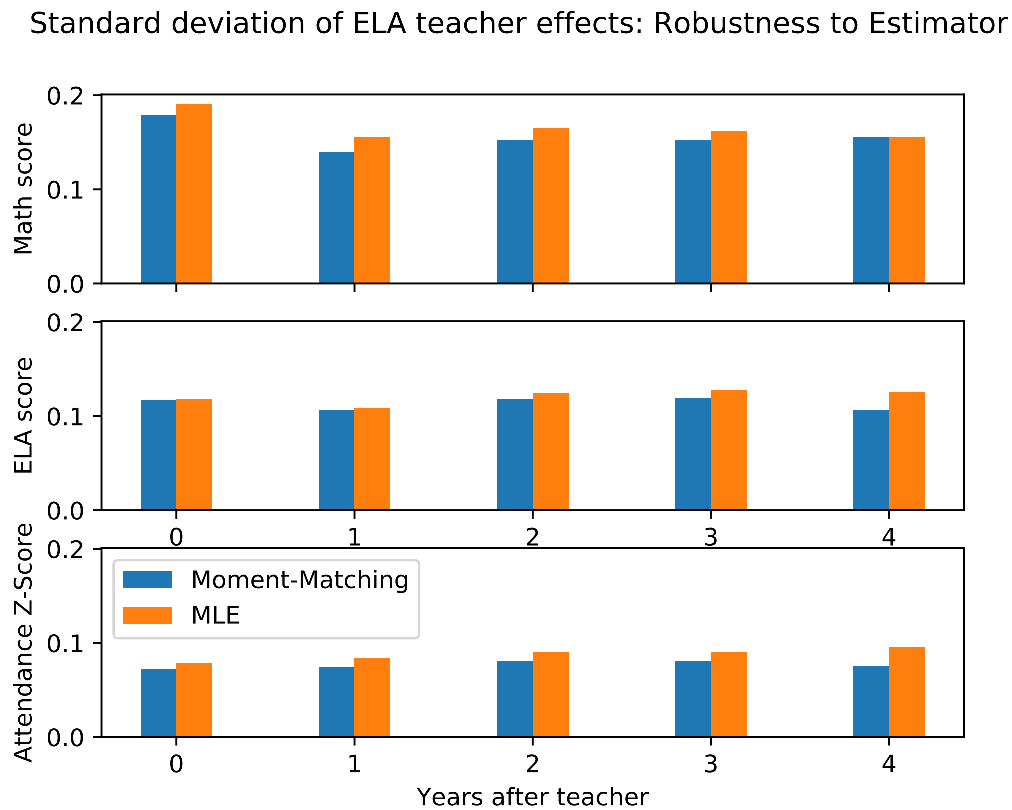


Table 21: *Standard deviation of English teacher effects on attendance: Robustness to choice of controls and estimator.*

Controls	Estimator	Year 0	1	2	3	4
Baseline	Moment-Matching	0.073	0.074	0.081	0.081	0.075
Baseline	MLE	0.078	0.084	0.090	0.090	0.096
Polynomial	Moment-Matching	0.077	0.077	0.085	0.093	0.118
Polynomial	MLE	0.087	0.086	0.092	0.097	0.125
All Lags	Moment-Matching	0.070	0.072	0.079	0.079	0.075
All Lags	MLE	0.083	0.082	0.089	0.090	0.100

Figure 18: *Standard deviation of English teacher effects on present and future outcomes, for the baseline controls used above, baseline controls plus all available lags, and for the controls from Chetty et al. (2014a), which include third-degree polynomials.*

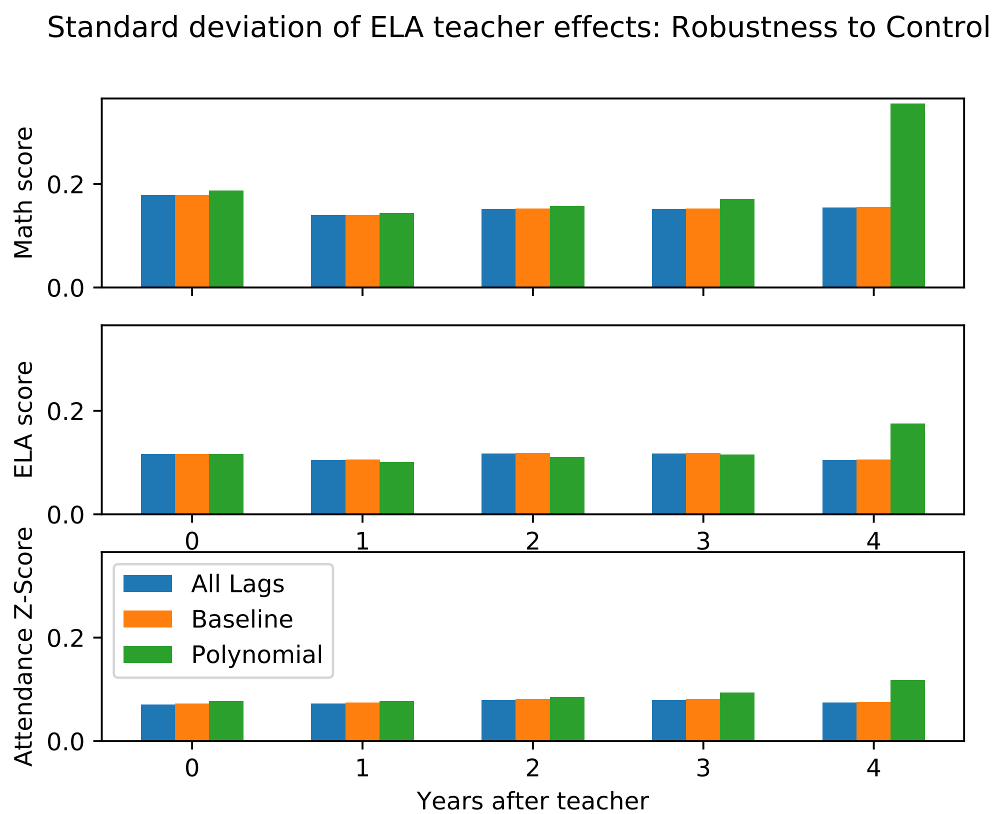


Table 22: *Standard deviation of English teacher effects on graduation: Robustness to choice of controls and estimator.*

Controls	Estimator	Graduated	Regents Diploma	Advanced Regents Diploma
Baseline	Moment-Matching	0.049	0.074	0.065
Baseline	MLE	0.051	0.076	0.065
Polynomial	Moment-Matching	0.059	0.073	0.077
Polynomial	MLE	0.059	0.072	0.062
All Lags	Moment-Matching	0.047	0.073	0.065
All Lags	MLE	0.052	0.078	0.065