Teacher Effects on Student Attendance (Preliminary and Incomplete)

Elizabeth Santorella *

August 28, 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.

Although teachers have substantial effects on test scores four years in the future, these effects are not highly correlated with teacher effects on same-year test scores, implying that the effects of having a teacher who raises same-year test scores fades out quickly, and short-term teacher effects on test scores do not predict long-term effects well. Teacher effects on same-year attendance, by contrast, are highly predictive of longer-term attendance. Teacher effects on test scores are only weakly correlated with teacher effects on attendance. Teachers who are one standard deviation above average at improving high school graduation increase graduation rates by 2% to 8%, although graduation data is limited.

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 teacher effects are in different domains, such as math test scores and attendance; how

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

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 several days per year. These effects fade out slowly; by contrast, students of a teacher who improves test scores have barely improved scores four years later. However, teacher quality is an important determinant of future test scores; 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 very weak, 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 2% to 8%, but I have less power to study graduation.

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 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.

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 if identification restrictions are satisfied; evidence suggests that they are. 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; I confirm that, conditional on controls, value-added does not predict past attendance or test scores, but does predict present and future attendance and test scores.

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, and their influence on future student achievement. Section 6 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 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. Observations are at the student-classroom level, so I index all variables by student i and classroom c, and use $j(i,c)$ to refer to the teacher of student i in classroom c. Subscripts index observations, and superscripts index outcomes. There are $H = 16$ outcomes: Math test scores, reading test scores, attendance z-scores, for years of leads for each of those variables, and a high school graduation indicator. Outcomes are $y_{ic} \in \mathbb{R}^H$ and $x_{ic} \in \mathbb{R}^{H \times K}$. Teacher j's effect on outcomes, $\mu_j \in \mathbb{R}^H$, is defined as a best linear predictor coefficient. For each outcome h,

$$E^* \left[y_{ic}^h \middle| x_{ic}, \mu_{j(i,c)}^h \right] = \alpha^h + x_{ic}^T \beta^h + \mu_{j(i,c)}^h \tag{1}$$

We can also define errors ν_{ic}:

$$y_{ic}^h \equiv \alpha^h + x_{ic}^T \beta^h + \mu_{j(i,c)}^h + \nu_{ic}$$

In order to ascribe a casual interpretation to parameter estimates, we need sorting on observables. This means, first, that variation in teacher effects that cannot be captured in covariates must be orthogonal to ν. This restriction would be violated if, say, better teachers tend to teacher students with higher previous test scores, students with higher past test scores tend to have higher future test scores, and past test scores are not included in 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 — $\mu_j = 0 \forall j$ — 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: $\mathbb{E}\left[\nu_{ic}\nu_{kc'}\right] = 0$ when $c \neq c'$. The parameter of interest is the covariance matrix of teacher effects,

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

3 Data, Setting, and Descriptive Statistics

The data includes almost all New York City public school students in grades 1 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, dropout status, number of credits attempted and passed, and what type of diploma the student received. I am currently only using test scores, attendance, and a binary indicator for college graduation as outcomes but would like to expand this analysis to include regents exam scores, number of classes passed, and whether the student receives a local, Regents, or advanced Regents diploma.

While my data includes N student-year observations, 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. Students who pass additional 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 74% 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 14% 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.

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

	Mean	St. Dev	Min	Max	Missing
Grade	5.88	1.43	4	8	0%
Year	2009.31	2.26	2006	2013	0%
Disabled	0.16	0.37	0	1	0%
Female	0.51	0.50	0	1	0%
English Language Learner	0.11	0.31	0	1	0%
Free Lunch	0.75	0.43	0	1	0%
Days absent	10.61	11.59	0	186	0.02%
Days present	169.96	14.63	0	195	0.63%
Days Absent Lag (Z-Score)	-0.13	0.83	-1.16	17.07	14.39%
Math Score (Z-Score)	0.20	0.92	-6.35	3.89	0.67%
Math score lag (Z-Score)	0.20	0.90	-6.35	3.89	31.32%
ELA Score (Z-Score)	0.19	0.91	-10.48	7.76	1.67%
ELA Score Lag (Z-Score)	0.19	0.92	-11.1	7.76	30.83%
ELA Class Size	13.80	8.54	1	40	0%
Math Class Size	13.80	8.52	1	40	0%
4-Year Graduation	0.69	0.46	0	1	87.52%
August 4-Year Graduation	0.74	0.44	0	1	87.52%
5-Year Graduation	0.78	0.41	0	1	93.76%

Table 2: *Teacher Summary Statistics*

Mean	St. Dev	Min	Max	% Missing

Teachers also have a unique anonymized identifier that is consistent across years. I observe each teacher's sex, ethnicity, birth date, salary, and years of experience in the district. Table 2 summarizes teachers' observable characteristics.

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 X days in an approximately 180- to 185-day school year, or X%. By comparison, the average student nationally is only 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

Figure 1: *Empirical CDF of absences in grades 1-4 and grades 4-8.*

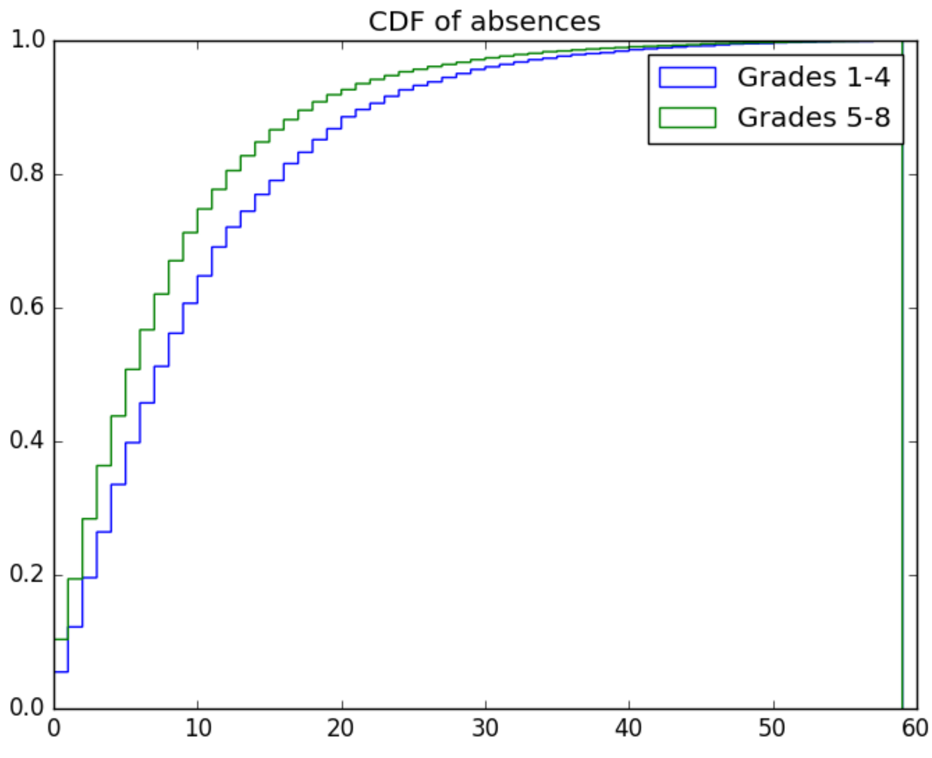
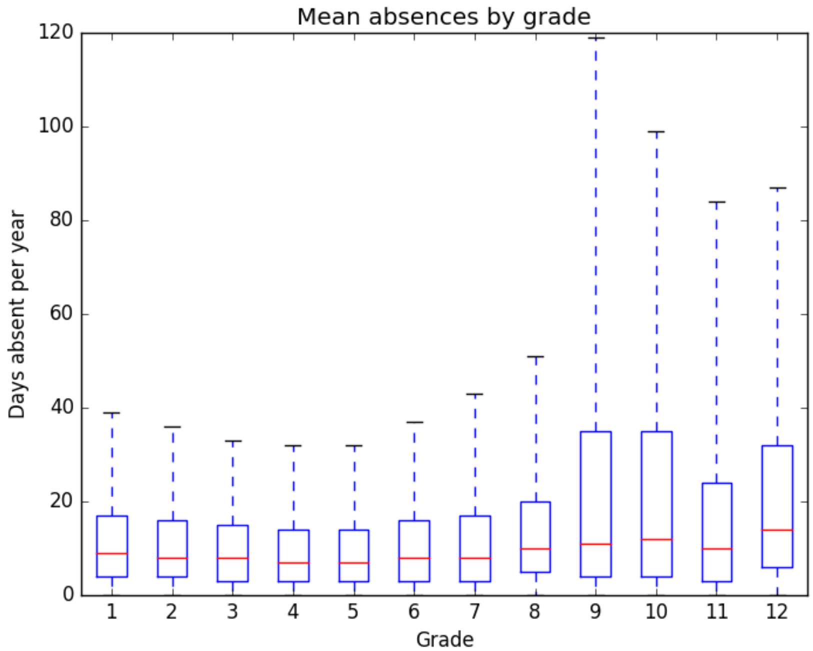


Figure 2: *Red lines show the median, box edges show the 25th and 75th percentiles, and whiskers show the 5th and 95th percentiles.*



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

Tables 3, 4, and 5 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, 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 are similar in both regressions with test scores as the dependent variable, they are not in Table 5, with attendance as the dependent variable. For example, special education students are significantly *less* likely to be absent, and 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 0.316 in a regression that includes lagged values of attendance and test scores.

Figures 3 and 4 show the association between attendance and test scores by plotting

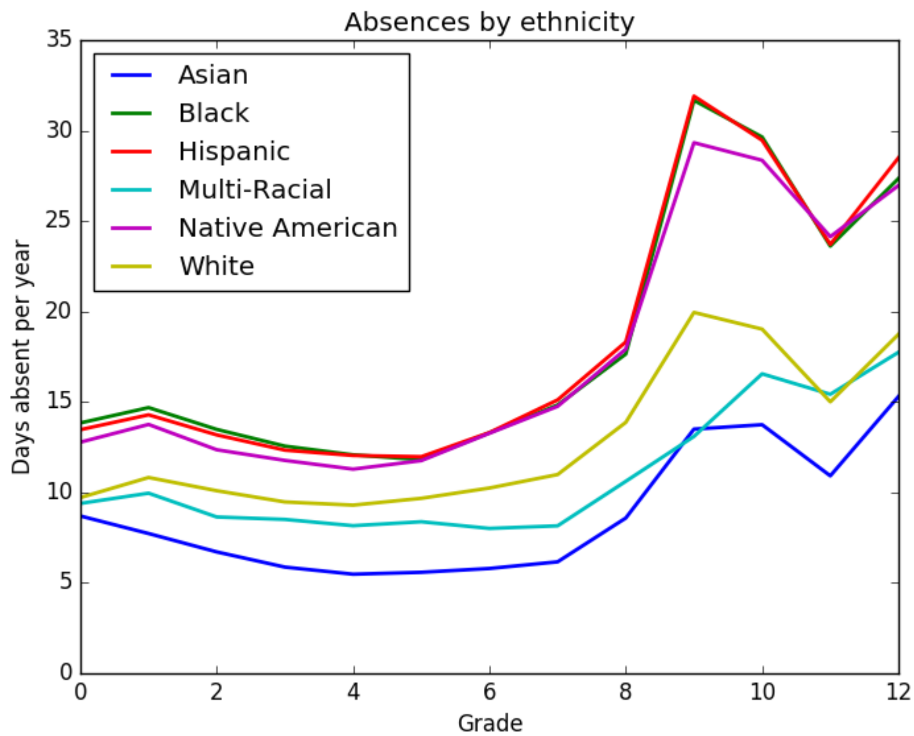


Table 3: *Predictors of English Language Arts scores. *** indicates significance at the 0.1% level.*

	ELA Score	ELA Score	ELA Score
Constant	-0.142***	0.183***	0.167***
Disability	0.223***	0.178***	0.181***
English Language Learner	-0.295***	-0.174***	-0.171***
Free Lunch	-0.013***	-0.004***	0.005***
Female	0.042***	0.032***	0.03***
Special Ed	-0.601***	-0.469***	-0.448***
Special Ed Missing	-1.656***	-1.44***	-1.444***
Attendance (z score)			**0.079*** **
L1 Attendance (z score)		**0.046*** **	**-0.006*** **
L1 Attendance Missing		-0.767***	-0.776***
L1 ELA Score		0.092***	0.089***
L1 Score Missing		0.19***	0.189***
L1 Math Score		0.356***	0.356***
Ethnicity	X	X	X
Home Language	X	X	X
Year Indicators	X	X	X
N	4.2 M	4.2 M	4.2 M
R-squared	0.761	0.813	0.817

Table 4: *Predictors of math scores. *** indicates significance at the 0.1% level.*

	Math Score	Math Score	Math Score
Constant	-0.193***	-0.369***	-0.389***
Disability	0.156***	0.139***	0.143***
English Language Learner	-0.128***	-0.049***	-0.046***
Free Lunch	-0.021***	-0.008***	0.002***
Female	0.023***	0.014***	0.012***
Special Ed	-0.613***	-0.481***	-0.454***
Special Ed Missing	-1.632***	-1.502***	-1.507***
Attendance (z score)			**0.1***
L1 Attendance (z score)		**0.056***	**-0.01***
L1 Attendance Missing		-0.165***	-0.177***
L1 ELA Score		0.064***	0.06***
L1 Score Missing		0.171***	0.17***
L1 Math Score		0.357***	0.357***
Ethnicity	X	X	X
Home Language	X	X	X
Year Indicators	X	X	X
N	4.2 M	4.2 M	4.2 M
R-squared	0.812948	0.843909	0.85028

Table 5: *Predictors of attendance. *** indicates significance at the 0.1% level.*

	Attendance (z score)	Attendance (z score)
Constant	0.468***	0.208***
disability	-0.077***	-0.059***
English Language Learner	-0.087***	-0.043***
Free Lunch	-0.114***	-0.064***
Female	0.045***	0.025***
Special Ed	-0.492***	-0.245***
Special Ed Missing	-0.031***	0.019***
L1 Attendance		0.666***
L1 Attendance Missing		0.193***
L1 ELA Score		0.038***
L1 Score Missing		-0.157***
L1 Math Score		-0.025***
Ethnicity	X	X
Home Language	X	X
Year Indicators	X	X
N	8.7M	8.7M
R-squared	0.061	0.316

Figure 3: *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.*

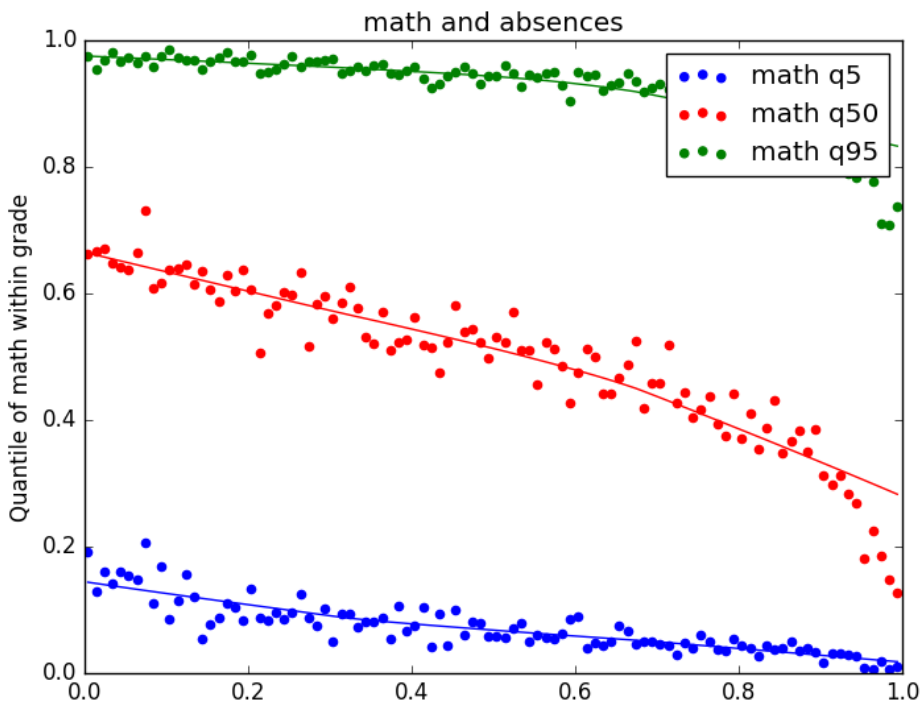
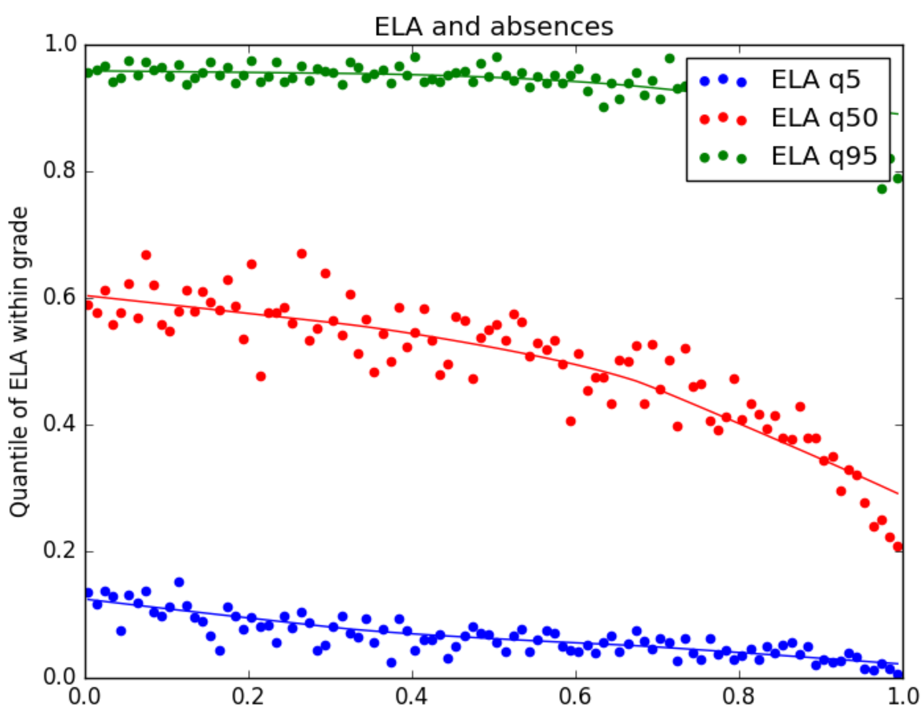


Figure 4: *Dots plot the fifth, fiftieth, and ninety-fifth percentiles of English Language Arts test scores, within each percentile of attendance. The solid line is a quadratic trend.*



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 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). The first part of this step is residualizing outcomes by estimating the α and β of Equation 1 by regressing outcomes on teacher indicators and covariates:

$$\hat{\alpha}^h, \hat{\beta}^h, m^h = \underset{a,b,m}{\arg\min} \sum_{i,c} \left(y_{ic} - x_{ic}^T b - m_{j(i,c)} - a \right)^2 \quad \text{subject to} \quad \sum_j m_j^h = 0$$

Key to estimating Σ_μ is that error in Equation 1 are independent in different classrooms. 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

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

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:

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

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 $\omega^n \sim \text{Dirichlet}(1, 1, \ldots, 1)$. First, estimate

$$\hat{\alpha}^{h,n}, \hat{\beta}^{h,n}, \hat{m}^{h,n} = \underset{a,b,m}{\arg\min} \sum_{i,c} \omega_{j(i,c)}^n \left(y_{ic} - x_{ic}^T b - m_{j(i,c)} - a \right)^2 \quad \text{subject to} \quad \sum_j m_j^h = 0$$

Then the n^{th} draw of Σ_μ is

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

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 account for ethnicity, gender, age, limited English proficiency status, free or reduced-price lunch eligibility, repeating a grade, twice-lagged absences, and twice-lagged test scores, as well as classroom- and school-level averages of these variables.

Although this 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 6 and 11, and in Tables 7 and 11.

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 mostly elementary school teachers who teach math and English. Figures for math teachers are here and figures for English teachers are in Appendix A.

The diagonal of Σ_μ is the magnitude teacher effects on each outcome. For example, the diagonal element of Σ_μ corresponding to four-year high school graduation rates is 0.0068, indicating that the variance of teacher effects on high school graduation is 0.0068, and a teacher who is one standard deviation above average at improving graduation increases graduation rates by the square root of 0.0068, 8.2 percentage points. (Table 10 shows that the standard deviation of English teachers' effects on on high school graduation rates is only about 2 percentage points; these outcomes may be measured with noise, as relatively few students have data on both teachers and graduation.) Figure 5 shows 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 0.11, in line with previous results, indicating that math teachers account for 1.3% of variance in same-year math test scores (0.11^2). The standard deviation of math teacher effects is slightly larger one and two years later. This does not indicate that math teachers are better at preparing their students for future tests than present tests, but rather that math teachers vary more in their effects on future test scores. One interpretation of these results is that teachers prepare similarly for their year's math test, but vary more in the degree to which they teach lasting skills. Three and four years later, effect sizes shrink. The pattern is similar for ELA test scores. For attendance, the standard deviation of math teacher effects is about 0.055 zero to two years out and then fades out. Appendix Figure 10 shows the analogous figure for language arts teachers. Table 6 displays these results for math teachers, and Table 10 displays these results for ELA teachers.

My findings on test scores are consistent with previous literature and but do not

Figure 5: *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 90% credible interval.*

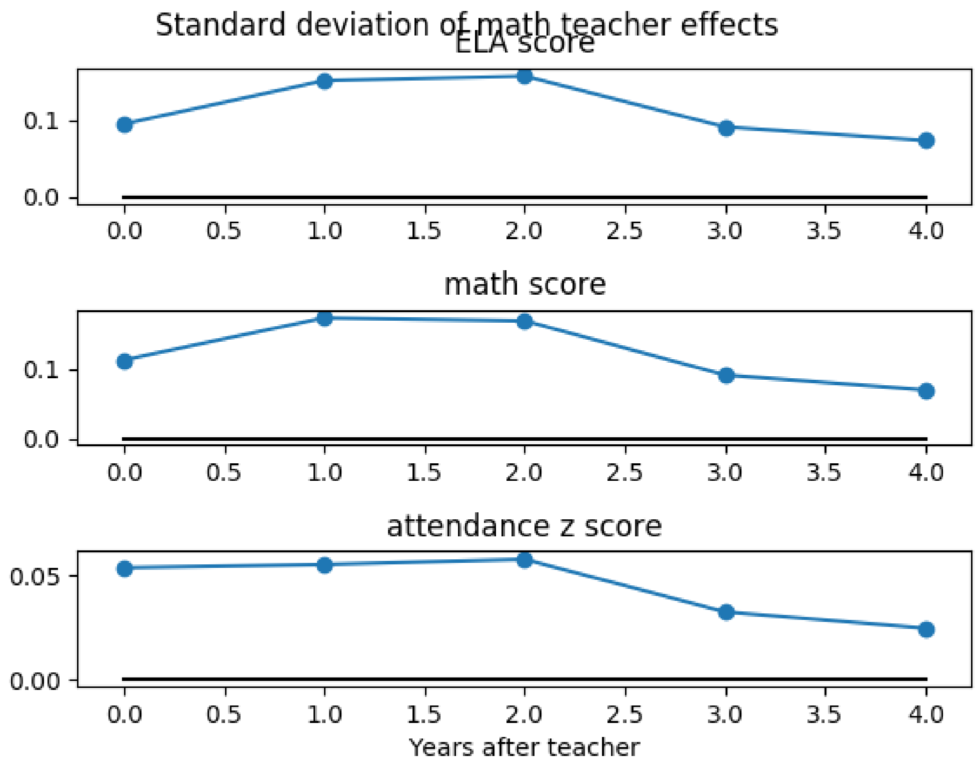
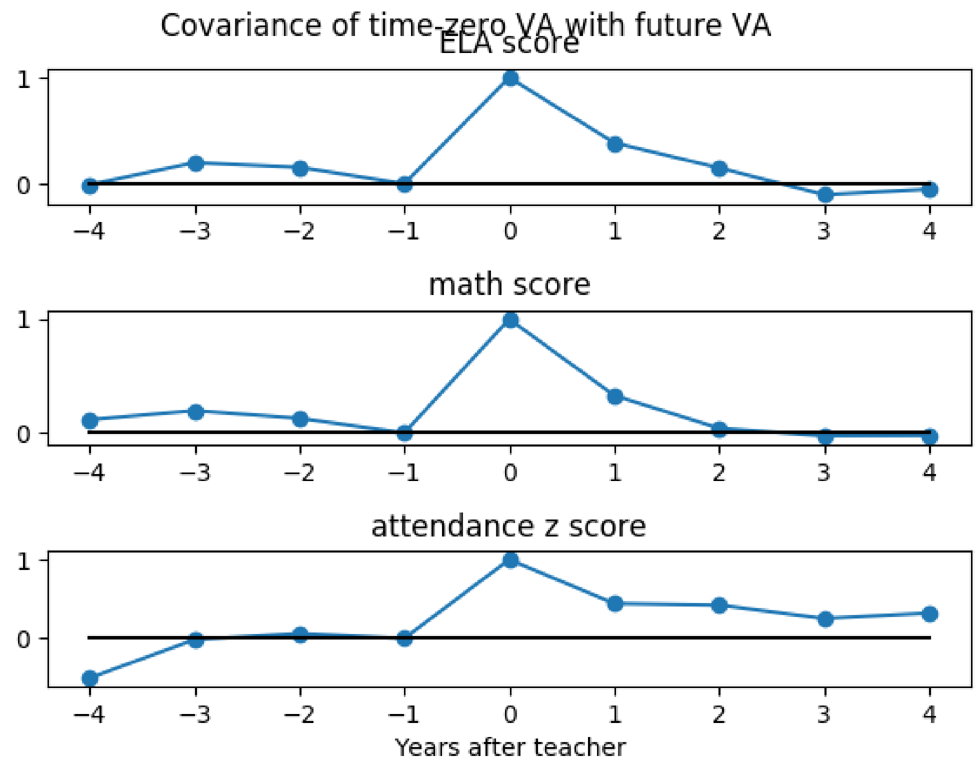


Table 6: *The standard deviation of math teacher effects on test scores and attendance, one to four years out, and on four-year high school graduation. This table displays the same information as 5, but in standard deviation units. 90% credible interval in parentheses.*

	Year				
	0	1	2	3	4
ELA score	0.099	0.076	0.068	0.037	0.013
	(0.081, 0.12)	(0.062, 0.086)	(0.049, 0.077)	(0.0, 0.05)	(0.0, 0.04)
math score	0.099	0.075	0.06	0.049	0.047
	(0.089, 0.109)	(0.064, 0.084)	(0.047, 0.07)	(0.029, 0.064)	(0.025, 0.059)
attendance z score	0.032	0	0	0	0
	(0.021, 0.04)	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)
graduated	0				
	(0, 0)				

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.

Figure 6: *The best linear predictor coefficient of a teacher's effect on a future outcome given her effect on a present outcome. Error bars plot the 90% credible interval.*



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:

$$\mathbb{E}\left[\mu_j\mu_j^T|x_i\right] = \Sigma_\mu. \tag{3}$$

Table 7: *Best linear predictor coefficients. 90% credible interval in parentheses.*

	Year				
	-4	-3	-2	-1	0
ELA score	0.327	0.369	0.324	0	1
	(0.196, 0.738)	(0.255, 0.638)	(0.21, 0.45)	(-0.0, 0.0)	(1.0, 1.0)
math score	0.065	-0.04	0.016	-0	1
	(-0.151, 0.219)	(-0.146, 0.065)	(-0.057, 0.063)	(-0.0, -0.0)	(1.0, 1.0)
attendance z score	0.329	0.735	0.275	0	1
	(-0.286, 1.12)	(0.155, 2.3)	(-0.211, 0.739)	(-0.0, 0.0)	(1.0, 1.0)
	Year				
	1	**2**	**3**	**4**	**Graduated**
ELA score	0.727	0.605	0.355	0.294	0.177
	(0.622, 0.811)	(0.424, 0.702)	(0.202, 0.437)	(0.19, 0.409)	(0.099, 0.337)
math score	0.485	0.389	0.301	0.287	0.173
	(0.381, 0.569)	(0.284, 0.485)	(0.193, 0.4)	(0.188, 0.428)	(0.087, 0.265)
attendance z score	0.238	0.135	0.217	0.501	-0.83
	(-0.614, 0.723)	(-0.819, 0.517)	(-0.411, 0.739)	(0.23, 1.05)	(-2.23, -0.333)

Combining Equations 1 and 3,

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

Figure 6 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 6 and Table 7 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 5, 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 11 and Table 11 repeat the same figure and table for English teachers.

Table 8: *Goodness of proxy for graduation and four-year-lead test scores and attendance, using same-year test scores and attendance. 90% credible set in parentheses.*

	Goodness of Proxy
graduation	2.891877
+4 year ELA score	0.006666
+4 year math score	0.005839
+4 year attendance z score	0.587426

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 assuming 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 μ_j^Q for outcomes in set Q. Then the expectation of μ_j^h given μ_j^Q is

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

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.

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

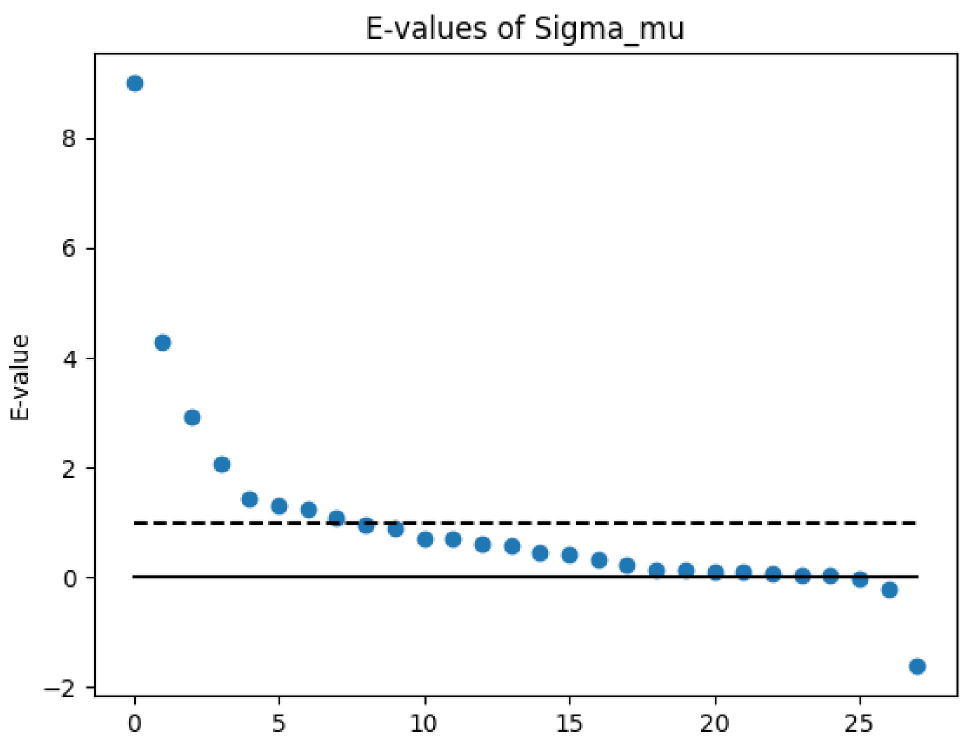
Table 8 and Appendix Table 12 show the goodness of proxy for outcomes graduation and four-year-lead test scores and attendance, using same-year test scores and attendance as a proxy. These tables show that teacher effects on present test scores do not predict teacher effects on current test scores well, but that attendance effects are much more predictable. This is consistent with Tables 6 and 10, which show that teachers vary substantially in their effects on test scores four years in the future, and with Tables 7 and 11, which show that teachers who improve test scores do not have persistent effects, while teachers who improve attendance do.

5.2 Factor analysis

Can Σ_μ be represented well by several factors? What are they? I would like to answer this question with a factor analysis, but have not yet completed it.

There are several rules of thumb for choosing the number of factors. One is to count the number of eigenvalues that are greater than one, after normalizing the covariance matrix to have all ones on the diagonal. This heuristic suggests eight eigenvalues for both ELA and math teachers. Another heuristic suggests plotting the eigenvalues, as in Figures 7 and 8, and look for a point where the eigenvalues start to level off. This heuristic suggests about four factors for math teachers and three for ELA teachers.

Figure 7: *Eigenvalues of Σ_μ for math teachers.*



After choosing the number of factors, the factor structure can be estimated by maximum likelihood analysis (Jöreskog, 1970).

I have not performed the factor analysis yet, but one hint of what the factors may look like comes from Table 9, 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. Table 9 shows correlations between teachers' effects on different outcomes.

Figure 8: *Eigenvalues of Σ_μ for English teachers.*

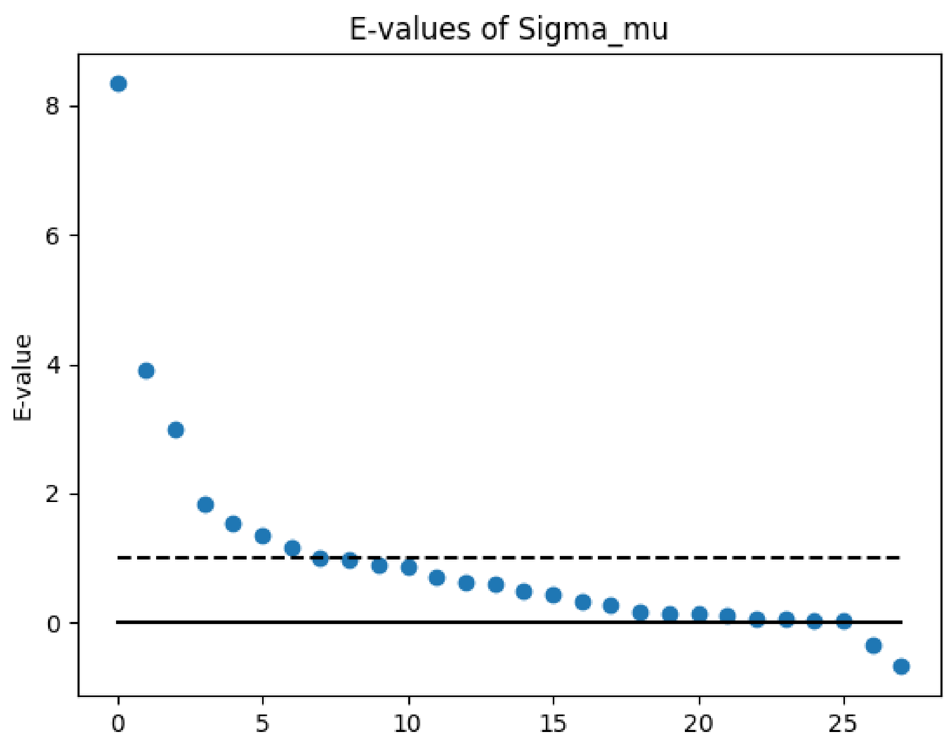


Table 9: *Correlations between teachers' value-added on different outcomes.*

	Attendance (Elem)	Attendance (Mid)	ELA scores	Math scores
Attendance (Elem)	1	NA	0.0275	0.0364
Attendance (Mid)	-	1	0.0143	0.0594
ELA	-	-	1	0.2158
Math	-	-	-	1

5.3 Identification and robustness checks

To-dos:

- Hypothesis test on pre-trends
- Inference: Currently implemented on 10% sample, waiting for things to run on whole sample
- Sensitivity of parameter estimates to covariates

6 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.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, **57** (2), 239–251.

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 9: *Histogram of student absence frequency.*

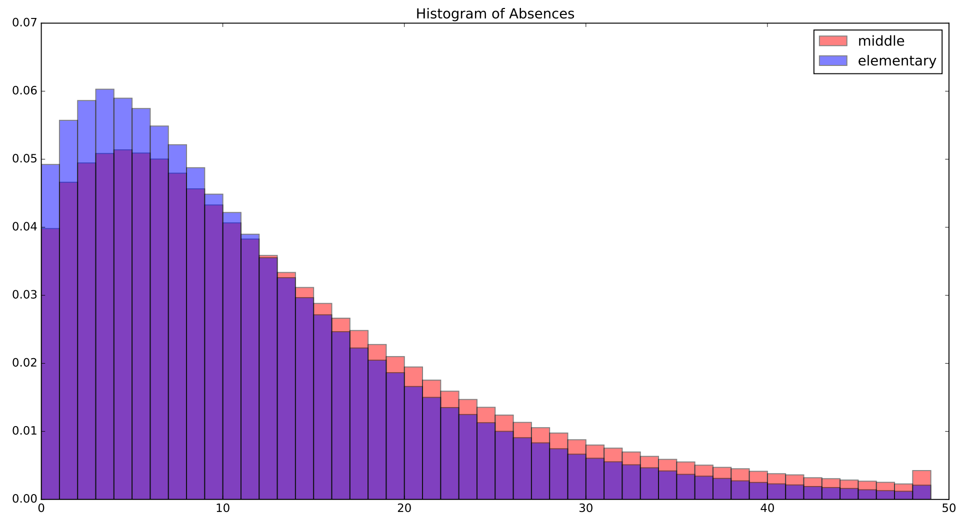


Table 10: *The standard deviation of ELA teacher effects on test scores and attendance, one to four years out, and on four-year high school graduation.*

	Year				
	0	1	2	3	4
ELA score	0.152	0.139	0.129	0.116	0.122
	(0.151, 0.155)	(0.135, 0.143)	(0.129, 0.133)	(0.115, 0.119)	(0.121, 0.129)
math score	0.161	0.153	0.138	0.126	0.127
	(0.158, 0.167)	(0.151, 0.158)	(0.134, 0.141)	(0.122, 0.129)	(0.123, 0.134)
attendance z score	0.066	0.058	0.054	0.045	0.046
	(0.064, 0.071)	(0.056, 0.061)	(0.053, 0.06)	(0.041, 0.046)	(0.04, 0.048)
graduated	0.037				
	(0.033, 0.039)				

Figure 10: *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 next figures repeat this for math scores and attendance.*

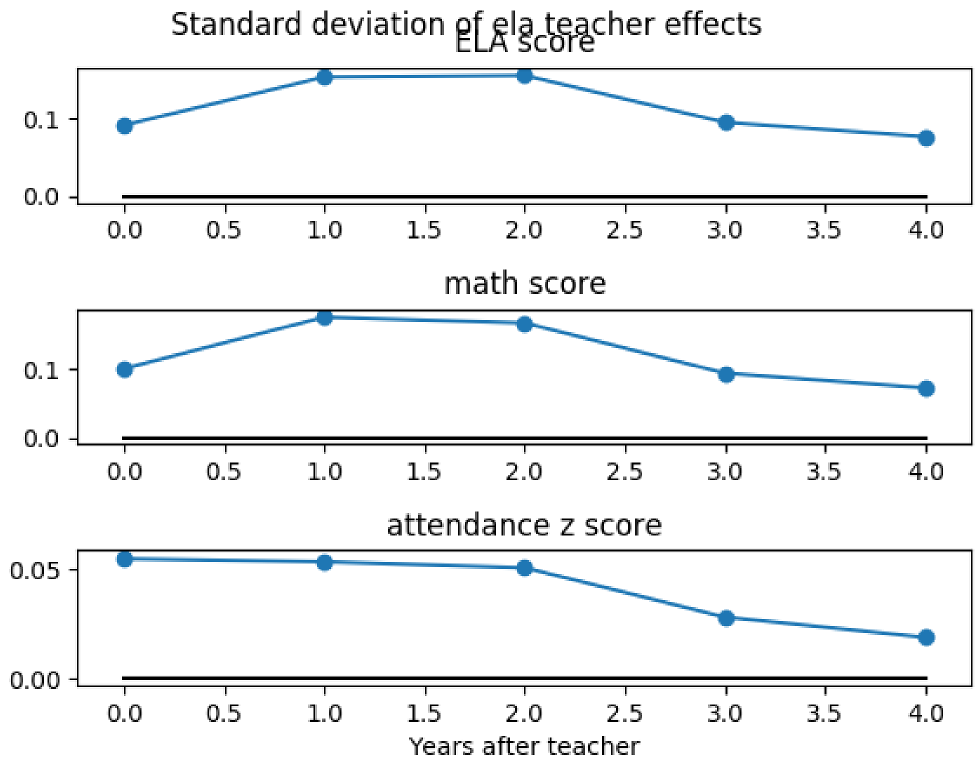


Figure 11: *The best linear predictor coefficent of a teacher's effect on a future outcome given her effect on a present outcome.*

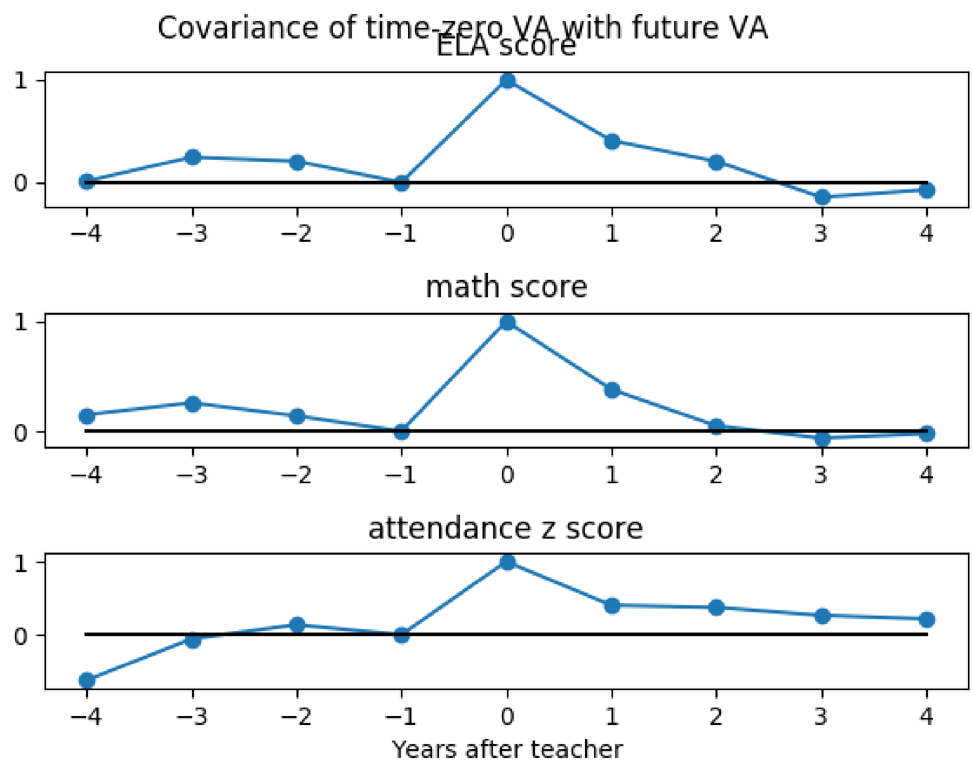


Table 11: *Best linear predictor coefficients.*

	Year				
	-4	-3	-2	-1	0
ELA score	0.182	0.214	0.212	-0	1
	(0.16, 0.191)	(0.18, 0.222)	(0.192, 0.243)	(-0.0, 0.0)	(1.0, 1.0)
math score	0.07	0.113	0.073	0	1
	(0.028, 0.092)	(0.083, 0.145)	(0.061, 0.097)	(0.0, 0.0)	(1.0, 1.0)
attendance z score	0.382	0.391	0.254	-0	1
	(0.375, 0.422)	(0.365, 0.455)	(0.228, 0.304)	(-0.0, -0.0)	(1.0, 1.0)
	Year				
	1	2	3	4	Graduated
ELA score	0.765	0.694	0.622	0.637	0.211
	(0.746, 0.787)	(0.695, 0.713)	(0.605, 0.651)	(0.621, 0.72)	(0.203, 0.237)
math score	0.688	0.587	0.586	0.551	0.183
	(0.666, 0.706)	(0.562, 0.624)	(0.56, 0.604)	(0.536, 0.614)	(0.164, 0.207)
attendance z score	0.573	0.499	0.364	0.308	-0.305
	(0.547, 0.604)	(0.48, 0.563)	(0.307, 0.397)	(0.238, 0.332)	(-0.336, -0.256)

Table 12: *Goodness of proxy for graduation and four-year-lead test scores and attendance, using same-year test scores and attendance.*

	Goodness of Proxy
graduation	0.669786
+4 year ELA score	0.009912
+4 year math score	0.005556
+4 year attendance z score	0.522906