THE UNIVERSITY OF CHICAGO

THE MULTIDIMENSIONAL IMPACT OF TEACHERS ON STUDENTS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECONOMICS

BY

NOLAN G. POPE

CHICAGO, ILLINOIS

JUNE 2017

TABLE OF CONTENTS

FIGURES

TABLES

ABSTRACT

For decades, policymakers and researchers have used value-added models that rely solely on student test scores to measure teacher quality. However, since teaching ability is multidimensional, test-score value-added measures of teacher quality may not fully capture the impact of teachers on students. In this paper, we use test-score and non-test-score measures of student achievement and behavior from over a million students in the Los Angeles Unified School District to estimate multiple dimensions of teacher quality. We find that test-score and non-test-score measures of teacher quality are only weakly correlated, and that both measures of teacher quality affect students' performance in high school. A teacher-removal policy simulation that uses both dimensions of teacher quality improves most long-term student outcomes by over 50 percent compared to a policy that uses test scores alone. Our results also show that the long-term effects of teachers in later grades are larger than in earlier grades, that performance in core elementary school subjects matters more for long-term outcomes than other subjects, and that non-test-score dimensions of ability capture additional sources of peer effects.

1 Introduction

Teacher quality has garnered the attention of policymakers and researchers for many years. Researchers have primarily measured teacher quality using a test-score value-added framework.[1] Although the use of test-score value-added has substantially impacted education research, people have long recognized that good teachers likely affect a wide range of student outcomes. In fact, early theoretical formulations of value-added used an education production function that modeled educational output as a "multidimensional factor" (Hanushek 1971). Consequently, measures of teacher quality that rely solely on student test scores may not fully capture the impact of teachers on students.

In this paper, we are interested in whether teachers can noticeably impact measures of student achievement beyond just test scores, and whether the impact on non-test-score measures is important for the future success of students. Specifically, how much heterogeneity is there in teachers' ability to impact non-test-score measures of student achievement (e.g., suspensions or attendance)? Are these non-test-score value-added measures correlated with test-score value-added measures? Are these additional measures as predictive of the future success of students as test-score value-added measures? All of these questions have important implications for education policy and how we think about measuring teacher quality.

To answer these questions, we gather administrative data from the Los Angeles Unified School District (LAUSD) for students in grades K-12 from 2003 to 2015. These data link over a million students to teachers, and track students over time as they progress through the LAUSD system. Our three measures of student achievement are constructed from (1) student math and English state test scores, (2) measures of student behavior, including suspensions, attendance, GPA, and grade retention, and (3) teacher assessments of student effort and 14 learning skills. The learning skills include teacher assessments such as whether a student makes good use of time, exercises self-control, and resolves conflicts appropriately. We measure the long-term effects of teachers

[1] An important exception is a working paper by Kirabo Jackson (2016) that estimates non-test-score measures of teacher quality. This paper is described fully in the literature review section along with the relative contributions of his and our paper.

using student performance in high school, including dropping out of high school, taking the SAT, SAT scores, high school exit exam scores, GPA, teacher assessments of effort and cooperation, attendance, suspensions, and grade retention.

We first document that elementary school students with better test scores, behavior, and learning skills perform better in high school. We then estimate teacher value-added measures of three dimensions of teacher quality – student test scores (using math and English state tests), student behavior (using GPA, attendance, suspensions, and grade retention), and student learning skills (using teacher assessments of effort and 14 learning skills). Using these value-added measures, we show that teachers affect both test-score and non-test-score dimensions of student achievement. To avoid bias and potential teacher manipulation when using teacher-reported non-test-score variables, we modify the standard value-added framework to use student outcomes from the year after the student was in a teacher's class, instead of the contemporaneous year.

We find that having a high test-score value-added teacher in elementary school improves students' high school performance. These long-term effects of test-score value-added are not substantially reduced by adding teachers' behavior or learning-skills value-added to the model. This result suggests that the long-term effects of test-score value-added may not be biased by omitting non-test-score teaching ability.

We also find that behavior value-added is only weakly correlated with test-score value-added, and has a similarly large effect on students' long-term outcomes. Therefore, test-score value-added misses the dimensions of teacher quality captured by behavior value-added that matter for long-term outcomes. Consequently, test-score value-added underestimates the total effect of teachers on students. The low correlation between the two value-added measures also suggests that using behavior value-added in conjunction with test-score value-added may substantially improve the accuracy with which overall teacher quality is measured.

We illustrate how behavior value-added improves the measurement of teacher quality using a hypothetical policy simulation that replaces teachers in the bottom 5 percent of the teacher quality distribution with district average teachers. Relative to relying on test-score value-added alone, a simple rule that equally weights the test-score and behavior value-added of a teacher results in at least a 50 percent improvement in the likelihood of dropping out of high school, taking the SAT, high school GPA, suspensions, absences, and on-time progression. These gains are obtained with little to no decline in student test scores, are similar to the gains obtained if an optimal weighting scheme is used, and do not require administering additional tests or using data beyond what schools typically collect.

In addition to test-score and behavior value-added, we find significant effects of learning-skills value-added alone on some high school outcomes. However, these effects dissipate after we control for both test-score and behavior value-added. Learning-skills value-added captures a dimension of teacher quality that is also measured by a linear combination of test-score and behavior value-added. Since the variables used to construct learning-skills value-added are plausible measures of noncognitive skills, the other two value-added measures may measure some aspects of noncognitive teaching ability.

Finally, we use test-score and non-test-score measures of ability in three applications. First, we estimate the effect of test-score and behavior value-added for each grade 3 to 12. We find that middle school and high school teachers have a larger effect on outcomes measured in 11th or 12th grade than elementary school teachers. This result suggests that teachers in later grades may play a more important role in improving long-term student outcomes than teachers in earlier grades. Assuming little tracking, these results imply large cumulative benefits of teacher value-added. For example, giving students a standard deviation better test-score value-added teacher each year from grades 3 to 12 increases the likelihood of taking the SAT by 8.1 percentage points, and reduces the likelihood of dropping out of high school by 0.5 percentage points. Giving students a standard deviation better behavior value-added teacher over the same period increases the likelihood of taking the SAT by 8.3 percentage points, and reduces the likelihood of dropping out of high school by 5.9 percentage points.

Second, the focus on test scores has limited the study of teacher effects to a few regularly tested subjects (e.g., math and English). We instead use subject-specific GPAs to compute value-added measures of teacher quality in 10 elementary school subjects. We find that students with higher value-added teachers in the subjects of math, reading, writing, and health perform better in high school, whereas having a higher value-added teacher in the subjects of speaking and science have negative effects on high school performance. Hiring teachers who are relatively better at teaching these subjects and spending more time on these subjects could potentially improve long-term student outcomes. Finally, we show that consistent with our main results, high-achieving test-score and behavior peers improve their classmates' short-term outcomes, whereas high-achieving learning-skills peers have little effect.

From a policy perspective, there are potentially large benefits from adopting a measure of teacher quality that includes both test-score and non-test-score dimensions. For example, policymakers can use non-test-score value-added to measure teacher quality for all teachers, not just math and English teachers. In addition, since focusing on only one output of the multidimensional education production function (i.e., test scores) may distort the efficient allocation of teachers' time and resources, using a broader measure of teacher quality may help alleviate this distortion. Finally, using a better measure of overall teacher quality can improve school districts' hiring and tenure decisions.

The rest of the paper will proceed as follows. Section 2 reviews the related literature. Section 3 describes the LAUSD data that we use and, in particular, describes the variables used to measure test-score, behavior, and learning-skills value-added. Section 4 outlines the empirical method for estimating teacher value-added measures and estimating the effect of teacher value-added on long-term student outcomes. Section 5 presents the descriptive results of the test score, behavior, and learning-skills value-added of teachers, and then reports the results for how teachers affect students' concurrent and long-term outcomes. The gains from teacher-removal policies that use multiple dimensions of teacher quality are also presented. Section 6 presents the relative value of higher quality teachers over the students' educational life cycle, in specific subjects, and from students' peers. Section 7 concludes.

2 Literature Review

Since the early 1970s, researchers have used test-score valued-added to measure teacher quality (Hanushek 1971). This research led states and school districts to use test-score value-added in teacher evaluations as early as the 1990s (Horn and Sanders 1994). Since then, the use of test-score value-added has expanded, and 27 states require that teacher evaluations include "growth measures as a significant criterion" (The National Council of Teacher Quality 2015). This increased use of test-score value-added has largely been due to a lack of other predictors of teacher quality (Hanushek and Rivkin 2010). Much of the recent work in the value-added literature focuses on the validity of value-added models (Bacher-Hicks, Kane, Staiger 2014; Chetty, Friedman, and Rockoff 2014a; Kane and Staiger 2008; Kane et al. 2013; Rockoff 2004; Rothstein 2010; Rothstein 2015), gains from using them in personnel decisions (Goldhaber and Hansen 2010; Gordon, Kane, and Staiger 2006; Hanushek 2011), and theoretical and empirical studies of their use in pay-for-performance (Fryer 2013; Goodman and Turner 2013; Neal 2011). This literature was recently highlighted by Chetty, Friedman, and Rockoff (2014b), who find that students with higher test-score value-added teachers earn significantly more by their late 20s, have fewer births as teenagers, and are more likely to attend college.

A small number of recent papers in multiple disciplines push beyond test-score value-added by examining teacher effects on contemporaneous non-test-score measures of student achievement. These non-test-score measures include social and behavioral skills (Jennings and DiPrete 2010), motivation (Ruzek et al. 2014), absences (Gershenson 2016), belief in the ability to do math and happiness in math class (Blazar and Kraft 2016), and grit, growth mindset, effort, and answering open-ended questions (Kraft and Grace 2016). Other studies assess multidimensional teacher effects using non-value-added approaches (Mihaly et al. 2013; Rockoff and Speroni 2010). However, these papers do not analyze whether the effects of non-test-score value-added measures persist through a student's education and matter for long-term outcomes, or whether test-score value-added captures the portion of these abilities that matter for long-term outcomes. Consequently, the extent to which non-test-score teacher quality influences the future success of students is unclear.

An important exception, and the paper most closely related to ours, is Jackson (2016). Analyzing test-score and behavioral value-added for 9th graders in North Carolina, he estimates the effects of teacher quality in 9th grade on outcomes in 12th grade. He finds behavioral value-added increases high school graduation rates and reduces high school dropout rates, whereas the effect of test-score value-added is statistically insignificant, conditional on a teacher's behavioral value-added. In addition, he finds ambiguous effects of both value-added measures on GPA, taking the SAT, and plans for college.[2] These results suggest behavioral value-added matters for some high school outcomes, but provide limited evidence that 9th-grade teacher quality is multidimensional because test-score value-added does not significantly affect high school outcomes after conditioning on behavioral value-added.

Our paper contributes to the literature in a number of ways. First, while Jackson's (2016) results suggest that only behavioral value-added affects long-term student outcomes, our results demonstrate that teacher quality is multidimensional and that significant long-term effects occur for both test-score and behavior value-added. Second, our policy simulation not only shows that teacher quality is multidimensional, but also that using better measures of teacher quality (both test-score and behavior value-added) substantially increases the benefits of selective dismissal policies. Third, our paper uses an additional, direct measure of noncognitive skills (i.e., learning-skills valued-added), which illuminates the relationship between behavior value-added and noncognitive skills. Our results suggest that behavior value-added, in part, measures noncognitive teaching ability but also captures other abilities (Heckman, Pinto, and Savelyev 2013). Finally, a concern with Jackson's (2016) analysis is that his behavioral value-added captures the direct effect of a teacher on an outcome (e.g., a strict teacher failing a student in a math class) rather than the effect of a teacher on the underlying ability of the student (e.g., a teacher negatively affecting a student's

[2]For these three outcomes, Jackson (2016) estimates effects for four value-added measures: math test scores, English test scores, math behavior, and English behavior. One of the six test score value-added measures has a statistically significant effect with the expected sign. All three English behavior factors have significant or marginally significant effects with the expected sign, but the math behavior estimates are larger in magnitude with the opposite sign, although they are not statistically significant.

knowledge of math, through poor teaching, such that the student fails her math class). We use student outcomes for the year after a student was in a teacher's class instead of the concurrent year to avoid bias from capturing the direct effects of teachers on high school outcomes unrelated to their effects on student ability.

This paper also uses non-test-score value-added in three additional applications. First, previous research on the effect of test-score value-added (and non-test-score value-added) on long-term outcomes by grade is limited. Chetty, Friedman, and Rockoff (2014b) examine grades 4 through 8 for test-score value-added, and Jackson (2016) studies grade 9 for both test-score and behavior value-added. We analyze the effects of both test-score and behavior value-added in grades 3 through 12, finding substantial differences between grades. Second, because of the focus on test scores, little work has examined the effects of teachers by subject, except for math and English (Jackson, Rockoff, and Staiger 2014). We examine how teachers affect 10 separate elementary school subjects, and the relative benefit each of these subjects provides to students. Lastly, we analyze how students who vary in their test-score, behavior, and learning-skills achievement affect their peers.

3 Los Angeles Student Data and Background

The LAUSD is the second largest school district in the United States, educating over 600,000 students each year. In 2003, the school district was 71.9 percent Hispanic, 12.1 percent black, and 9.4 percent white.[3] We use a panel of student-level administrative data on all public school students in the LAUSD. The panel links students to teachers over time and includes the 2002-03 to 2014-15 school years, which we reference by year of graduation (e.g., we refer to the 2002-03 school year as 2003). Our analysis focuses on the over 110,000 3rd to 5th grade students studying in the LAUSD each year.

These data are unique in the level of detail they provide about each student's academic perfor-

[3]Statistics can be found at http://dq.cde.ca.gov/dataquest.

mance. For grades 2 through 11, math and English California state test (CST) scores are available for each student. The testing regime is relatively consistent over this period, with the only major change being an essay section added to the 4th- and 7th-grade English test in 2011. For all grades, these data contain the number of days a student was suspended, the number of days a student was absent, and whether a student did not progress on time to the next grade (i.e., held back). Both elementary and high school students received progress reports with their grades by subject and a number of additional teacher assessments of student performance.

Elementary school progress reports (grades K-5) are given each trimester, and contain achievement grades in 10 subjects (e.g., reading, mathematics, art, etc.), effort grades for the same 10 subjects, grades for five "work and study habits" (e.g., "makes good use of time," "organizes materials," etc.), and grades for nine "learning and social skills" (e.g., "resolves conflicts appropriately," "exercises self-control," etc.). All grades are on a 4-point scale, with no fractional points given. We compute an annual GPA for each of the four groups listed above. Figure 1 shows a template of the progress report.

Starting in the 6th grade, middle school and high school students receive progress reports each semester, with three categories of grades for each of their classes: achievement (i.e., academic performance), "work habits," which we term effort (i.e., "effort," "responsibility," "attendance," and "evaluation"), and "cooperation" (i.e., "courtesy," "conduct," "improvement," and "class relations"). Achievement is graded on a 4-point scale, and effort and cooperation are graded on a 3-point scale, with no fractional points given. We compute annual GPAs for each of these three groups of measures. Appendix Figure A.1 shows additional details on grading criteria.

Additional data are available for middle and high school students, including whether a student dropped out of high school (i.e., the student enrolled in the LAUSD in grade 9 and did not graduate high school in the LAUSD within 5 years), graduated from the LAUSD conditional on enrolling in the LAUSD in 12th grade, SAT scores, PSAT scores, math and English California High School Exit Examination (CAHSEE) scores, science CST scores (grades 5, 8, and 10), social science CST scores (grades 8, 11, and world history), and the number of AP courses taken. All test scores are normalized to be mean zero and standard deviation one at the grade-year level, except both SAT and PSAT scores, which we place on a 600-2400 scale (PSAT is normally on a 60-240 scale, and for some years, the SAT was on a 400-1600 scale). We top code days absent at 180 days per year, and report log absences as the log of one plus the number of absences.

Described fully in sections 4.1 and 4.2, we compute test-score value-added measures using math and English CST scores, and behavior value-added using log days absent, achievement GPA, an indicator for suspensions, and an indicator for being held back. For elementary school teachers, we compute learning-skills value-added using the three additional types of elementary school GPAs. We reduce the dimensionality of both the inputs to the value-added variables and the value-added variables themselves by creating equally weighted indices.

Our main outcome variables are measures of high school performance, including an indicator for dropping out of high school, an indicator for taking the SAT, SAT scores, the three high school GPA measures averaged from grades 9-12, math and English California high school exit-exam scores, days suspended in grades 9-12, log absences in grades 9-12, and an indicator for being held back in grades 9-12. We treat graduation as a supplemental measure because it is conditional on enrolling in the LAUSD in 12th grade.

Summary statistics for these data are shown in Table 1. Panel A shows summary statistics for all students in grades 3-5 from the 2004 to 2010 school years. Panel B shows high school summary statistics for the students in Panel A who attended high school in the LAUSD. The LAUSD dropout variable overestimates the dropout rate by a factor of about 1.5 because it includes both dropouts and students who transferred to schools outside of the LAUSD. The actual LAUSD dropout rate, estimated by the LAUSD, was 37.2 percent in 2010. The graduated variable is an overestimate of the actual graduation rate because it is conditioned on entering the 12th grade.

4 Empirical Method

4.1 Estimating Teacher Value-Added

Let S_{ijt} be a measure of student i's test scores, behavior, or learning skills in year t (in teacher j's class). For example, S_{ijt} could be a standardized test score, an indicator for whether the student was suspended, or a teacher's assessment of a particular learning skill. The goal is to estimate the effect of a teacher on several measures of students' test scores, behavior, and learning skills. Recent research estimating teacher test-score value-added and its affects on long-term outcomes has used slightly different estimation strategies (Chetty, Friedman, and Rockoff 2014b; Jackson 2016; Rothstein 2015). Our approach combines elements of each, and we show in the Appendix that the main results are robust across a range of estimation strategies.

For test-score value-added, we use the following estimation procedure.[4] Although very similar, a small but important adjustment is made when estimating behavior and learning-skills value-added, which is discussed below. We construct value-added measures by first residualizing the achievement measure, S_{ijt}, by regressing it on a vector of controls, X_{ijt}, for lagged student achievement and the classroom environment. The controls include lags of a third-order polynomial of the student's math test score, English test score, GPA, log days absent, an indicator for suspensions, an indicator for being held back, work and study habits GPA, learning and social skills GPA, and effort GPA. We include third-order polynomials of lagged class and grade level means of each of those variables, and English learner status for the individual, class, and grade. We also fully interact each of these variables except English learner status with grade fixed effects, and include a control for class size. For middle and high school students, we exclude the effort GPA, work and study habits GPA, and learning and social skills GPA controls for lack of data:

$$S_{ijt} = \Gamma X_{ijt} + \varepsilon_{ijt}, \tag{1}$$

[4]We compute value-added measures using a program written by Michael Stepner for Chetty, Friedman, and Rockoff (2014a, 2014b).

where $\varepsilon_{ijt} = \mu_{jt} + \alpha_c + \gamma_{it}$.

Suppose the error term is an additively separable function of teacher quality (μ_{jt}), classroom shocks (α_c), and idiosyncratic student-year shocks (γ_{it}). Following Chetty, Friedman, and Rockoff (2014a), we assume teacher quality and student achievement follow a stationary process such that the expected value of teacher quality is not a function of time, and the serial correlation of each of the components of ε_{ijt} is determined by the number of years between them. Embedded in this approach is the special case often used in the value-added literature, where a teacher's quality is the same in each year (i.e., $\mu_{jt} = \mu_j$ for all t).

Let v_{ijt} be the residualized student achievement computed using this procedure:

$$v_{ijt} = S_{ijt} - \hat{\Gamma} X_{ijt}. \tag{2}$$

The residualization purges S_{ijt} of measures of prior achievement at the student, class, and grade levels. We then take the mean of the residuals, $\bar{v}_{jt}$, in each year for all students taught by teacher j, which provides an estimate of the teacher's ability to affect student achievement in each year t under some assumptions.

The key assumption is that students are not sorted to teachers on unobservable components of student achievement, that is, $\mathrm{E}[\alpha_c + \gamma_{it} | j] = \mathrm{E}[\alpha_c + \gamma_{it}]$. If some teachers systematically have students with better or worse unobservable components of student achievement, the estimated value-added measures will pick up differences in unobservables and not just the teacher's causal impact. Although this is a strong assumption, it is plausible in this context because the value-added model includes extensive controls for students' prior achievement and behavior in school that have been shown to account for most student sorting (Bacher-Hicks, Kane, and Staiger 2014; Chetty, Friedman, and Rockoff 2014a). To help alleviate concerns about student sorting based on unobservable determinants of student achievement, in section 5.6, we check for forecast bias, examine the effect of teacher value-added on predicted outcomes as a placebo test, and perform a quasi-experimental analysis that uses teachers switching grades and schools.

We then predict teacher quality in year t, $\hat{v}_{jt}$, using measures of teacher quality in all years except t to avoid biasing estimates of the long-term effects of teacher quality on student outcomes (Jacob, Lefgren, and Sims 2010). Including year t would likely bias the long-term estimates because the same estimation error might be introduced to both the long-term outcome and the estimated value-added measure. We allow the weight placed on the value-added measure in each year to vary by the number of years before or after year t. We compute the weights by minimizing the mean-squared error of the difference between $\bar{v}_{jt}$ and predictions of $\bar{v}_{jt}$, using the $\bar{v}$ of teacher j in all years except t. This approach is equivalent to fitting a least squares model using estimated teacher quality in year t as the outcome with leads and lags of estimated teacher quality as the independent variables, and then using the coefficients produced by this model to predict teacher quality in year t. This procedure produces leave-year-out, jackknife, value-added estimates that allow for drift in teacher quality.

We modify this procedure when we calculate non-test-score value-added measures by using the lead of the achievement measure, $S_{ij(t+1)}$, as the outcome variable. This approach contrasts with Jackson (2016) and most of the non-test-score value-added literature that uses contemporaneous student outcomes instead of outcomes in the next year, but is closely related to approaches used to calculate the value-added of professors (Carrell and West 2010; Figlio, Schapiro, and Soter 2015). This approach requires the main assumption – that students are not sorted to teachers on unobservable components of student achievement – holds for two years instead of just one.

We use $S_{ij(t+1)}$ because using S_{ijt} creates the potential for the non-test-score value-added measures to capture aspects of teacher behavior unrelated to teachers' ability to affect students' behavior or learning skills. For example, grades are likely affected not only by how much a teacher helps a student learn and work diligently, but also by how strictly the teacher grades. Similarly, suspensions are affected both by whether a teacher causes improvements in student behavior and how harshly or leniently a teacher chooses to punish a student. These types of measurement error could lead to biased estimates of the effect of teacher value-added on student outcomes.

A related concern is that teachers could directly affect long-term outcomes without affecting a student's behavior or learning skills. For example, if a teacher is more likely than other teachers to recommend a student be held back, that student may be more likely to drop out of high school even if the teacher actually has no effect on the student's behavior or learning skills. This potential direct effect could bias the effect of teacher value-added on long-term outcomes in the direction of affecting long-term outcomes.

We remove bias from variation in teacher strictness (or leniency), and the direct effect of teachers on long-term student outcomes, by using the lead of the student achievement measure (i.e., achievement in year $t+1$, rather than the measure of student achievement in year t):

$$S_{ij(t+1)} = \Gamma X_{ijt} + \varepsilon_{ijt}. \tag{3}$$

This approach introduces noise to our estimates because it partially captures the effect of the teacher in year $t+1$, but removes systematic bias from teachers evaluating their own students. In addition, using student achievement in year $t+1$ makes it more difficult for teachers to manipulate their behavior or learning-skills value-added.

4.2 Estimating the Long-Term Effects of Teacher Value-Added

Once we have leave-year-out estimates of teacher quality, $\hat{v}_{jt}$, we ask how having either a higher- or lower-quality teacher along some dimension of teacher quality affects a student in the long term. Let y_i be a long-term outcome of interest such as whether a student is a high school dropout, an indicator for taking the SAT, or a score on a test required for high school graduation. Let k index a number of distinct leave-year-out value-added measures of test scores, behavior, or learning skills. We regress outcome, y_i, on a number of value-added measures and our controls from equation (1). The estimates of $\hat{\beta}_k$ for each value-added measure assess how each dimension of teacher quality affects the outcome of interest:

$$y_i = \sum_{k=1}^{K} \beta_k \hat{v}_{jkt} + \Gamma X_{ijt} + \eta_{ijt}. \tag{4}$$

We reduce the dimensionality of the estimates of teacher quality by constructing three indices of the value-added variables. The first index is computed using teacher math and English test-score value-added, which we call the test-score value-added, or $\hat{\theta}^{s}_{jt}$. The second value-added index is computed using value-added for suspensions, log days absent, GPA, and not progressing to the next grade on time (i.e., held back), which we call the behavior value-added, or $\hat{\theta}^{b}_{jt}$. The third value-added index is computed using the value-added from effort GPA, work and study habits GPA, and learning and social skills GPA, which we call learning-skills value-added, or $\hat{\theta}^{l}_{jt}$. We chose these three groups because they separate test scores from non-test scores, and because the behavior value-added includes variables that are available for all grades, whereas the learning-skills value-added is available only for elementary school students. The indices are computed by summing the standardized value-added variables, recoded so each has the same expected sign, and then standardizing the resulting index to be mean zero, standard deviation one. In the Appendix, we show that the main results are robust to grouping GPA with learning skills, using factor analysis to construct the three indices, and using exploratory factor analysis to choose the factors and the factor load on each value-added variable.

We estimate the long-term effect of these value-added measures using the following specification:

$$y_i = \beta^s \hat{\theta}^{s}_{jt} + \beta^b \hat{\theta}^{b}_{jt} + \beta^l \hat{\theta}^{l}_{jt} + \Gamma X_{ijt} + \eta_{ijt}. \tag{5}$$

These estimates inform us of the extent to which different dimensions of teacher quality matter for long-term student outcomes. We also compare the estimates from equation (5) with the estimates from a model that omits non-test-score value-added indices. This comparison allows us to sign the bias from omitting non-test-score measures in papers that estimate the effect of teachers' test-score value-added on long-term outcomes. If we find that $\hat{\beta}^s$ falls when we move from a model that excludes $\hat{\theta}^{b}_{jt}$ and $\hat{\theta}^{l}_{jt}$ to one that includes them, it suggests that typical estimates of the long-term effects of test-score value-added are biased upward by omitted measures of behavioral or noncognitive skills. Alternatively, if $\hat{\theta}^{b}_{jt}$ or $\hat{\theta}^{l}_{jt}$ affect long-term outcomes, and the estimate of $\hat{\beta}^s$ is unaffected by adding $\hat{\theta}^{b}_{jt}$ or $\hat{\theta}^{l}_{jt}$, the long-term effects of test-score value-added may be unbiased, but estimates of the total effect of teachers on students is larger than the effects found when using test-score value-added alone.

Let tildes denote residualized student value-added indices, for example, $\tilde{\theta}^s_{jt} = \hat{\theta}^s_{jt} - \hat{\Gamma}X - \hat{\beta}^b\hat{\theta}^b_{jt} - \hat{\beta}^l\hat{\theta}^l_{jt}$. To interpret the estimates in equation (5) as causal, we must assume $\text{cov}(\tilde{\theta}^s_{jt}, \eta_{ijt}) = \text{cov}(\tilde{\theta}^b_{jt}, \eta_{ijt}) = \text{cov}(\tilde{\theta}^l_{jt}, \eta_{ijt}) = 0$; the residualized leave-year-out measures of teacher value-added and student unobservables that affect the outcome, y_i, are uncorrelated. Although a strong assumption, it allows for some sorting on unobservables. For example, suppose y_i is an individual's SAT score and η_{ijt} is a scalar that captures the effect of an SAT preparation course (if any) on an SAT score that is orthogonal to X. People who have unexplained variation in SAT preparation courses in high school could systematically sort into the classrooms of particular types of elementary school teachers without biasing the results, as long as they do not systematically sort into higher value-added elementary school teachers. To help alleviate some of the concerns with this assumption, in section 5.6, we examine the effect of teacher value-added on predicted outcomes as a placebo test and perform a quasi-experiment analysis that uses teachers switching grades and schools.

In addition, there are reasons to believe that this approach is conservative. First, we find somewhat larger, although much less precisely estimated, effects using a quasi-experimental design that uses variation in teachers switching between grades and schools. Second, we estimate smaller effects than if we use the approach taken by Chetty, Friedman, and Rockoff (2014b).

We also extend this analysis in two ways. First, we examine the dynamic effects of a teacher on student outcomes for years $\tau \in [0, 1, .., 7]$:

$$y_{i(t+\tau)} = \beta^s\hat{\theta}^s_{jt} + \beta^b\hat{\theta}^b_{jt} + \beta^l\hat{\theta}^l_{jt} + \Gamma X_{ijt} + \eta_{ijt}. \tag{6}$$

The model shows the extent to which the effect of teacher value-added on student outcomes persists or fades over a number of years. Second, we assess the effects on long-term outcomes by grade to see in which grades high-quality teachers have the most impact on students.

5 Results

5.1 Descriptive Results

5.1.1 Descriptive Relationships in Student Data

To assess whether multiple dimensions of teacher quality might matter for long-term outcomes, we estimate the relationship between measures of student achievement, both with each other and with long-term outcomes. Appendix Table A.1 shows bivariate correlations between each of the measures of student achievement. English and math test scores are highly correlated. The relationship between test scores and students' GPA, learning-skills GPA, and effort GPA is weaker but the correlation still ranges from 0.45 to 0.68. The correlations of each of these variables with attendance, days suspended, and being held back are substantially weaker, and suggests these variables largely capture different aspects of student achievement. These correlations suggest that test scores, behavior, and learning skills are related, but that some room remains for them to have an independent effect on long-term outcomes. Reducing the dimensionality of these variables by separately computing test-score, behavior (i.e., attendance, days suspended, being held back, and GPA), and learning-skills (i.e., learning-skills GPA and effort GPA) indices, as described in section 4.2, yields correlations between 0.46 and 0.55 (Table 2).

Next, we assess whether these measures of student achievement are related to long-term outcomes, conditional on the same set of controls we use to compute the value-added measures. English and math test scores, GPA, learning-skills GPA, suspensions, and log days absent measured in grades 3 through 5 typically have a statistically significant relationship with high school outcomes (Appendix Table A.2). After reducing the dimensionality to the three indices of student achievement, test scores, behavior, and learning skills measured in grades 3 through 5 nearly always have a statistically significant effect on high school outcomes (Table 3). For many of the high school outcomes, behavior and learning skills are as predictive of the outcome as test scores.

These results are consistent with test scores, behavior, and learning skills each independently affecting long-term outcomes. However, despite the fact that we control for a wide range of measures of student achievement, these estimates may be biased because of unobservables. Consequently, these results may not hold in situations in which there is exogenous variation in students' test scores, behavior, and learning-skills achievement. To address this concern, we move to a teacher value-added framework in which omitted variables are less likely to bias the results.

5.1.2 Descriptive Relationships in Teacher Value-Added Data

We compute teacher value-added as described in section 4.1. Appendix Table A.3 shows the relationship between the value-added measures. English and math test-score value-added are highly correlated. The correlations between test-score value-added and all other variables are much weaker, but they are positively correlated with GPA, effort GPA, and learning-skills GPA, which have correlations between 0.14 and 0.20. The GPA, effort GPA, and learning-skills GPA value-added are highly correlated with each other. The three value-added measures of student behavior – log absences, days suspended, and held back – are all weakly correlated with each other, test scores, and GPA measures. These correlations suggest that math and English test-score measures of teacher quality are closely related, as are GPA-based measures of teacher quality, whereas the ability to influence student behavior relates less closely. Table 4 shows similar results. The correlation between test-score value-added and behavior value-added is 0.15, the correlation of test-score value-added with learning-skills value-added is 0.17, and the correlation of behavior value-added with learning-skills value-added is 0.46.

5.2 Effects of Teacher Quality on Long-Term Outcomes

5.2.1 Single Value-Added Effects

Figures 2 through 4 show the effect of teachers' test-score, behavior, and learning-skills value-added individually on each high school outcome, conditional on the set of controls used to compute value-added measures. The plotted points show the relationship between the mean residualized outcome and the mean residualized value-added variables (with the unconditional mean of the outcome and value-added variables added back in) for 20 equally sized bins of teacher value-added measures. The coefficients and standard errors reported in the figures are from an OLS regression, using the micro data, of the outcome variable on the value-added variable, conditional on the same set of controls.

Figure 2 shows that students with better teachers in grades 3 to 5, as measured by the test-score value-added, score significantly higher on the SAT, have significantly higher achievement, effort, and cooperation GPAs, and score significantly higher on the high school exit exams. We find no significant effects on dropping out of high school, taking the SAT, being held back, log days absent, or being suspended. These results are consistent with the existing literature that shows benefits in adulthood from higher test-score value-added teachers, although research that demonstrates positive effects of elementary school teachers on high school outcomes is rare (Rothstein 2015).

Figure 3 shows the effect of teachers' behavior value-added on each outcome. We observe at least marginally statistically significant effects in the expected direction on all of the outcome variables except high school dropout. This indicates that in the absence of test-score value-added, having a teacher with a higher behavior valued-added impacts the high school outcomes of students in a meaningful way. Figure 4 shows the effect of teachers' learning-skills value-added on each outcome. We find less evidence of an effect than for the other two value-added measures, but the coefficient on the learning-skills value-added typically has the expected sign, and we find either marginally significant or significant effects for GPA, the English high school exit exam, days suspended, and absences. These results suggest that elementary school teachers affect students' long-term outcomes by improving student achievement as measured by both test-score and non-test-score data.

Comparing the magnitudes across the analyses, we tend to find that test-score value-added has a large effect on outcomes that have more cognitive content than the behavior or learning-skills value-added, whereas the pattern of results is reversed for outcomes that have more noncognitive content. For example, having a teacher with a standard deviation higher test-score value-added increases the math high school exit exam score by 0.023 standard deviations, whereas the increase for behavior value-added is 0.014 standard deviations, and the statistically insignificant increase for the learning-skills value-added is 0.004 standard deviations. However, the effect of having a teacher with a standard deviation higher test-score value-added on days suspended is less than 0.001, whereas behavior and learning-skills value-added both reduce days suspended by a statistically significant 0.003 days, a 2 percent decrease.

The test-score value-added estimates appear to have two sets of potential nonlinear effects. First, the effect of test-score value-added on all three GPA measures is positive until teachers become above average, and then the relationship is, if anything, negative. Second, there is suggestive evidence that the top vingtile or two of the test-score value-added distribution has a smaller effect on several outcomes than would be predicted from the rest of the test score value-added distribution (Figure 2). Chetty, Friedman, and Rockoff (2014b) find a similar anomaly in their data, and drop the top 1 percent of teachers because of evidence of "test manipulation." We leave those teachers in, although including them biases the effects of test-score value-added toward zero if "test manipulation" exists. We find less evidence of non-monotonicities for both behavior and learning-skills value-added, and outliers in the top vingtile are less common. One explanation for this finding is that because non-test-score value-added measures are constructed using student achievement in year $t+1$, teachers are unable to manipulate their non-test-score value-added measure unless they influence the actions of their students' teachers in the subsequent year.

The results shown in Figures 2 through 4 suggest that multiple components of teacher quality affect long-term outcomes. Our findings also indicate that in situations in which no test-score data are available, but other administrative data such as grades, attendance, suspensions, and held back are available, creating estimates of teacher quality that are associated with long-term benefits to students is possible. Some evidence also suggests non-test-score value-added measures calculated using our approach are less prone to manipulation by teachers, although they might begin to be manipulated if used in high-stakes settings.

5.2.2 Multi-Value-Added Effects

Now that we have found that each of the three dimensions of teacher quality affect high school outcomes, we can determine whether more than one value-added measure independently affects long-term outcomes. Significant effects of more than one value-added measure would suggest that teacher quality is multidimensional in a way that both matters for long-term outcomes and is measurable using a value-added approach. In addition, this analysis informs the extent to which the long-term effects of test score value-added measures are driven by teachers' affect on behavior and learning skills.

Table 5 shows the effect that each of the three elementary school value-added measures have on high school outcomes in an OLS regression in which all three value-added measures are included simultaneously along with the baseline controls (equation 5). Including behavior and learning-skills value-added only slightly affects the coefficients on the test-score value-added measures. For example, the coefficient in the SAT-score regression falls from 6.39 to 6.24 SAT points (or a constant 0.021 standard deviations), the coefficient in the math high school exit exam regression falls from 0.023 to 0.022 standard deviations, and the coefficient in the high school GPA regression falls from 0.004 to 0.002 GPA points. These results indicate that the long-term effects of test-score value-added are likely not driven by teachers' effects on students' behavior and learning skills that are correlated with test-score value-added.

The effects of behavior value-added on most outcomes is also not affected substantially by conditioning on the test-score and learning-skills value-added. Behavior value-added picks up a dimension of teacher quality that is largely unrelated to the other two value-added measures and that matters for long-term outcomes.

Adding the other value-added measures considerably weakens the evidence for an independent effect of teachers on long-term outcomes through learning skills. None of the coefficients on the learning-skills value-added in Table 5 are statistically significant with the expected sign. Effort GPA and the math graduation test score are negative and marginally statistically significant, and cooperation GPA is negative and statistically significant. These results suggest that test-score and behavior value-added capture the effect that teachers have on long-term outcomes through students' learning skills.

The size of these results can be interpreted using the cross-sectional relationship between test scores and earnings. Hanushek and Wossman (2008) find consistent evidence that in the cross section, a standard deviation increase in test scores at the end of high school increases earnings by 12 percent (Lazear 2003; Mulligan 1999; Murnane et al. 2000). Approximately the same relationship holds between elementary and middle school test scores and earnings. Chetty, Friedman, and Rockoff (2014b) show that direct estimates of a standard deviation improvement in teacher value-added on earnings, and a back of the envelope estimate using the cross-sectional relationship between primary school test scores and earnings, yield estimates of the effect of teachers on earnings in the 1.3 to 1.5 percent range. We find effects of approximately the same size or larger using the effect of teachers on contemporaneous test scores. If we instead use the effect of having a standard deviation higher test-score value-added teacher in elementary school on high school test scores, which ranges from 0.016 to 0.022 standard deviations, the estimated increase in earnings would be 0.23 percent. This much smaller effect is likely driven by the fade-out in the effect of teachers on test scores over time.

Combined, these results indicate that teacher quality is multidimensional in a way that matters for long-term outcomes. Importantly, this multidimensionality can be measured using a combination of test scores and other data that schools routinely collect.

5.3 Policy Implications of Multidimensional Teacher Quality

Policies that use teachers' test-score value-added to hire, fire, or incentivize teachers have been widely criticized because making decisions using only one (potentially gameable) dimension of teacher quality is considered unfair, or even counterproductive. However, the effect on long-term outcomes of having higher test-score and behavior value-added teachers implies that policies that shift the distribution of teacher quality upward in these dimensions benefit students. In comparison to just using test-score value-added, we show that using multiple dimensions of teacher quality in teacher-removal policies substantially improves the measurement of teacher quality and students' long-term outcomes.

Figure 5 shows scatter plots of teacher quality as measured by value-added in a given year. The dashed lines show the 5th percentile of teachers for a given value-added measure. The first panel plots test-score and behavior value-added, and shows that although both dimensions of teacher quality are positively correlated, the correlation is relatively weak, and some teachers who perform poorly as measured by test-score value-added perform well on the behavior value-added dimension. For example, the majority of teachers who are in the bottom 5 percent of teachers as measured by the test-score value-added are not in the bottom 5 percent of teachers as measured by the behavior value-added. Therefore, a linear combination of a teacher's value-added measures might be a better predictor of teacher quality and measure for teacher-removal policies.

One way to assess the value of using multiple measures of teacher quality is to ask to what extent students' long-term outcomes could be improved under a policy that replaces a school district's bottom 5 percent of teachers with average teachers as measured by only test-score value-added versus different linear combinations of the three value-added measures. Panel A of Table 6 shows the effect on a student's high school outcomes of being assigned an average teacher instead of a teacher in the bottom 5 percent of teachers as measured by a teacher's true value-added (realized value-added *ex post*, $\bar{v}_{jt}$). This panel shows the upper bound on the effects of the teacher-removal policy. The simulation uses estimated effects of teacher value-added on high school outcomes (Figures 2 through 4 and Table 5) and the within-teacher correlations between the three teacher value-added measures (Table 4). Standard errors for the estimated forecasts are shown in parentheses.

Each cell in row 1 shows the effect on students' high school outcomes if their bottom 5 percent test-score value-added elementary school teacher was replaced by an average teacher. For example, the students that the policy would affect (about 5 percent) would see their SAT scores increase by 13 points and their high school GPA by 0.008 GPA points. Row 2 shows the effect on students' high school outcomes if their bottom 5 percent behavior value-added elementary school teacher was replaced by an average teacher. The benefits from using behavior value-added are comparable to using test-score value-added, and in some cases, the benefits are larger.

Row 3 uses the average of teachers' test-score and behavior value-added. When this combined measure is used, students affected by the policy would see beneficial effects on all but one of the high school outcomes. Row 5 shows the percent improvement in students' outcomes if the replacement of the bottom 5 percent of teachers is performed using the average of teachers' test-score and behavior value-added instead of just teachers' test-score value-added. There is over a 100 percent increase in the beneficial effects on students for dropping out of high school, taking the SAT, GPA, effort GPA, days suspended, log absences, and held back. Importantly, these gains are accompanied by only small decreases in English exit exam test scores and SAT scores.

Row 4 uses a maximization procedure to choose the optimal weights to be placed on a linear combination of teachers' test-score, behavior, and learning-skills value-added to determine the bottom 5 percent of teachers for the indicated outcome variable. The optimal weights vary depending on the outcome variable, so simultaneously improving all outcomes by the calculated amount would not be possible. However, for most of the outcomes, a policy that uses the optimal weights for a particular outcome only slightly outperforms a simple policy that places equal weights on test-score and behavior value-added.

Panel B of Table 6 shows analogous results using teachers' estimated value-added based on the three previous years of student data. These results reflect the potential student gains if the teacher removal policy were to be implemented for teachers who had taught for three years. Similar to Panel A, student gains can be obtained if both test-score and behavior value-added are used to make the teacher-removal decision. Because the autocorrelation between years for the behavior measure is smaller than for the test-score measure (Appendix Figure A.2), the percent gain from using both value-added measures instead of just the test-score value-added is smaller.

These results suggest that the dimensions of teacher quality captured by behavior value-added are roughly as important for long-term outcomes as test-score value-added, and in combination, can improve student outcomes. Most of these benefits do not require new tests or assessments – only a new use for data that schools already collect.

5.4 Which Behaviors Matter for Long-Term Outcomes?

Behavior value-added includes several weakly correlated value-added measures, some of which may matter more than others for long-term outcomes. A straightforward way to assess which variables matter most is to regress high school outcomes on the full set of value-added measures that we use to construct the lower dimensional representation of teacher quality plus the usual set of controls. Focusing on the components of behavior value-added, we find suspensions and absences have significant effects with the expected sign for a number of outcomes, whereas held back only has a statistically significant effect on high school GPA (Appendix Table A.4). Interpreting the results for GPA is more difficult because GPA is highly correlated with the components of learning-skills value-added. We see some positive and significant effects for GPA, and negative and significant effects for the various components of learning-skills value-added, with a net effect of approximately zero. This suggests that the behavior value-added results are driven primarily by teachers' effects on suspensions and absences.

Another way to illustrate this is to move GPA value-added from behavior value-added to learning-skills value-added, and conduct the main analysis again. The new behavior value-added constructed only from absences, suspension, and being held back has a significant effect in the expected direction on all high school outcomes except high school dropout, SAT score, and the English high school exit exam (Appendix Table A.5). The new GPA-based value-added affects only taking the SAT and suspensions. The point estimates of the GPA-based value-added are often smaller than the significant effects of the other value-added measures, and have relatively tight confidence intervals, which suggests that the null effect for the GPA-based value-added is not simply due to a lack of power.

These results suggest multiple dimensions through which teachers affect long-term student outcomes, one that is closely related to improved performance on tests and others related to reduced absences and suspensions. The abilities reflected in achievement GPA, effort GPA, and learning-skills GPA matter for long-term outcomes, but the portion of these abilities that teachers are able to affect is largely captured by test scores and the ability to keep the students in the classroom.

5.5 Dynamic Effects of Teacher Quality

Figure 6 shows how the value-added measures affect a number of outcomes that can be tracked over time beginning in elementary school. The effect of test-score value-added on test scores shows the expected pattern of results. Having a standard deviation higher test-score value-added teacher has a large effect on math and English test scores in year zero that largely, but not completely, fades out over the next seven years.

Behavior value-added and learning-skills value-added show less evidence of fade out, but our approach to constructing these variables should result in measures with less fade out than test-score value-added. By measuring behavior and learning-skills value-added using the effect of a teacher this year on student achievement in the next year, we are effectively removing the first year of fade out from the estimates. In addition, because some of the student achievement variables are grades, and students may be graded on a curve, seeing little effect of behavior and learning-skills value-added on GPA measures in year zero would not be surprising.

5.6 Checking for Bias in Long-Term Effects

We conduct four analyses to look for evidence of bias in the estimates of the long-term effects of teachers. Consistent with Chetty, Friedman, and Rockoff (2014a), most tests show no evidence of bias, and the magnitude of the bias in the remaining tests is sufficiently small such that it does not substantially affect our conclusions.

First, we show that the value-added measures are forecast unbiased, specifically that each of the leave-year-out value-added variables cause an increase in the corresponding residualized achievement variable that is statistically indistinguishable from one (Appendix Table A.6). Only math test scores are marginally different than one, for which a one unit increase in the math test-score value-added causes a 0.99 increase in math test scores.

Second, we show that after conditioning on the main controls, students expected to perform better in elementary school based on their twice-lagged values of achievement are largely not sorting to higher value-added teachers. The estimated forecast bias from selection on student characteristics is between -1.6 to 1.3 percent, which is smaller than Chetty, Friedman, and Rockoff's (2014a) point estimate of 2.2 percent. The forecast bias is only statistically significant for GPA, and marginally significant for math test scores, with point estimates of 1.3 and 0.3 percent, respectively (Appendix Table A.6). An analogous calculation using predicted high school outcomes from the twice-lagged values of the control variables shows no evidence of upward bias. The only significant point estimates are for behavior value-added, but each suggests that better students are sorted to worse teachers (Table 7).

Finally, we aggregate these data to the school-grade-year level and estimate long-term effects using quasi-experimental variation in teacher value-added caused by teachers switching between grades or schools. The analysis removes variation in teacher value-added caused by students sorting to teachers within a grade. Following Chetty, Friedman, and Rockoff (2014b), we regress changes in school-grade-year high school outcomes on changes in the mean teacher value-added weighted by the number of students.[5] Appendix Table A.7 shows that the signs on the estimated coefficients are generally consistent with the main results in Table 5, and the point estimates tend to be larger. However, the estimates are much less precise. Despite this loss in statistical power, we observe a significant effect of test-score value-added on math exit exams, and either significant or marginally significant effects in the expected direction of behavior value-added for taking the SAT, the three GPA outcomes, and math exit exams. As in the main table, learning-skills value-added is often wrong-signed and statistically significant.

5.7 Robustness Checks

We conduct a number of robustness checks in which we use alternative approaches to constructing the value-added indices, different specifications to estimate the long-term effects, and additional high school outcomes. The results of the robustness checks are qualitatively consistent with the main results.

[5]The sample is limited to cases in which we have value-added measures for all teachers in a given school-grade year in two consecutive years, to the subset of students for which we have both the long-term outcome variable and a teacher value-added measure, and to value-added measures that can be computed leaving out both year t and $t-1$.

Appendix Table A.8 reports the effect of teacher value-added on additional high school outcomes such as graduating from the LAUSD if enrolled in the LAUSD in 12th grade, took the PSAT, PSAT score, 11th grade English CST score (the last grade the CST is administered), 11th grade math CST score, 8th grade science CST score, 10th grade science CST score, 8th grade social studies CST score, 11 grade social studies CST score, world history CST score, and the number of AP courses. We see significant effects of test-score value-added on all outcomes except LAUSD graduation and took the PSAT. Test-score value-added affects test-score outcomes by between 0.013 and 0.018 standard deviations, and does not vary noticeably by subject. Having a high test-score value-added teacher in elementary school improves long-term performance across a number of subjects, not just English and math. The coefficients on behavior value-added are typically the expected sign, and are either significant or marginally significant for 4 of the 11 outcomes, whereas the coefficients on the learning-skills value-added are typically wrong-signed and are marginally or statistically significant for 3 of the 11 outcomes. There is also little evidence that teacher value-added in grades 3 through 5 affects whether students leave the school district in subsequent years (Appendix Figure A.3).

We now show that the results are robust to a number of changes to our approach to computing the value-added indices and long-term effects. The main results are larger and more often statistically significant if we follow Chetty, Friedman, and Rockoff (2014b) when computing value-added measures by residualizing the achievement data using within-teacher variation in the controls (Appendix Table A.9). The effects are even larger if we use Chetty, Friedman, and Rockoff's (2014b) approach to computing long-term effects by residualizing the outcome variables using within-teacher variation in the controls and then regressing the residualized high school outcomes on teacher value-added with no controls (Appendix Table A.10). The main results are essentially unchanged if we use factor analysis to construct the three indices for teacher ability (Appendix Tables A.11 and A.12), and only somewhat weaker for the non-test score factor if we use exploratory factor analysis to construct two orthogonal factors (Appendix Tables A.13 and A.14). Finally, the results are largely robust to converting test score value-added into deciles and test score outcomes into percentiles, which means we use ordinal, rather than cardinal measures of teacher value-added and test score outcomes. However, we no longer observe effects of test-score value-added on SAT scores or either value-added variable on the GPA measures (Appendix Table A.16).

Lastly, Appendix Table A.15 shows that if we remove behavior value-added from the main analysis, learning-skills value-added no longer has any wrong-signed, statistically significant effects, and we observe negative and significant effects on days suspended and a negative and marginally significant effect on absences. This suggests that part of the reason for the unintuitive results for learning-skills value-added is that behavior value-added and learning-skills value-added are moderately correlated.

6 Applications of Non-Test-Score Measures

We take the general approach from the last section to demonstrate that non-test-score measures of achievement are useful for answering additional questions related to the effects of teachers and peers on students. We first examine teacher effects over the educational life cycle, and ask to what extent there could be long-term gains from moving high value-added teachers between grades. We then take the approach used to construct non-test-score value-added measures to compute GPA value-added for specific subjects in order to test the long-term value of having a better teacher in different subjects. This analysis could also be interpreted as suggesting how gains could be obtained from reallocating time between subjects. Finally, we estimate the effect of a student's peers on contemporaneous outcomes.

6.1 Long-Term Effects of Teacher Quality over the Educational Life Cycle

The approach in this analysis is to compute the test-score and behavior value-added for teachers in grades 3 through 12, and ask how having a standard deviation better teacher in each grade affects long-term outcomes as measured in 11th or 12th grade.[6] We do not compute learning-skills value-

[6]We cannot compute teacher value-added in 12th grade, so value-added measures for teachers in 12th grade use estimates of teacher quality in earlier grades.

added because we do not have learning-skills data for middle and high school students. Previous work on long-term teacher effects by grade has estimated the effects of test-score value-added for grades 4 through 8 (Chetty, Friedman, and Rockoff 2014b), and test-score and non-test-score value-added for grade 9 in North Carolina (Jackson 2016).

Figure 7 reports the results of this analysis for outcome variables measured as late as possible in a student's career. In each graph, we also report the sum of the coefficients across all grades. With some assumptions, particularly no tracking of students and no diminishing returns to having consecutive high quality teachers, this sum reflects the cumulative effect of having a standard deviation higher test-score or behavior value-added teacher in each grade from 3rd through 12th grade. If tracking students plays a large role, or if diminishing returns exist, this sum overestimates the cumulative effect. However, Chetty, Friedman, and Rockoff (2014b) find evidence for only a small amount of tracking.

We find that having a standard deviation higher test-score teacher in grades 3 through 12 has a beneficial effect on taking the SAT, SAT scores, and math and English test scores. The cumulative effect for each of these outcomes is quite large. Using the cross-sectional relationship between test scores and earnings, and the cumulative effects on math and English tests scores, having a standard deviation higher test-score value-added teacher in each grade increases a student's adult earnings by 2.7 to 5.2 percent.

We also find that having a standard deviation higher behavior value-added teacher in grades 3 through 12 has a large beneficial effect on dropping out of high school, graduation, taking the SAT, the three GPA measures, absences, suspensions, and grade retention. For example, the cumulative effects suggest that having a standard deviation higher behavior value-added teacher in each grade decreases the likelihood of dropping out of high school by 8.9 percentage points. Once adjusted for dropping out of high school being over estimated due to students leaving the LASUD, this effect is still a 5.9 percentage point decrease.

Since we have data on all grades 3 to 12, we can also examine whether having a high value-added elementary school teacher or high school teacher is more important. Models of human capital formation in which past human capital production is complementary to current human capital production suggests that having a high value-added elementary school teacher is more important. Alternatively, the substantial fade out we observe in the effect of teacher value-added suggests that high school teachers will have a larger effect on student outcomes.

The results generally show that having a high value-added English and math teacher in middle school or high school has a bigger impact on long-term outcomes than having a high value-added elementary school teacher. This pattern of results is especially clear for dropping out of high school, test scores, the GPA measures, absences, suspensions, and grade retention. For example, having a one standard deviation higher behavior value-added teacher has little effect on whether a student drops out of high school in grades 3 to 5, but reduces the likelihood of dropping out by about one percentage point per year in grades 6 to 12. Exceptions exist, notably for taking the SAT, but the pattern of results is fairly clear. The strength of the middle school and high school effects is somewhat surprising because we only calculate value-added for English and math teachers, with whom a student spends less than half of her school day, whereas in elementary school, the value-added measures are calculated using a classroom teacher with whom students spend much more time.

Focusing on 9th graders, which is the subpopulation analyzed by Jackson (2016) using North Carolina data, we observe significant effects of both value-added measures on being a high school dropout, graduation, and taking the SAT, and of behavior value-added on 12th grade GPA (Figure 7). Jackson (2016) finds significant effects of behavior on graduation and dropping out of high school, without clear evidence of effects on the other three outcomes. In comparison to Jackson (2016), we find evidence that teachers affect a broader range of outcomes, and more evidence of the importance of the abilities measured by test-score value-added.

6.2 Long-Term Effects of Teacher Quality by Subject

A significant shortcoming of the test-score value-added framework is that to measure a teacher's quality in a subject, a test must be administered in that subject. Consequently, we cannot evaluate teachers using value-added measures in subjects or grades that are untested, or compare the importance of high quality teachers across untested subjects. We extend the approach for calculating non-test-score value-added described in section 4.1 to compute value-added measures for elementary school teachers by subject using students' grades in each subject. We measure a teacher's quality using the grade each student receives in a subject in year $t+1$, controlling for the baseline controls from year $t-1$.

Appendix Table A.17 shows long-term student outcomes regressed on students' grades and the standard set of controls. Better grades in virtually all subjects improve students' long-term outcomes. Two exceptions are speaking and PE, which indicate that students who perform better in speaking or PE perform worse in high school, even conditional on their grades in other subjects and prior achievement. However, these estimates may not be detecting the true effect of ability in a particular subject, but unobserved characteristics associated with both elementary school grades and high school outcomes.

Table 8 reports the results after redoing this analysis using teacher value-added for each subject. Teachers who excel at teaching math, reading, writing, and health have students who perform better in high school. The effects for math-GPA value-added are somewhat subject specific. The effect is only positive and marginally significant for SAT scores, and significant for math high school exit exam scores. Unexpectedly, students with better elementary school math teachers are more likely to be held back in high school. The effects of reading-GPA value-added are more widespread, with significant or marginally significant effects on taking the SAT, SAT score, and both math and English high school exit exam scores. Writing-GPA value-added matters for the three high school GPA measures, and the probability of being held back, but has a negative effect on taking the SAT. Health-GPA value-added matters for taking the SAT, effort GPA, cooperation GPA, and both math and English high school exit exam scores.

Speaking-GPA value-added has negative effects on SAT scores, both high school exit exam measures, and all three GPA measures. Perhaps talking in class is not well rewarded in high school. We also observe negative and significant or marginally significant effects on science-GPA value-added for four outcomes. We find little evidence that social studies, PE, and arts affect outcomes in either direction.

These results broadly support the traditional view that math, reading, and writing are building-block skills that have long-term benefits. The health results are unexpected, suggesting that improved health knowledge at a young age could have long-term benefits, though this explanation should be interpreted with caution. Besides the negative effect of speaking and science, we find relatively little evidence of effects of other subjects. A plausible explanation for these results is that certain subjects, as taught in elementary schools, do not impart skills that have long-term value, and teachers who focus too heavily on them have students with poorer reading, writing, and math skills. These results suggest that elementary schools could potentially create long-term benefits for students by hiring and retaining strong math, reading, and writing teachers, and focusing more of their time on teaching those subjects.

6.3 Peer Effects

In addition to teachers, the test score, behavior, and learning-skills achievement of a student's peers may be influential. We test whether a student's peers influence her achievement by estimating the effect of having peers with varying levels of average test scores, behavior, and learning skills on a student's outcomes.

This section relates to the broader peer effects literature, which shows effects on primary and secondary school students from peer ability measured by test scores, peer characteristics such as race and gender, and peer behaviors such as substance use on both test-score and social outcomes (Sarcedote 2011). There is less work using direct measures of students' in-school behavior and noncognitive abilities, though some research uses suspensions (Figlio 2007), suspensions and absences (Liu 2010), retained students (Lavy, Passerman, Schlosser 2012), and variables associated with classroom behavior problems such as percent male (Lavy and Schlosser 2011) and students from violent homes (Carrell and Hoekstra 2010, 2016).

We use an approach closely related to the main analysis to compute the effect of peers on student outcomes. Let θ^s_{it}, θ^b_{it}, and θ^l_{it} be students' test-score, behavior, or learning-skills indices, computed using student data rather than teacher value-added measures. We measure the achievement of a student's peers using the leave-out mean value of an index in the previous year for all members of a student's current class. For example, the test-score index leave-out mean for student i in a class with N students is:

$$\bar{\theta}^s_{ijt} = \frac{1}{N-1}\sum_{n=1}^{N} \theta^s_{n(t-1)} 1[i \neq n]. \tag{7}$$

We then estimate the effect of peers on same-year outcomes using the following model:

$$y_{ijt} = \beta^s \bar{\theta}^s_{jt} + \beta^b \bar{\theta}^b_{jt} + \beta^l \bar{\theta}^l_{jt} + \Gamma Z_{ijt} + \delta_i + \varepsilon_{ijt}, \tag{8}$$

where Z_{ijt} is a vector of controls that includes a third-order polynomial of lagged measures of the student's math CST score, English CST score, GPA, effort GPA, work and study habits GPA, learning and social skills GPA, log days absent, an indicator for suspensions, an indicator for being held back, an indicator for English learner status, grade fixed effects, and year fixed effects. The model also includes a student fixed effect δ_i. The student fixed effects are critical because it is unlikely that the achievement of a student's peers is uncorrelated with a student's own achievement (Hanushek et al. 2003; Sacerdote 2011). After conditioning on student fixed effects, for the results to be interpreted as causal, we need year-to-year variation in the quality of a student's peers to be unrelated to the time varying unobservables that affect a student's performance in school.

Table 9 shows the estimated coefficients from equation (8). Having peers with higher lagged math and English standardized test scores improves student performance on those exams. Interestingly, stronger competition from peers significantly reduces achievement and effort GPAs, and increases absences. The positive effect on test scores and the negative effect on GPA are consistent with teachers curving their grades. Having better peers, as measured by behavior value-added, reduces the frequency of being held back and being absent. We observe only a marginally significant effect of having peers with better learning skills on effort GPA, and otherwise observe no evidence of a positive effect on student performance. These results indicate that test scores, and to a lesser extent behavior achievement of peers, matter for student outcomes, whereas we find little evidence that learning skills matter after conditioning on other variables.

7 Conclusion

The results demonstrate that teacher quality is multidimensional. We show that teachers' test-score value-added has significant effects on long-term outcomes, and that adding controls for behavior and learning-skills value-added does not influence the estimated effects. This finding indicates the long-term effects of having a high test-score value-added teacher may not be biased upward by omitting measures of behavior or learning-skills value-added.

We also find that a teacher value-added measure that combines the teacher valued-added effects on GPA, absences, suspensions, and grade retention affects many high school outcomes. These effects are similar in magnitude to test-score value-added. This second dimension of teacher quality is only weakly correlated with test-score value-added, and allows for substantial improvement in the measurement of teacher quality. For example, a policy that uses both dimensions and three years of data to identify the bottom 5 percent of teachers and replaces them with average teachers improves dropout rates, the likelihood of taking the SAT, GPA, effort GPA, absences, and being held back by over 50 percent versus a policy that uses only test-score value-added. Despite substantial gains in many areas, high school test scores experience only minimal declines.

We find that learning-skills value-added individually has significant effects on a number of high school outcomes. However, once we control for test-score and behavior value-added, the effects dissipate. These results suggest that although the teacher effect on learning skills matters for long-term outcomes, test-score and behavior value-added capture that measure of teacher quality.

We then demonstrate that this value-added framework can be extended to analyze effects by grade, all elementary school subjects, and peers. We find that high test-score and behavior value-added middle school and high school teachers have a greater effect on end of high school outcomes than elementary school teachers. We also show that teachers who are relatively better at teaching math, reading, writing, and health improve their elementary school students' high school outcomes, whereas teachers who are better at teaching speaking and science worsen them. Teaching these subjects may have long-term benefits for students, which suggests schools should focus on improving teaching quality in those areas. Finally, we find that having peers with higher test scores and better behavior improves contemporaneous student outcomes.

References

Bacher-Hicks, Andrew, Thomas J. Kane, and Douglas O. Staiger. 2014. “Validating Teacher Effect Estimates Using Changes in Teacher Assignments in Los Angeles.” National Bureau of Economic Research Working Paper No. 20657.

Blazar, David, and Matthew A. Kraft. 2016. “Teacher and Teaching Effects on Students’ Attitudes and Behaviors.” Harvard University Working Paper.

Carrell, Scott E., and James E. West. 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy*, 118(3): 409-432.

Carrell, Scott E., and Mark L. Hoekstra. 2010. “Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone’s Kids.” *American Economic Journal: Applied Economics*, 2(1): 211–228.

Carrell, Scott E., Mark L. Hoekstra, and Elira Kuka. 2016. “The Long-Run Effects of Disruptive Peers.” Working Paper.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014a. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review,* 104(9): 2593-2632.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014b. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review,* 104(9): 2633-2679.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2015. “Measuring the Impacts of Teachers: Response to Rothstein (2014).” CEPR Discussion Paper No. DP10768.

Figlio, David N. 2007. “Boys Named Sue: Disruptive Children and Their Peers.” *Education Finance and Policy,* 2(4): 376–94.

Figlio, David N., Morton O. Schapiro, and Kevin B. Soter. 2015. "Are Tenure Track Professors Better Teachers?" *The Review of Economics and Statistics*, 97(4): 715-724.

Fryer, Roland G. 2013. “Teacher Incentives and Student Achievement: Evidence from New York City Public Schools.” *Journal of Labor Economics,* 31(2): 373-407.

Gershenson, Seth. 2016. “Linking Teacher Quality, Student Attendance, and Student Achievement.” *Education Finance and Policy,* 11(2): 123-149.

Goldhaber, Dan, and Michael Hansen. 2010. “Assessing the Potential of Using Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions.” National Center for Analysis of Longitudinal Data in Education Research Working Paper No. 31.

Goodman, Sarena, and Lesley Turner. 2013. "The Design of Teacher Incentive Pay and Educational Outcomes: Evidence from the New York City Bonus Program." *Journal of Labor Economics,* 31(2): 409-420.

Gordon, Robert, Thomas J. Kane, and Douglas O. Staiger. 2006. “Identifying Effective Teachers using Performance on the Job.” Washington, DC: The Brookings Institution.

Hanushek, Eric A. 1971. “Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data.” *American Economic Review Papers and Proceedings,* 61(2): 280-288.

Hanushek, Eric A. 2011. “Valuing Teachers: How Much Is a Good Teacher Worth?” *Education Next,* 11(3): 41-45.

Hanushek, Eric A., and Ludger Woessmann. 2008. “The Role of Cognitive Skills in Economic Development.” *Journal of Economic Literature*, 46(3): 607- 668.

Hanushek, Eric A., and Steven G. Rivkin. 2010. “Generalizations about Using Value-Added Measures of Teacher Quality.” *American Economic Review Papers and Proceedings,* 100(2): 267-271.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. “Does Peer Ability Affect Student Achievement?” *Journal of Applied Economics,* 18(5): 527-544.

Heckman, James J., Rodrigo Pinto, and Peter A. Savelyev. 2013. “Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes.” *American Economic Review,* 103(6): 2052-2086.

Horn, Sandra P., and William L. Sanders. 1994. “The Tennessee Value-Added Assessment System (TVAAS): Mixed-Model Methodology in Education Assessment.” *Journal of Personnel Evaluation in Education,* 8(3): 299-311.

Jackson, C. Kirabo. 2016. “What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes.” National Bureau of Economic Research Working Paper No. 22226.

Jackson, C. Kirabo, Jonah E. Rockoff, and Douglas O. Staiger. 2014. "Teacher Effects and Teacher-Related Policies." *Annual Review of Econonomics,* 6(1): 801-825.

Jacob, Brian A., Lars Lefgren, and David P. Sims. 2010. "The Persistence of Teacher-Induced Learning." *Journal of Human Resources,* 45(4): 915-943.

Jennings, Jennifer L., and Thomas A. DiPrete. 2010. “Teacher Effects on Social and Behavioral Skills in Early Elementary School.” *Sociology of Education,* 83(2): 135-159.

Kane, Thomas J., and Douglas O. Staiger. 2008. “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation.” National Bureau of Economic Research Working Paper No. 14607.

Kane, Thomas J., Daniel F. McCaffrey, Trey Miller, and Douglas O. Staiger. 2013. "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment." In Research Paper. MET Project. Bill & Melinda Gates Foundation.

Kraft, Matthew A., and Sarah Grace. 2016. "Teaching for Tomorrow’s Economy? Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies." Brown University Working Paper.

Lavy, Victor, and Analia Schlosser. 2011. “Mechanisms and Impacts of Gender Peer Effects at School.” *American Economic Journal: Applied Economics,* 3(2): 1-33.

Lavy, Victor, Daniele Paserman, and Analia Schlosser. 2012. “Inside the Black Box of Ability Peer Effects: Evidence from Variation in Low Achievers in the Classroom.” *Economic Journal*, 122(559): 208-237.

Lazear, Edward P. 2003. “Teacher Incentives.” *Swedish Economic Policy Review,* 10(3): 179-

214.

Liu, Keke. 2010. “Peer Group Effects on Student Outcomes: Evidence from Randomized Lotteries.” PhD dissertation, Vanderbilt University.

Mihaly, Kata, Daniel F. McCaffrey, Douglas O. Staiger, and J. R. Lockwood. 2013. "A Composite Estimator of Effective Teaching." Seattle, WA: Bill & Melinda Gates Foundation.

Mulligan, Casey B. 1999. “Galton versus the Human Capital Approach to Inheritance.” *Journal of Political Economy,* 107(6): S184-224.

Murnane, Richard J., John B. Willett, Yves Duhaldeborde, and John H. Tyler. 2000. “How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings.” *Journal of Policy Analysis and Management,* 19(4): 547-568.

National Council on Teacher Quality. 2015. “2015 State Teacher Policy Yearbook: National Summary.” Technical Report.

Neal, Derek. 2011. “The Design of Performance Pay in Education.” In *Handbook of the Economics of Education* Volume 4, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 495-550. Amsterdam: North-Holland, Elsevier.

Rockoff, Jonah E. 2004. “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data.” *American Economic Review Papers and Proceedings,* 94(2): 247-252.

Rockoff, Jonah, E., and Cecilia Speroni. 2010. “Subjective and Objective Evaluations of Teacher Effectiveness.” *American Economic Review Papers & Proceedings,* 100(2): 261-266.

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics,* 125(1): 175-214.

Rothstein, Jesse. 2015. "Revisiting the Impacts of Teachers." Working Paper.

Ruzek, Erik A., Thurston Domina, AnneMarie M. Conley, Greg J. Duncan, and Stuart A. Karabenick. 2015. “Using Value-Added Models to Measure Teacher Effects on Students’ Motivation and Achievement.” *Journal of Early Adolescence,* 35(5-6): 852-882.

Sacerdote, Bruce. 2011. "Peer Effects in Education: How Might They Work, How Big Are

They and How Much Do We Know Thus Far?" *Handbook of the Economics of Education*, 3: 249-277.

Figure 1: LAUSD Elementary School Progress Report

PROGRESS REPORT

School Year:

Principal:

Teacher: **Room:**

Grade Level:

Birth Date:	Grade Reporting Period					
	1		2		3	
Academic Subjects	AC	EF	AC	EF	AC	EF
Reading						
ELD Reading						
Writing						
ELD Writing						
Listening						
ELD Listening						
Speaking						
ELD Speaking						
Mathematics						
History/Social Science						
Science						
Health Education						
Physical Education						
Arts						

ACHIEVEMENT SCORES	(ELD) ENGLISH LANGUAGE DEVELOPMENT SCORES	EFFORT SCORES
*Meets Standards		
4 = Advanced*	4 = Advanced Progress	4 = Strong
3 = Proficient*	3 = Average Progress	3 = Consistent
2 = Partially Proficient	2 = Partial Progress	2 = Inconsistent
1 = Not Proficient	1 = Limited Progress	1 = Poor

Work and Study Habits	Reporting Period		
	1	2	3
Makes good use of time			
Works independently			
Organizes materials			
Presents neat and careful work			
Completes homework on time			
Learning and Social Skills			
Follows directions and procedures			
Accepts and respects authority			
Cooperates well in a group situation			
Shows dependability			
Takes responsibility			
Exercises self-control			
Resolves conflicts appropriately			
Demonstrates appropriate social interaction with peers			
Demonstrates fairplay			

Student Assessment

Instructional Programs
Master Plan Program

ELD Level **Start Date** **End Date** **Grade Period**

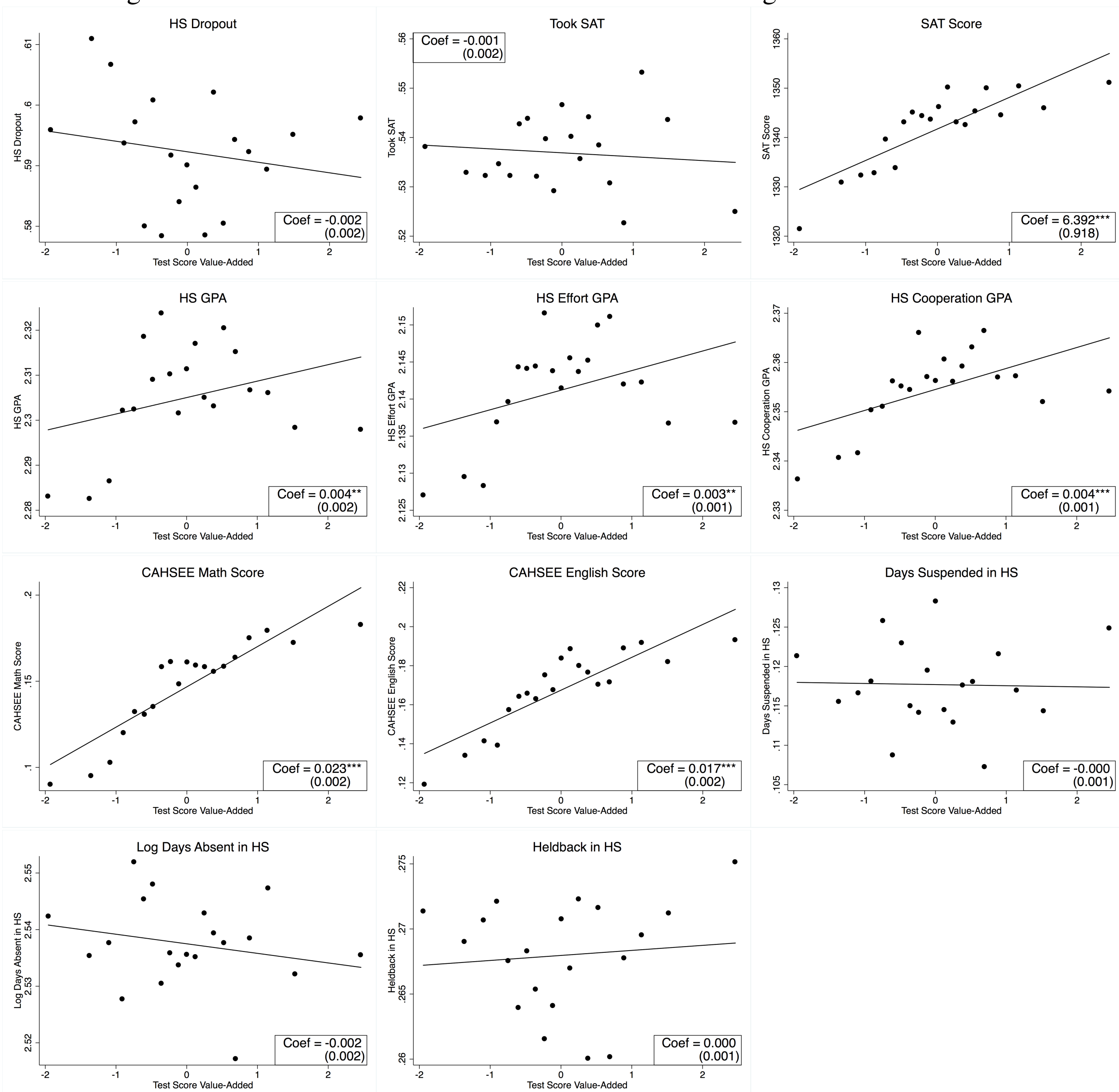
Instructional Services
Interventions

Intervention **Date**

Office Copy

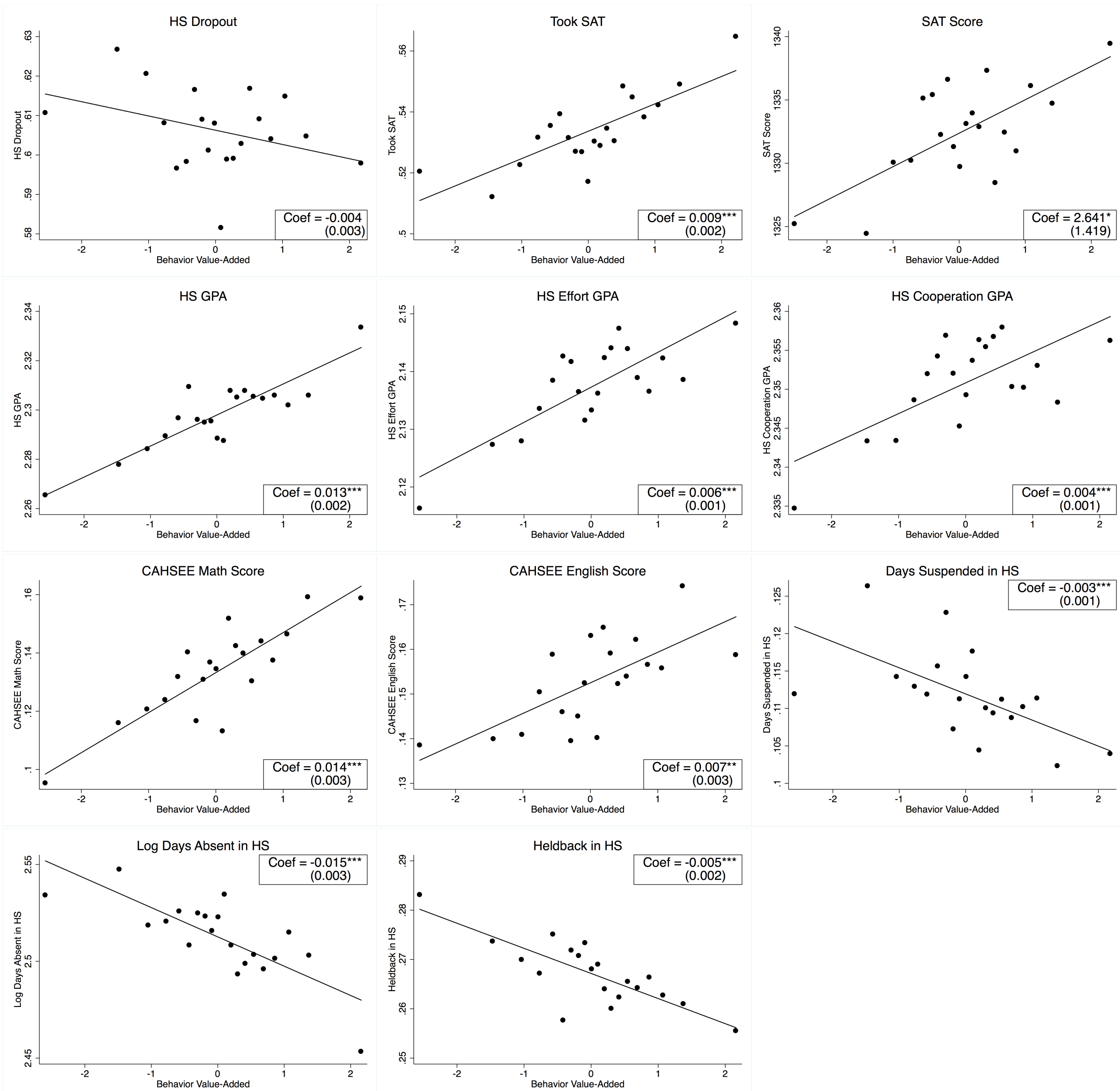
Note: The figure displays a blank copy of an LAUSD elementary school progress report. Each row labels the academic subject, work and study habits, or learning and social skill each student is graded on by their teacher. Columns 1, 2, and 3 correspond to each of the three trimesters students receive a grade. For the academic subjects, the AC column stands for achievement scores and the EF column stands for effort scores. For all academic subjects, work and study habits, and learning and social skills, students receive a grade ranging from 1 (the poorest performing) to 4 (the best performing).

Figure 2: Effect of Teacher Test-Score Value-Added on High School Outcomes



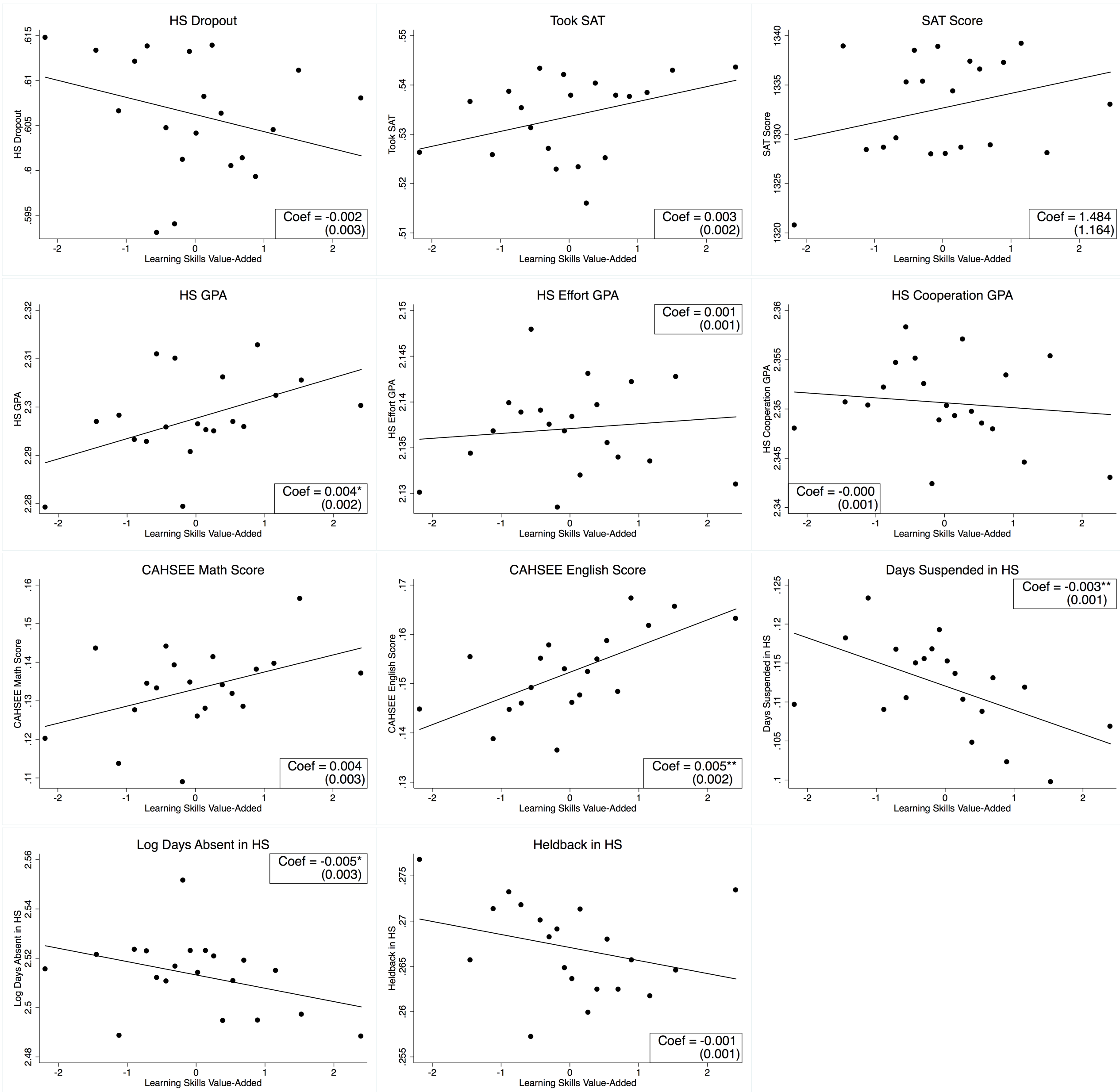
Note: Figure 2 shows binned scatter plots of residualized high school outcome variables and normalized teacher test-score value-added for grades 3-5. We construct these plots by first residualizing the outcome and teacher value-added variables using the controls shown in equation (5). We then plot the mean values of both variables in 20 equally sized bins. Lastly, we add back the unconditional mean of both variables. We also plot the best linear fit estimated prior to binning the data, and report its slope coefficient and standard error, clustered at the school-cohort level. $***$ $p < 0.01$, $**$ $p < 0.05$, and $*$ $p < 0.10$.

Figure 3: Effect of Teacher Behavior Value-Added on High School Outcomes



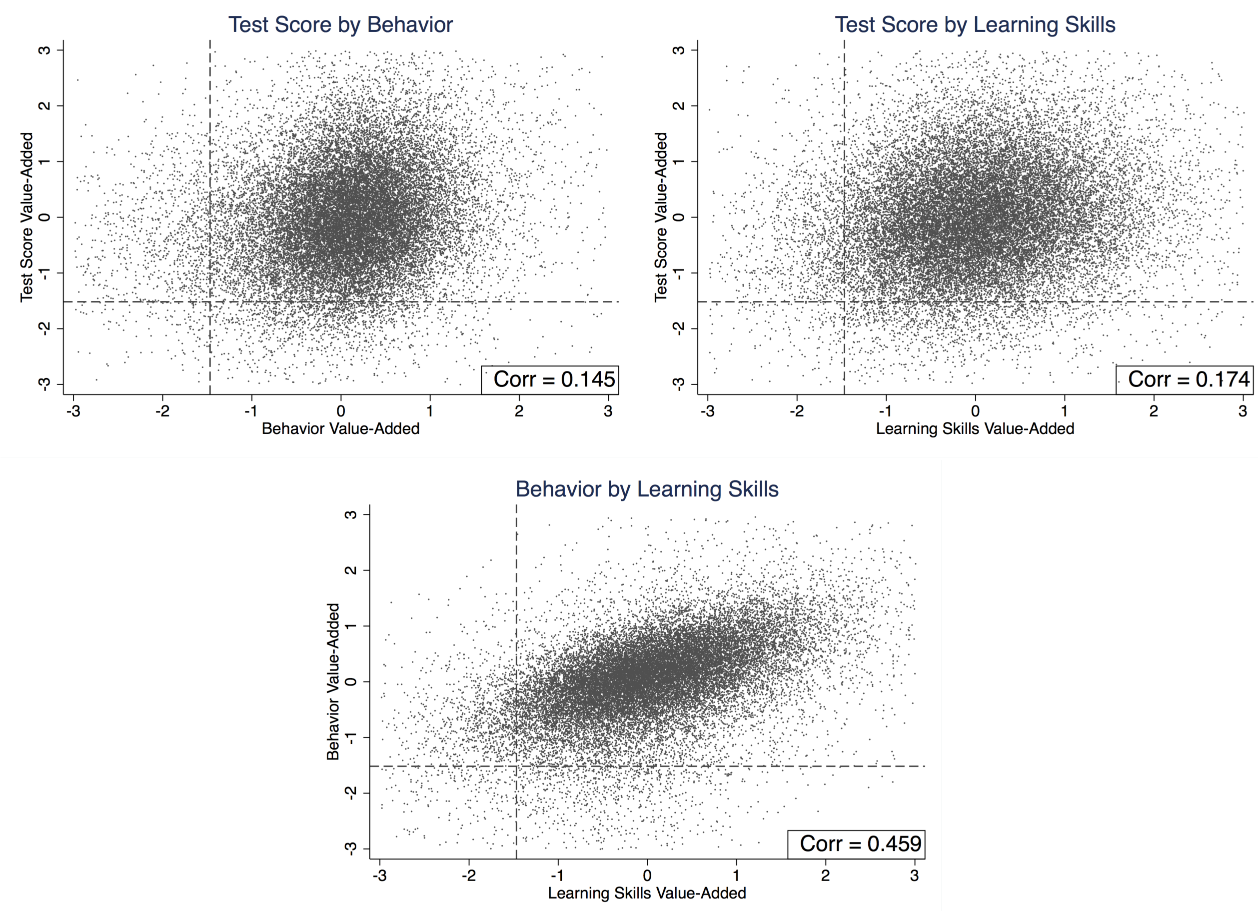
Note: Figure 3 shows binned scatter plots of residualized high school outcome variables and normalized teacher-behavior value-added for grades 3-5. We construct these plots by first residualizing the outcome and teacher value-added variables using the controls shown in equation (5). We then plot the mean values of both variables in 20 equally sized bins. Lastly, we add back the unconditional mean of both variables. We also plot the best linear fit estimated prior to binning the data, and report its slope coefficient and standard error, clustered at the school-cohort level. $*** p < 0.01$, $** p < 0.05$, and $* p < 0.10$.

Figure 4: Effect of Teacher Learning-Skills Value-Added on High School Outcomes



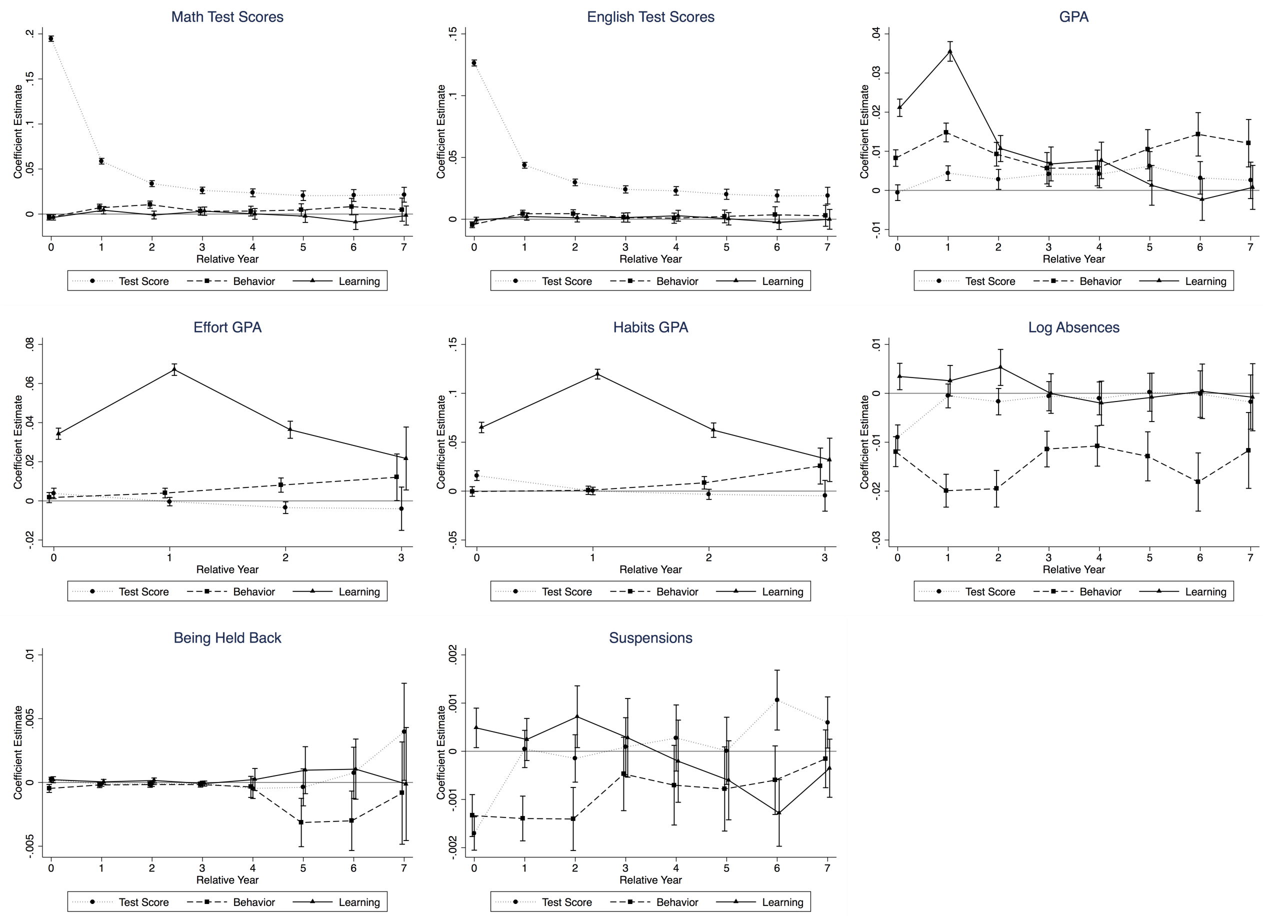
Note: Figure 4 shows binned scatter plots of residualized high school outcome variables and normalized teacher learning-skills value-added for grades 3-5. We construct these plots by first residualizing the outcome and teacher value-added variables using the controls shown in equation (5). We then plot the mean values of both variables in 20 equally sized bins. Lastly, we add back the unconditional mean of both variables. We also plot the best linear fit estimated prior to binning the data, and report its slope coefficient and standard error, clustered at the school-cohort level. $***$ p < 0.01, $**$ p < 0.05, and $*$ p < 0.10.

Figure 5: Two-Dimensional Cross Teacher Value-Added Plots



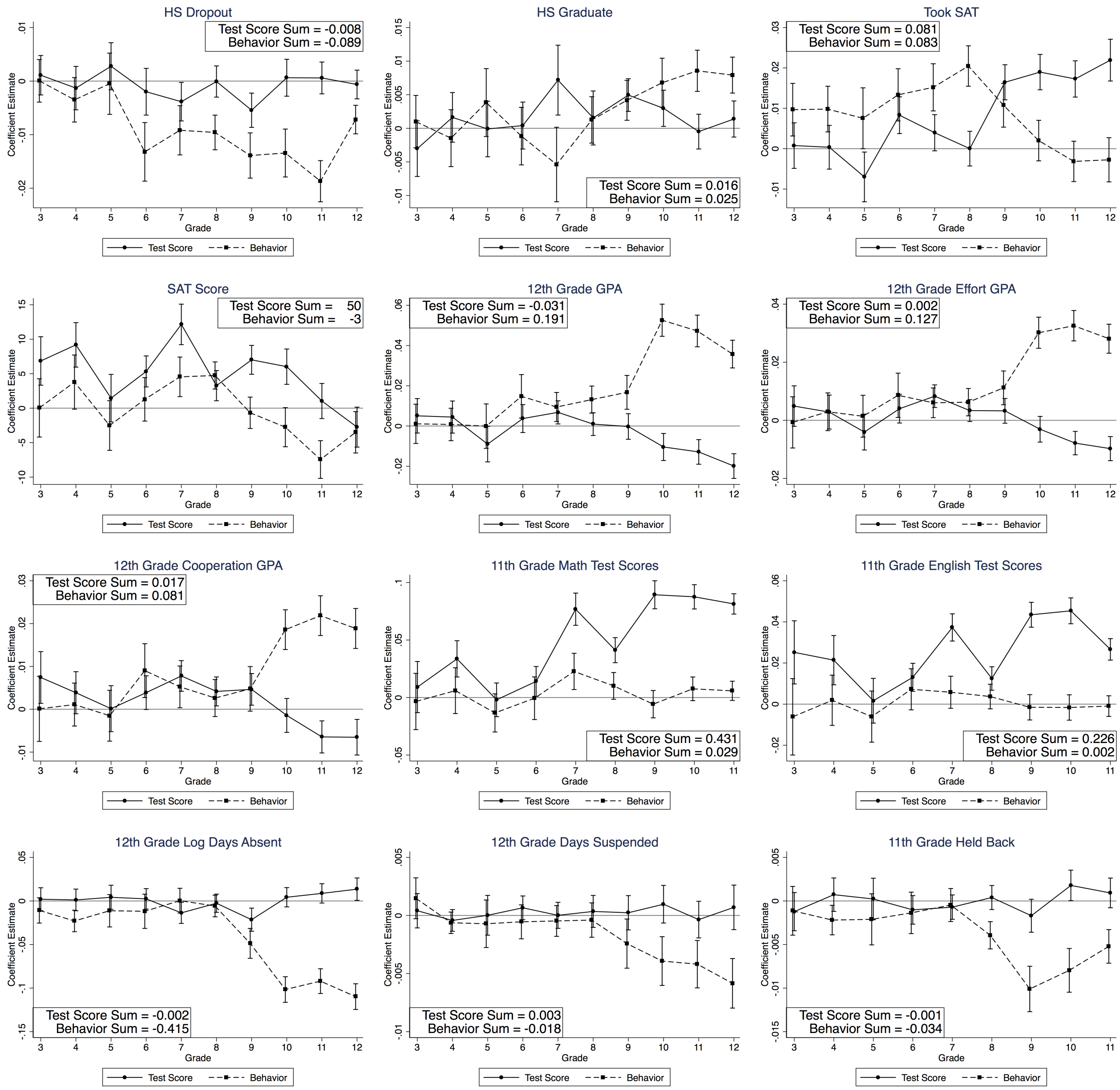
Note: The first scatter plot shows a plot of elementary school teachers' annual, normalized test-score and behavior value-added within three standard deviations of the mean. The two dashed lines show the cutoffs for the 5th percentile of the test-score and behavior teacher value-added, respectively. The second and third scatter plots are constructed analogously for test-score value-added versus learning-skills value-added, and behavior value-added versus learning-skills value-added, respectively.

Figure 6: Dynamic Effects of Test-Score, Behavior, and Learning-Skills Teacher Value-Added



Note: Each plot shows the effect of test-score, behavior, and learning-skills value-added of teachers in grades 3-5 on student outcomes in the concurrent year (the year a student was in a teacher's classroom) and future years (the years after a student was in a teacher's classroom). The estimated effects are obtained by regressing leads of outcome variables on teacher test-score, behavior, and learning-skills value-added and the baseline controls as specified in equation (6). The coefficients on test-score, behavior, and learning-skills value-added are plotted along with 95 percent confidence intervals, with standard errors clustered at the school-cohort level.

Figure 7: Effects of Test-Score and Behavior Teacher Value-Added by Grade



Note: The figure shows plots of the effect of test-score and behavior teacher value-added on high school outcomes by the grade level of the student. The plotted coefficients and standard errors (clustered at the school-cohort level) are from a regression of the high school outcome variable on test-score and behavior teacher value-added, and the vector of controls for high school students specified in section 4.1, estimated separately for each grade.

Table 1: Summary Statistics

Variable	Mean	Standard Deviation	Observations
Panel A: Grades 3 - 5			
Math CST Score	0.01	1.00	891,643
English CST Score	0.00	1.00	891,751
GPA	2.88	0.42	861,977
Effort GPA	3.14	0.46	861,977
Learning Skills GPA	3.10	0.58	614,532
Fraction of Days Absent	3.9%		858,308
Days Suspended	0.05	0.38	906,193
Held Back	0.7%		837,401
English Learner	42.0%		906,193
Panel B: High School Outcomes			
LAUSD Dropout	54.6%		333,513
Took SAT	50.5%		249,436
SAT Score	1,330	298	145,265
GPA	2.25	0.96	536,868
Effort GPA	2.12	0.52	476,548
Cooperation GPA	2.33	0.45	476,548
Math CAHSEE Score	0.07	1.01	331,266
English CAHSEE Score	0.08	0.98	329,980
Days Suspended	0.18	0.83	588,273
Fraction of Days Absent	7.8%		542,959
Held Back a Grade	29.0%		449,533
Graduated if Entered 12th Grade	88.6%		190,278
Took PSAT	68.9%		470,703
PSAT Score	1,110	248	348,992
Math CST Score	-0.05	0.99	124,044
English CST Score	0.00	0.99	135,769
Grade 8 Science CST Score	0.02	1.00	599,880
Grade 10 Science CST Score	0.09	1.00	296,069
Grade 8 Social Science CST Score	0.03	1.00	548,439
Grade 11 Social Science CST Score	0.07	0.99	160,483
World History CST Score	0.06	0.98	270,403
Number of AP Courses	0.73	1.70	588,273

Note: Panel A reports summary statistics for all LAUSD students in grades 3-5 from 2004 to 2010. Panel B reports high school summary statistics for all LAUSD students who were in grades 3-5 from 2004 to 2010 and attend high school in the LAUSD. Elementary school GPA, effort GPA, and learning-skills GPA are on a 4-point scale. GPA in high school is on a 4-point scale, and effort GPA and cooperation GPA in high school are on a 3-point scale. All test scores except the SAT and PSAT are normalized at the grade-year level. Both the SAT score and PSAT score are on a 600-2400 scale. The LAUSD Dropout variable is the fraction of students who enrolled in an LAUSD school in 9th grade and did not graduate from the LAUSD within five years. The Graduated if Entered 12th Grade variable shows the fraction of students who enrolled in an LAUSD school in 12th grade and graduated from the LAUSD.

Table 2: Correlation of Elementary School Student Achievement Measures

Measure	Test Scores	Behavior	Learning Skills
Test Scores	1		
Behavior	0.463	1	
Learning Skills	0.532	0.550	1

Note: Table 2 reports the correlations between the three measures of student achievement for grades 3-5. Each of the three measures of student achievement are equally weighted indices. The test-score index is computed using the students' normalized math and English test scores. The behavior index is computed using students' GPA, suspensions, log days absent, and not progressing to the next grade on time (held back). The learning-skills index is computed using students' effort GPA, work and study habit GPA, and learning and social skills GPA.

Table 3: Relationship between Elementary School Achievement and High School Outcomes

Measure	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Test Scores	-0.032***	0.091***	143.271***	0.208***	0.103***	0.089***	0.472***	0.398***	-0.009***	-0.112***	-0.052***
	(0.003)	(0.003)	(1.772)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.002)	(0.004)	(0.002)
Behavior	-0.020***	0.028***	10.826***	0.062***	0.032***	0.030***	0.028***	0.020***	-0.028***	-0.130***	-0.017***
	(0.002)	(0.002)	(1.328)	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)
Learning Skills	-0.040***	0.068***	-2.393*	0.228***	0.133***	0.121***	0.045***	0.064***	-0.045***	-0.062***	-0.061***
	(0.002)	(0.003)	(1.286)	(0.003)	(0.002)	(0.001)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
Observations	134,356	100,691	59,582	321,243	251,460	251,460	160,385	159,911	343,727	302,903	240,204
R-squared	0.421	0.179	0.686	0.321	0.315	0.326	0.572	0.577	0.045	0.281	0.134

Note: This table reports the predictive effect of a standard deviation increase in each of the three measures of student achievement (see Table 2) in grades 3-5 on high school student outcomes. Specifically, each column of the table reports the coefficients on each of the three achievement measures of students from an OLS regression of the high school student outcome on the students' three achievement measures in grades 3-5, along with the baseline controls described in section 4.1. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table 4: Correlation of Elementary School Teacher Value-Added Measures

Grades 3-5 VA	Test Score VA	Behavior VA	Learning Skills VA
Test Score VA	1		
Behavior VA	0.145	1	
Learning Skills VA	0.174	0.459	1

Note: Table 4 reports the correlations between the three measures of teacher value-added for grades 3-5. Each of the three measures are equally weighted indices. The test-score value-added is computed using teachers' value-added for math and English test scores. The behavior value-added is computed using teachers' value-added for GPA, suspensions, log days absent, and not progressing to the next grade on time (held back). The learning-skills value-added is computed using teachers' value-added for effort GPA, work and study habit GPA, and learning and social skills GPA.

Table 5: Effect of Elementary School Teacher Value-Added on High School Outcomes

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Test Score VA	-0.002	-0.002	6.237***	0.002	0.003**	0.005***	0.022***	0.016***	0.001	0.001	0.001
	(0.002)	(0.002)	(1.023)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)
Behavior VA	-0.003	0.010***	1.955	0.013***	0.007***	0.005***	0.013***	0.004	-0.003*	-0.016***	-0.006***
	(0.003)	(0.003)	(1.494)	(0.003)	(0.002)	(0.001)	(0.004)	(0.003)	(0.001)	(0.003)	(0.002)
Learning Skills VA	-0.000	-0.001	-0.547	-0.002	-0.003*	-0.003**	-0.005*	0.001	-0.002	0.002	0.001
	(0.003)	(0.002)	(1.173)	(0.003)	(0.002)	(0.001)	(0.003)	(0.003)	(0.001)	(0.003)	(0.002)
Observations	135,786	102,517	60,694	293,028	233,078	233,078	152,345	151,820	316,123	277,333	221,757
R-squared	0.293	0.145	0.617	0.244	0.234	0.239	0.500	0.512	0.040	0.267	0.108

Note: This table reports the effect of a standard deviation increase in the three measures of elementary school teacher value-added (see Table 4) on students' high school outcomes. Specifically, each column of the table reports the coefficients on each of the three normalized measures of teacher value-added from an OLS regression of the students' high school outcome on the three measures of teacher value-added for the students' teachers in grades 3-5, along with the baseline controls described in section 4.1. The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table 6: Effect of Replacing the Bottom 5 Percent of Elementary School Teachers on High School Outcomes

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Panel A: Using True Value-Added											
Test Score VA	-0.004	-0.002	13.198***	0.008**	0.005**	0.009***	0.048***	0.035***	-0.000	-0.003	0.001
	(0.004)	(0.003)	(1.895)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.002)	(0.005)	(0.002)
Behavior VA	-0.007	0.019***	5.446*	0.026***	0.013***	0.008***	0.028***	0.014**	-0.007***	-0.031***	-0.011***
	(0.006)	(0.004)	(2.926)	(0.005)	(0.003)	(0.003)	(0.007)	(0.005)	(0.003)	(0.007)	(0.003)
$\frac{1}{2}$(Test Score + Behavior)	-0.008	0.012***	12.723***	0.023***	0.013***	0.012***	0.052***	0.034***	-0.005*	-0.023***	-0.007**
	(0.006)	(0.004)	(2.812)	(0.005)	(0.003)	(0.003)	(0.007)	(0.005)	(0.003)	(0.007)	(0.003)
Optimal Three VA	-0.008	0.019***	13.981***	0.027***	0.014***	0.013***	0.054***	0.037***	-0.008**	-0.031***	-0.011***
Weighted Avg.	(0.008)	(0.005)	(2.565)	(0.006)	(0.003)	(0.002)	(0.006)	(0.005)	(0.003)	(0.008)	(0.004)
% Gain (Row 1 to Row 3)	128%	200%+	-4%	200%+	137%	39%	7%	-3%	200%+	200%+	200%+
Panel B: Using Previous Three Years of Student Data											
Test Score VA	-0.002	-0.001	7.695***	0.004**	0.003**	0.005***	0.028***	0.020***	-0.000	-0.002	0.000
	(0.002)	(0.002)	(1.105)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)	(0.003)	(0.001)
Behavior VA	-0.002	0.005***	1.529*	0.007***	0.004***	0.002***	0.008***	0.004**	-0.002***	-0.009***	-0.003***
	(0.002)	(0.001)	(0.821)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
$\frac{1}{2}$(Test Score + Behavior)	-0.003	0.003	6.661***	0.008***	0.005***	0.006***	0.025***	0.018***	-0.001	-0.006**	-0.002
	(0.003)	(0.002)	(1.168)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)	(0.003)	(0.001)
Optimal Three VA	-0.003	0.005***	7.696***	0.008***	0.005***	0.006***	0.028***	0.021***	-0.003**	-0.009***	-0.003***
Weighted Avg.	(0.003)	(0.001)	(1.292)	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)	(0.001)	(0.002)	(0.001)
% Gain (Row 1 to Row 3)	53%	200%+	-13%	79%	52%	9%	-10%	-13%	200%+	200%+	200%+

Note: Panel A shows the effect on a student's high school outcomes of being assigned an average teacher instead of a teacher in the bottom 5 percent as measured by the teacher value-added variable indicated in each row. The simulation uses the estimated effects of teacher value-added measures on high school outcomes (from Figures 2-4 and Table 5), the within-teacher correlations between the three teacher valued-added measures (shown in Table 4), and assumes teacher value-added measures are normally distributed. Row 1 in Panel A uses a measure of teachers' test-score value-added to replace the bottom 5 percent of teachers. Therefore, each cell in row 1 shows the effect on a student's high school outcome (shown in the column) if she were to move from a teacher in the bottom 5 percent of test-score teacher value-added to an average teacher. Row 2 of Panel A shows the improvement in outcomes for a move from a teacher in the bottom 5 percent to an average teacher as measured by behavior value-added. Row 3 of Panel A uses the average of teachers' test-score value-added and their behavior value-added. Row 4 of Panel A uses a maximization procedure to choose the optimal weights to be place on teachers' test-score, behavior, and learning-skills value-added to determine the bottom 5 percent of teachers for the indicated outcome variable. Row 5 of Panel A shows the percent improvement in students' outcomes if the replacement of the bottom 5 percent of teachers is performed by using the average of teachers' test-score and behavior value-added instead of just teachers' test-score value-added. Panel B shows analogous results using teachers' value-added based on only the three previous years of student data. These valued-added measures are estimated using three prior years of data on each teacher, along with the autocorrelations in teachers' value-added across years shown in Figure A.2. The standard errors on the estimated forecast are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table 7: Effect of Elementary School Teacher Value-Added on Predicted High School Outcomes

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Test Score VA	0.000	0.000	0.429	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.001)	(0.881)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)	(0.001)	(0.000)
Behavior VA	0.001**	-0.001*	-2.021**	-0.003***	-0.002**	-0.001**	-0.006***	-0.006***	0.000**	0.002**	0.001**
	(0.000)	(0.001)	(0.988)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)	(0.001)	(0.000)
Learning Skills VA	0.000	0.000	0.225	0.000	0.000	0.000	-0.001	0.001	0.000	0.001	0.000
	(0.000)	(0.001)	(1.050)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)	(0.001)	(0.000)
Observations	66,354	50,137	29,269	158,327	123,347	123,347	78,651	78,434	169,929	149,194	117,763
R-squared	0.962	0.709	0.750	0.681	0.666	0.667	0.701	0.750	0.767	0.833	0.737

Note: This table reports the effect of a standard-deviation increase in the three measures of elementary teacher value-added (see Table 4) on students' predicted high school outcomes using double lagged student achievement. The predicted outcomes are created by estimating an OLS regression of the high school outcome indicated in each column on third-order polynomials of double lagged (year $t-2$) math and English test scores, GPA, effort GPA, learning skills GPA, log absences, an indicator for suspension, an indicator for being held back, and an indicator of English learner status. The coefficients obtained from this OLS regression are then used to predict students' high school outcomes. Each column of the table reports the coefficients on each of the three normalized measures of teacher value-added from an OLS regression of the students' predicted high school outcome on the three measures of teacher value-added for the students' teachers in grades 3-5, along with the baseline controls described in section 4.1. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table 8: Effect of Elementary School Teacher Subject Emphasis on High School Outcomes

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Math GPA	0.004	-0.006	5.997*	-0.005	-0.002	0.001	0.021**	0.009	-0.000	0.001	0.015***
	(0.007)	(0.005)	(3.191)	(0.006)	(0.004)	(0.003)	(0.008)	(0.007)	(0.003)	(0.007)	(0.003)
Reading GPA	-0.011	0.029***	9.830***	0.007	0.006	0.005	0.030***	0.028***	0.002	0.006	-0.005
	(0.009)	(0.006)	(3.748)	(0.007)	(0.004)	(0.004)	(0.009)	(0.007)	(0.003)	(0.009)	(0.004)
Writing GPA	-0.001	-0.012**	2.202	0.016**	0.008**	0.007*	-0.003	0.002	0.002	-0.005	-0.012***
	(0.009)	(0.006)	(3.508)	(0.007)	(0.004)	(0.004)	(0.009)	(0.008)	(0.003)	(0.009)	(0.004)
Listening GPA	-0.001	-0.010	4.579	0.003	0.002	0.004	0.017*	0.011	-0.003	-0.013	0.001
	(0.009)	(0.006)	(3.346)	(0.007)	(0.004)	(0.004)	(0.009)	(0.008)	(0.004)	(0.008)	(0.004)
Speaking GPA	0.014	-0.003	-11.322***	-0.025***	-0.013***	-0.014***	-0.028***	-0.025***	0.001	0.009	0.001
	(0.009)	(0.006)	(3.410)	(0.006)	(0.004)	(0.004)	(0.009)	(0.008)	(0.004)	(0.008)	(0.004)
History/Social Science GPA	-0.000	-0.003	1.483	0.009	0.008**	0.008**	-0.002	0.001	-0.001	0.006	-0.006
	(0.008)	(0.006)	(3.420)	(0.006)	(0.004)	(0.004)	(0.008)	(0.007)	(0.003)	(0.008)	(0.004)
Science GPA	-0.001	0.000	-6.470*	-0.000	-0.005	-0.009**	-0.024***	-0.013*	-0.003	-0.010	-0.003
	(0.009)	(0.006)	(3.516)	(0.006)	(0.004)	(0.004)	(0.008)	(0.007)	(0.003)	(0.008)	(0.004)
Health Education GPA	0.003	0.015***	5.572	0.007	0.008**	0.012***	0.023***	0.019***	-0.002	-0.001	0.002
	(0.008)	(0.006)	(3.462)	(0.006)	(0.004)	(0.003)	(0.008)	(0.007)	(0.003)	(0.008)	(0.004)
PE GPA	-0.007	0.011**	-0.223	0.003	-0.005	-0.011***	-0.008	0.000	-0.003	-0.012	-0.002
	(0.008)	(0.005)	(3.339)	(0.006)	(0.004)	(0.003)	(0.007)	(0.007)	(0.003)	(0.008)	(0.004)
Arts GPA	-0.002	0.000	-4.331	-0.004	-0.005	-0.007**	-0.009	-0.015**	0.003	0.007	0.003
	(0.008)	(0.005)	(3.454)	(0.006)	(0.004)	(0.003)	(0.008)	(0.006)	(0.003)	(0.008)	(0.004)
Observations	136,125	102,822	60,875	293,576	233,529	233,529	152,693	152,162	316,747	277,860	222,174
R-squared	0.293	0.146	0.617	0.244	0.234	0.240	0.500	0.512	0.039	0.266	0.108

Note: This table reports the effect of a standard deviation increase in the 10 subject elementary school teacher value-added measures on students' high school outcomes. Specifically, each column of the table reports the coefficients on each of the 10 subject teacher value-added measures from an OLS regression of the students' high school outcome on the 10 subject teacher value-added measures for the students' teachers in grades 3 through 5, along with the baseline controls described in section 4.1. The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table 9: Effect of Elementary School Peers on End-of-Year Outcomes

Peer Achievement	GPA	Effort GPA	Learning Skills GPA	Math CST	English CST	Days Suspended	Log Absences	Held Back
Peer Test Scores	-0.015***	-0.011***	0.003	0.025***	0.027***	-0.001	0.008***	0.000
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.001)	(0.003)	(0.000)
Peer Behavior	0.005	-0.000	0.000	-0.003	-0.004	-0.002	-0.011**	-0.001**
	(0.003)	(0.004)	(0.004)	(0.006)	(0.004)	(0.002)	(0.005)	(0.000)
Peer Learning Skills	0.003	0.006*	0.004	-0.002	-0.002	0.000	-0.002	0.000
	(0.003)	(0.003)	(0.003)	(0.005)	(0.004)	(0.001)	(0.004)	(0.000)
Observations	811,370	811,370	816,869	710,145	710,362	836,744	820,457	754,627
R-squared	0.843	0.775	0.825	0.878	0.907	0.654	0.806	0.757

Note: This table reports the effect of being in a classroom with a standard deviation higher peer group in the three measures of student achievement (see Table 2) on students' contemporaneous academic outcomes. Peer achievement is measured using a leave-out mean of peers' lagged test score, behavior, and learning skills (as shown in equation (7)) normalized to have a mean of zero and a standard deviation of one. Each column of the table reports the coefficients on each of the three normalized leave-out means of peers' lagged measures of achievement from an OLS regression of the students' contemporaneous outcome on the three normalized leave-out means, along with the baseline controls described in section 6.3 and a student fixed effect. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Appendix

Figure A.1: Secondary School Criteria for Marks

CRITERIA FOR MARKS

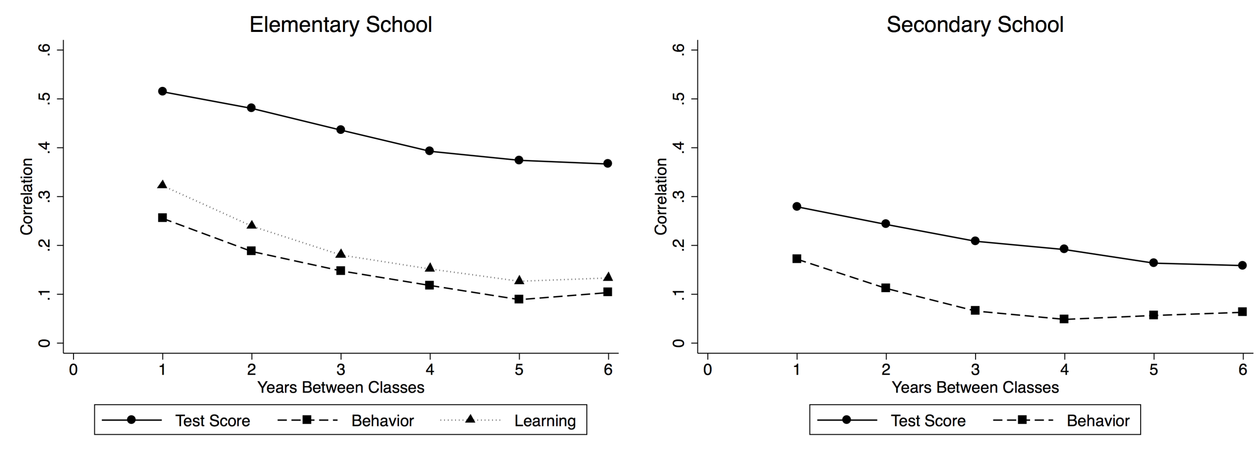
Academic Mark	A	B	C	D	FAIL
Quality of Work	Demonstrates an exemplary level of understanding of content standards and tasks.	Demonstrates a thorough understanding of the content standards and tasks.	Demonstrates an understanding of the content standards and tasks.	Demonstrates a limited understanding of the content standards and tasks.	Demonstrates an inability to understand the content standards and tasks.
Interpretation and Application	Demonstrates exceptional and fluent skills in analyzing, synthesizing, and drawing inferences from observations and other data or information.	Demonstrates fluent skills in analyzing, synthesizing, and drawing inferences from observations and other data or information.	Demonstrates satisfactory skills in analyzing, synthesizing, and drawing inferences from observations and data or information.	Demonstrates a limited ability to analyze, synthesize, and draw inferences from observations and other data or information.	Demonstrates an incomplete and/or inaccurate analysis of data or information that has been collected.
Thinking and Reasoning Skills	Demonstrates an insightful and thorough use of prior knowledge and skills to create innovative ideas, products or performances in a variety of contexts.	Demonstrates an insightful use of prior knowledge and skills to create innovative ideas, products or performances in a variety of contexts.	Demonstrates use of prior knowledge and skills to create innovative ideas, products or performances in a variety of contexts.	Demonstrates limited use of prior knowledge and skills to create innovative ideas, products or performances.	Demonstrates incomplete use of prior knowledge/skills to create innovative ideas, products or performances.
Quantity of Work	Produces extra work in addition to assigned work, of both teacher-generated and self-initiated toward achieving standards for the course.	Produces extra work in addition to all assigned work, usually teacher-generated and self-initiated toward achieving standards for the course.	Produces the assigned work in achieving standards for the course.	Demonstrates a need to improve in the amount of work completed and effort expended toward achieving standards for the course.	Demonstrates no improvement of the work completed and in the effort expended toward achieving standards for the course.

WORK HABITS	E	S	U
Effort	Demonstrates exceptional determination in accomplishing tasks and mastering standards.	Demonstrates determination in accomplishing tasks and mastering standards.	Demonstrates little determination in accomplishing tasks and mastering standards.
Responsibility	Accepts complete responsibility for personal actions and demonstrates honesty, fairness, and integrity.	Accepts responsibility for personal actions and frequently demonstrates honesty, fairness, and integrity.	Accepts little responsibility for personal actions.
Attendance	Maintains excellent attendance record by consistently avoiding unnecessary absences or tardies.	Maintains a satisfactory attendance record by avoiding unnecessary absences or tardies.	Makes little effort to maintain a satisfactory attendance record; is frequently absent or tardy without excuses.
Evaluation	Makes explicit effort to examine work using both teacher-generated and self-generated criteria.	Makes effort to examine work using teacher-generated criteria.	Makes use only of teacher-generated criteria to examine work on an inconsistent basis.

COOPERATION	E	S	U
Courtesy	Maintains courteous relations with the teacher and other students and consistently works without disturbing others.	Demonstrates courteous relations with the teacher and other students and generally works without disturbing others.	Demonstrates discourteous behavior towards the teacher and other students and consistently lacks consideration for others.
Conduct	Obeys rules, respects public and personal property and actively promotes the general welfare.	Obeys rules, respects public and personal property and supports the general welfare.	Shows disregard for rules; has little respect for public and personal property and often opposes the general welfare.
Improvement	Assumes responsibility for personal improvement and rarely needs correction.	Tries to improve and usually accepts corrections in an objective manner.	Makes little attempt to improve and shows indifference or resistance to corrections.
Class Relations	Demonstrates leadership ability to work with others in a variety of situations to set and achieve goals.	Demonstrates ability to work with others in a variety of situations to set and achieve goals.	Demonstrates little ability to work with others in a variety of situations to set and achieve goals.

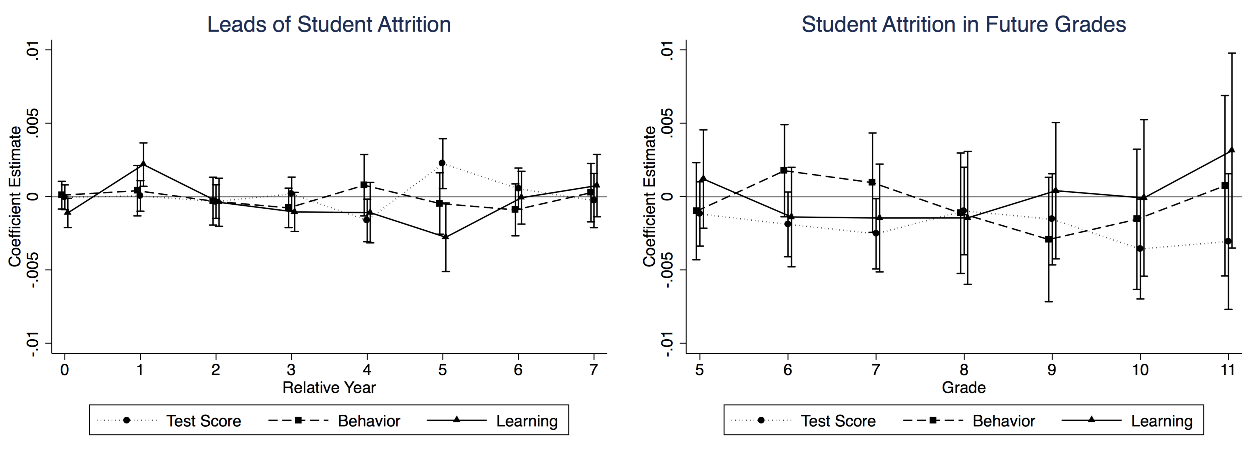
Note: The figure shows LAUSD's secondary school teacher guidelines for giving achievement, effort, and cooperation grades.

Figure A.2: Autocorrelations of Teacher Value-Added Across Years



Note: Each graph shows the correlation between mean test-score residuals, using the baseline controls, across classes taught by the same teacher in different years. The first graph plots the autocorrections for elementary school and the second graph for secondary school.

Figure A.3: Effects of Teacher Value-Added on Student Attrition



Note: The figure shows plots of the effect of a teacher's test-score, behavior, and learning-skills value-added in grades 3-5 on leads of student attrition (the first panel) and attrition in future grades (the second panel). In the first panel the plotted coefficients and standard errors (clustered at the school-cohort level) are from a regression of the leads of attrition out of the sample on test-score, behavior, and learning skills teacher value-added and the vector of controls specified in section 4.1, estimated separately for each lead. In the second panel the plotted coefficients and standard errors (clustered at the school-cohort level) are from a regression of student attrition in a particular grade on test-score, behavior, and learning skills teacher value-added and the vector of controls specified in section 4.1, estimated separately for each grade.

Table A.1: Correlation of All Student Achievement Measures

Measures	Math Test Scores	English Test Scores	GPA	Learning Skills GPA	Effort GPA	Log Absences	Days Suspended	Held Back
Math Test Scores	1							
English Test Scores	0.762	1						
GPA	0.636	0.676	1					
Learning Skills GPA	0.454	0.454	0.688	1				
Effort GPA	0.527	0.539	0.828	0.774	1			
Log Absences	-0.186	-0.110	-0.170	-0.181	-0.176	1		
Days Suspended	-0.081	-0.076	-0.098	-0.184	-0.116	0.076	1	
Held Back	-0.072	-0.077	-0.106	-0.069	-0.082	0.018	0.010	1

Note: Table A.1 reports the correlations between each measure of grades 3-5 student achievement.

Table A.2: Relationship between Elementary School Achievement and High School Outcomes for All Achievement Measures

Measures	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Math Test Scores	-0.023***	0.049***	58.139***	0.131***	0.065***	0.054***	0.364***	0.073***	-0.005***	-0.087***	-0.036***
	(0.002)	(0.003)	(1.489)	(0.003)	(0.002)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)	(0.002)
English Test Scores	-0.013***	0.049***	94.959***	0.094***	0.050***	0.049***	0.101***	0.371***	-0.013***	-0.027***	-0.020***
	(0.002)	(0.003)	(1.715)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.002)	(0.004)	(0.002)
GPA	0.002	0.012***	25.921***	0.006	-0.008***	-0.017***	0.062***	0.080***	0.010***	0.003	0.003
	(0.003)	(0.004)	(1.937)	(0.004)	(0.003)	(0.002)	(0.005)	(0.004)	(0.002)	(0.005)	(0.002)
Learning Skills GPA	-0.044***	0.065***	-12.383***	0.236***	0.144***	0.147***	0.025***	0.033***	-0.070***	-0.101***	-0.064***
	(0.003)	(0.003)	(1.646)	(0.003)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)	(0.004)	(0.002)
Effort GPA	0.002	0.002	-0.817	-0.007*	-0.006***	-0.018***	-0.005	-0.009**	0.020***	0.018***	0.000
	(0.003)	(0.004)	(1.808)	(0.004)	(0.002)	(0.002)	(0.005)	(0.004)	(0.002)	(0.005)	(0.002)
Suspended	0.043***	-0.045***	-17.708**	-0.130***	-0.070***	-0.099***	-0.024*	-0.054***	0.235***	0.075***	0.047***
	(0.007)	(0.011)	(6.972)	(0.011)	(0.006)	(0.006)	(0.012)	(0.012)	(0.017)	(0.014)	(0.007)
Log Absences	0.021***	-0.033***	-0.806	-0.092***	-0.051***	-0.041***	-0.028***	-0.002	0.004***	0.249***	0.027***
	(0.002)	(0.002)	(0.890)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)
Held Back	0.249***	0.117***	36.725**	0.069**	0.037**	0.025	0.289***	0.207***	-0.102***	-0.272***	-0.048**
	(0.024)	(0.040)	(17.136)	(0.028)	(0.017)	(0.016)	(0.038)	(0.034)	(0.013)	(0.039)	(0.019)
Observations	134,357	100,691	59,582	321,243	251,460	251,460	160,385	159,911	343,728	302,903	240,204
R-squared	0.422	0.180	0.689	0.324	0.320	0.334	0.579	0.587	0.049	0.298	0.136

Note: This table reports the predictive effect of each of the measures of student achievement (see Table A.1) in grades 3-5 on students' high school outcomes. Specifically, each column of the table reports the coefficients on each of the measures of student achievement from an OLS regression of the students' high school outcome on the students' measures in grades 3-5, along with the baseline controls described in section 4.1. Math and English test scores, GPA, learning-skills GPA, and effort GPA have each been normalized at the grade-year level. Suspended and Held Back are indicators for being suspended and being held back a grade, respectively. Log Absences is the log number of days absent. Standard errors clustered at the school-cohort level are reported in parentheses. $*** p < 0.01$, $** p < 0.05$, and $* p < 0.10$.

Table A.3: Correlation of All Elementary School Teacher Value-Added Measures

Grades 3-5 VA	Math Test Scores	English Test Scores	GPA	Learning Skills GPA	Effort GPA	Log Absences	Days Suspended	Held Back
Math Test Scores	1							
English Test Scores	0.761	1						
GPA	0.174	0.199	1					
Learning Skills GPA	0.145	0.158	0.683	1				
Effort GPA	0.142	0.154	0.767	0.760	1			
Log Absences	-0.064	-0.031	-0.122	-0.110	-0.111	1		
Days Suspended	-0.028	-0.033	-0.036	-0.078	-0.040	0.062	1	
Held Back	-0.023	-0.019	-0.016	-0.004	-0.011	0.010	0.006	1

Note: Table 4 reports the correlations between measures of grades 3-5 teacher value-added. Each measure of teacher value-added is created as described in section 4.1.

Table A.4: Effect of All Elementary School Teacher Value-Added Measures on High School Outcomes

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Math Test Scores VA	0.003	0.003	-8.234***	-0.014***	-0.008***	-0.006***	-0.000	-0.018***	0.002	0.011**	0.005**
	(0.005)	(0.004)	(2.095)	(0.004)	(0.002)	(0.002)	(0.005)	(0.004)	(0.002)	(0.005)	(0.002)
English Test Scores VA	-0.006	-0.007*	18.200***	0.019***	0.013***	0.013***	0.033***	0.043***	-0.001	-0.012**	-0.004
	(0.005)	(0.004)	(2.165)	(0.004)	(0.002)	(0.002)	(0.005)	(0.005)	(0.002)	(0.005)	(0.002)
GPA VA	0.001	0.024***	7.716**	0.003	0.001	0.002	0.008	0.016**	0.002	-0.002	-0.011***
	(0.008)	(0.006)	(3.717)	(0.006)	(0.004)	(0.004)	(0.008)	(0.007)	(0.003)	(0.008)	(0.004)
Learning Skills GPA VA	-0.007	-0.018***	-5.377	-0.012**	-0.007**	-0.003	-0.018**	-0.011*	0.000	-0.011	0.008**
	(0.009)	(0.006)	(3.314)	(0.006)	(0.004)	(0.003)	(0.008)	(0.006)	(0.003)	(0.008)	(0.004)
Effort GPA VA	0.003	0.002	-1.602	0.014**	0.005	-0.003	0.009	0.001	-0.007**	0.009	-0.001
	(0.010)	(0.007)	(3.982)	(0.007)	(0.004)	(0.004)	(0.009)	(0.007)	(0.003)	(0.009)	(0.004)
Log Absences VA	0.005	0.000	-10.348***	-0.024***	-0.020***	-0.017***	-0.027***	-0.010	0.001	0.012	0.014***
	(0.010)	(0.007)	(3.861)	(0.007)	(0.005)	(0.004)	(0.009)	(0.008)	(0.004)	(0.009)	(0.005)
Days Suspended VA	0.011	-0.021***	1.990	-0.017**	-0.007*	-0.002	-0.026***	-0.001	0.008**	0.053***	-0.002
	(0.008)	(0.007)	(4.179)	(0.007)	(0.004)	(0.004)	(0.010)	(0.008)	(0.003)	(0.008)	(0.004)
Held Back VA	0.002	-0.005	13.027*	-0.035***	-0.009	-0.004	0.013	0.021	0.005	-0.001	0.009
	(0.017)	(0.013)	(7.009)	(0.011)	(0.007)	(0.007)	(0.018)	(0.016)	(0.005)	(0.015)	(0.009)
Observations	135,786	102,517	60,694	293,028	233,078	233,078	152,345	151,820	316,123	277,333	221,757
R-squared	0.293	0.145	0.617	0.244	0.234	0.239	0.500	0.513	0.040	0.267	0.108

Note: This table reports the effect of a standard deviation increase in the measures of elementary school teacher value-added (see Table A.3) on students' high school outcomes. Specifically, each column of the table reports the coefficients on each of the normalized measures of teacher value-added from an OLS regression of the students' high school outcome on the measures of teacher value-added for the students' teachers in grades 3-5, along with the baseline controls described in section 4.1. The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.5: Effect of Elementary School Teacher Value-Added on High School Outcomes Using Alternative Grouping

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Test Score VA	-0.002	-0.002	6.238***	0.003	0.003**	0.005***	0.022***	0.016***	0.001	0.000	0.001
	(0.002)	(0.002)	(1.025)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)
Behavior VA	-0.003	0.005**	0.619	0.011***	0.006***	0.004***	0.010***	0.000	-0.002**	-0.015***	-0.003**
	(0.003)	(0.002)	(1.165)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)	(0.003)	(0.001)
All GPA VA	-0.001	0.004**	0.849	0.003	-0.000	-0.002	0.000	0.004	-0.003**	-0.004	-0.002
	(0.003)	(0.002)	(1.221)	(0.003)	(0.002)	(0.001)	(0.003)	(0.003)	(0.001)	(0.003)	(0.002)
Observations	135,786	102,517	60,694	293,028	233,078	233,078	152,345	151,820	316,123	277,333	221,757
R-squared	0.293	0.145	0.617	0.244	0.234	0.239	0.500	0.512	0.040	0.267	0.108

Note: This table reports the effect of a standard deviation increase in the three alternative measures of elementary school teacher value-added on students' high school outcomes. The test-score value-added is computed using the teachers' value-added for math and English test scores. The behavior value-added is computed using teachers' value-added for suspensions, log days absent, and not progressing to the next grade on time (held back). The All GPA value-added is computed using teachers' value-added for GPA, effort GPA, work and study habit GPA, and learning and social skills GPA. Each column of the table reports the coefficients on each of the three alternative measures of teacher value-added from an OLS regression of the students' high school outcome on the three alternative measures of teacher value-added for the students' teachers in grades 3 through 5 along with the baseline controls described in section 4.1. The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.6: Estimates of Forecast Bias

Grades 3-5 VA	Math Test Score	ELA Test Score	GPA	Habit GPA	Effort GPA	Log Absences	Suspended	Held back
Panel A: Using Actual Residualized Corresponding Outcome Variable								
Corresponding Teacher	0.989*	0.986	1.026	0.997	1.020	1.055	1.089	1.194
Value-Added Measure	(0.006)	(0.010)	(0.020)	(0.018)	(0.018)	(0.054)	(0.135)	(0.244)
Observations	669,779	670,013	408,062	403,686	403,686	470,540	478,901	370,654
Panel B: Using Predicted Corresponding Outcome Variable								
Corresponding Teacher	0.003*	0.001	0.013**	-0.001	0.004	-0.016	0.008	-0.001
Value-Added Measure	(0.002)	(0.003)	(0.005)	(0.003)	(0.003)	(0.016)	(0.009)	(0.002)
Observations	415,314	415,314	325,221	324,879	324,879	345,857	347,401	324,117

Note: Each cell in Panel A reports the coefficient on the indicated elementary school teacher value-added measure from an OLS regression of students' residualized (using the baseline controls) test score, GPA, or behavioral measure on the teacher value-added measure calculated using that outcome. Teacher valued-added is estimated using the leave-out year approach detailed in section 4.1, and scaled so both the outcome and value-added measure are in the same units. Each cell in Panel B reports the coefficient on the indicated teacher value-added measure from an OLS regression of students' predicted outcome (using double-lagged student achievement) on teachers' value-added measure after all variables are residualized using the baseline controls. The predicted outcomes are created by estimating an OLS regression of the outcome indicated in each column on third-order polynomials of double lagged (year $t-2$) math and English test scores, GPA, effort GPA, learning-skills GPA, log absences, an indicator for suspension, an indicator for being held back, and an indicator of English learner status. The coefficients obtained from this OLS regression are then used to predict students' outcomes. Standard errors clustered at the school-cohort level are reported in parentheses.

Table A.7: Effect of Elementary School Teacher Value-Added on High School Outcomes Using School-Grade Variation

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Test Score VA	-0.015	-0.008	3.082	-0.001	-0.004	0.000	0.033**	0.018	-0.004	-0.004	-0.002
	(0.013)	(0.008)	(5.395)	(0.009)	(0.005)	(0.005)	(0.015)	(0.014)	(0.006)	(0.011)	(0.006)
Behavior VA	-0.039	0.029*	2.100	0.043**	0.026**	0.021**	0.053*	0.036	-0.009	-0.030	-0.013
	(0.029)	(0.017)	(12.080)	(0.018)	(0.011)	(0.010)	(0.031)	(0.029)	(0.012)	(0.022)	(0.011)
Learning Skills VA	0.003	-0.022	-4.207	-0.049***	-0.020**	-0.018**	-0.047*	-0.039	0.017*	0.024	0.008
	(0.026)	(0.015)	(9.799)	(0.016)	(0.009)	(0.008)	(0.025)	(0.024)	(0.010)	(0.018)	(0.010)
Observations	3,106	3,094	2,878	6,250	5,175	5,175	4,044	4,043	6,290	6,233	5,141
R-squared	0.352	0.021	0.001	0.006	0.006	0.004	0.021	0.010	0.016	0.324	0.037

Note: This table reports the effect of a standard deviation increase in the three measures of elementary school teacher value-added (see Table 4) on students' high school outcomes using school-grade variation in teachers' value-added. Specifically, each column of the table reports the coefficients on the first difference in the school-grade level mean of each of the three measures of teacher value-added from a student weighted OLS regression of the first difference in the school-grade level mean of the outcome variable on the first difference in the school-grade level mean of each of the three measures of teacher value-added for grades 3 -5, along with year fixed effects. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.8: Effect of Elementary School Teacher Value-Added Measures on Additional High School Outcomes

Grades 3-5 VA	LAUSD Graduate	Took PSAT	PSAT Score	Grade 11 Math CST	Grade 11 English CST	Grade 8 Science CST	Grade 10 Science CST	Grade 8 Social Science CST	Grade 11 Social Science CST	World History CST	Number of AP Classes
Test Score VA	0.000	0.002	4.973***	0.016***	0.015***	0.017***	0.017***	0.018***	0.013***	0.014***	0.007**
	(0.001)	(0.001)	(0.523)	(0.005)	(0.004)	(0.002)	(0.003)	(0.002)	(0.004)	(0.003)	(0.004)
Behavior VA	0.001	-0.002	1.529**	0.005	0.003	0.008***	0.005	0.000	0.006	0.006*	0.014**
	(0.002)	(0.002)	(0.759)	(0.007)	(0.005)	(0.003)	(0.004)	(0.003)	(0.005)	(0.004)	(0.006)
Learning Skills VA	-0.001	0.000	-0.325	-0.008	-0.006	0.001	-0.008**	0.003	-0.013***	-0.006*	-0.006
	(0.002)	(0.002)	(0.683)	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)	(0.005)	(0.004)	(0.005)
Observations	64,039	228,103	184,898	48,330	53,448	331,200	136,922	273,642	54,880	113,571	316,123
R-squared	0.048	0.053	0.591	0.367	0.479	0.476	0.420	0.483	0.357	0.366	0.234

Note: This table reports the effect of a standard deviation increase in the three measures of elementary school teacher value-added (see Table 4) on students' high school outcomes. Specifically, each column of the table reports the coefficients on each of the three normalized measures of teacher value-added from an OLS regression of the students' high school outcome on the three measures of teacher value-added for the students' teachers in grades 3-5, along with the baseline controls described in section 4.1. The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.9: Effect of Elementary School Teacher Value-Added Estimated with Teacher Fixed Effects on High School Outcomes

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Test Score VA	-0.003	-0.002	9.273***	0.006***	0.005***	0.008***	0.030***	0.023***	0.001	-0.001	0.001
	(0.003)	(0.002)	(1.122)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)	(0.003)	(0.001)
Behavior VA	-0.005	0.010***	2.269	0.018***	0.010***	0.007***	0.016***	0.005*	-0.003**	-0.020***	-0.006***
	(0.003)	(0.003)	(1.512)	(0.003)	(0.002)	(0.001)	(0.004)	(0.003)	(0.002)	(0.004)	(0.002)
Learning Skills VA	-0.001	-0.000	-0.395	-0.002	-0.004**	-0.004***	-0.006*	0.001	-0.003*	0.004	0.002
	(0.004)	(0.002)	(1.233)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)	(0.002)
Observations	135,786	102,517	60,694	293,028	233,078	233,078	152,345	151,820	316,123	277,333	221,757
R-squared	0.293	0.145	0.617	0.244	0.234	0.239	0.501	0.512	0.040	0.267	0.108

Note: This table reports the effect of a standard deviation increase in the three measures of elementary school teacher value-added (see Table 4) on students' high school outcomes. The value-added estimates used in the three indices are estimated using Chetty, Friedman, and Rockoff's (2014a, 2014b) approach, which estimates the relationship between controls and student achievement using within-teacher variation. Specifically, equation (1) is modified to include a teacher fixed effect, but equation (2) is unchanged. Each column of the table reports the coefficients on each of the three normalized measures of teacher value-added from an OLS regression of the students' high school outcome on the three measures of teacher value-added for the students' teachers in grades 3-5, along with the baseline controls described in section 4.1. The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.10: Effect of Elementary School Teacher Value-Added on High School Outcomes Estimated Using Within-Teacher Variation

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Test Score VA	-0.017***	0.002	27.788***	0.026***	0.019***	0.024***	0.052***	0.050***	-0.002	-0.014***	-0.001
	(0.003)	(0.002)	(2.070)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.003)	(0.001)
Behavior VA	-0.001	0.013***	3.267	0.022***	0.012***	0.009***	0.026***	0.009**	-0.005***	-0.017***	-0.007***
	(0.004)	(0.003)	(2.017)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.002)	(0.004)	(0.002)
Learning Skills VA	-0.012***	0.000	7.429***	0.005	0.000	-0.001	-0.000	0.012***	-0.004**	-0.002	0.001
	(0.004)	(0.002)	(2.059)	(0.003)	(0.002)	(0.002)	(0.004)	(0.003)	(0.001)	(0.004)	(0.002)
Observations	135,786	102,517	60,694	293,028	233,078	233,078	152,345	151,820	316,123	277,333	221,757
R-squared	0.003	0.001	0.029	0.002	0.003	0.005	0.009	0.008	0.000	0.001	0.000

Note: This table reports the effect of a standard deviation increase in the three measures of elementary school teacher value-added (see Table 4) on students' high school outcomes. The value-added estimates used in the three indices are estimated using the Chetty, Friedman, and Rockoff (2014a, 2014b) approach, which estimates the relationship between controls and student achievement using within-teacher variation. Specifically, equation (1) is modified to include a teacher fixed effect, but equation (2) is unchanged. The outcome variables are residualized using the same approach. Each column of the table reports the coefficients on each of the three normalized measures of teacher value-added from an OLS regression of the students' residualized high school outcome on the three measures of teacher value-added for the students' teachers in grades 3-5 (and no controls). The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.11: Long-Term Effect of the Teacher Value-Added Measures Estimated Using a Factor Model

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Test Score VA	-0.002	-0.002	6.187***	0.002	0.003**	0.005***	0.022***	0.016***	0.001	0.001	0.001
	(0.002)	(0.002)	(1.023)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)
Behavior VA	-0.004	0.013***	2.336	0.013***	0.007***	0.004***	0.015***	0.006*	-0.003**	-0.021***	-0.005***
	(0.004)	(0.003)	(1.763)	(0.003)	(0.002)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)	(0.002)
Learning Skills VA	0.000	-0.004*	-0.790	-0.003	-0.003**	-0.003**	-0.008**	-0.001	-0.001	0.005	0.001
	(0.004)	(0.002)	(1.253)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.003)	(0.002)
Observations	135,786	102,517	60,694	293,028	233,078	233,078	152,345	151,820	316,123	277,333	221,757
R-squared	0.293	0.145	0.617	0.244	0.234	0.239	0.500	0.512	0.040	0.267	0.108

Note: This table reports the effect of a standard deviation increase in the three elementary school teacher value-added measures (created using a factor model) on students' high school outcomes. Each column of the table reports the coefficients on each of the three normalized measures of teacher value-added from an OLS regression of the students' high school outcome on the three measures of teacher value-added for the students' teachers in grades 3-5, along with the baseline controls described in section 4.1. The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. The factor loads are reported in Table A.12. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.12: Factor Loads for Elementary School Value-Added

Grades 3-5 VA	Test Score	Behavior	Learning Skills
Math Test Scores	0.804		
English Test Scores	0.804		
GPA		0.253	
Log Absences		-0.271	
Suspended		-0.160	
Held Back		-0.049	
Effort GPA			0.798
Work and Study GPA			0.917
Learning and Social GPA			0.886

Note: This table reports the factor loads for the test-score, behavior, and learning-skills teacher valued-added measures in elementary school (grades 3-5). Each of the three measures of teacher value-added is created using factor analysis. The test-score value-added is computed using the teachers' value-added for math and English test scores, and the factor loads are shown in column 1. The behavior value-added is computed using teachers' value-added for GPA, suspensions, log days absent, and not progressing to the next grade on time (held back), and the factor loads are shown in column 2. The learning-skills value-added is computed using teachers' value-added for effort GPA, work and study habit GPA, and learning and social skills GPA, and the factor loads are shown in column 3.

Table A.13: Effect of Elementary School Teacher Value-Added on High School Outcomes Using Exploratory Factor Analysis

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Factor 1 (Loads Primarily	-0.002	0.004**	1.394	0.004*	0.001	-0.000	0.003	0.005**	-0.003**	-0.006*	-0.002
on Non-test Score)	(0.003)	(0.002)	(1.207)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.001)	(0.003)	(0.001)
Factor 2 (Loads Primarily	-0.002	-0.001	6.507***	0.004**	0.003***	0.005***	0.023***	0.017***	0.001	-0.001	0.000
on Test Scores)	(0.002)	(0.002)	(1.029)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)
Observations	135,786	102,517	60,694	293,028	233,078	233,078	152,345	151,820	316,123	277,333	221,757
R-squared	0.293	0.145	0.617	0.244	0.234	0.239	0.500	0.512	0.040	0.266	0.108

Note: This table reports the effect of a standard deviation increase in the two alternative measures of elementary school teacher value-added on students' high school outcomes. The two factors are computed using a model with all of our measures of teacher value-added included simultaneously rather than by imposing that only certain value-added measures contribute to certain factors. The value-added measures used to compute the factors include math and English test scores, suspensions, log days absent, not progressing to the next grade on time (held back), GPA, effort GPA, work and study habits GPA, and learning and social skills GPA. The second factor primarily loads on the two test-score value-added measures and to a lesser extent GPA, whereas the first factor loads primarily on GPA and the other non-test-score variables. Each column of the table reports the coefficients on each of the two alternative measures of teacher value-added from an OLS regression of the students' high school outcome on the two alternative measures of teacher value-added for the students' teachers in grades 3-5, along with the baseline controls described in section 4.1. The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.14: Factor Loads for the Exploratory Factor Analysis

Grades 3-5 VA	Factor 1	Factor 2
Math Test Scores	0.102	0.807
English Test Scores	0.118	0.808
GPA	0.781	0.126
Log Absences	0.121	0.044
Suspended	0.065	0.029
Held Back	0.008	0.024
Effort GPA	0.845	0.068
Work and Study GPA	0.910	0.088
Learning and Social GPA	0.866	0.050

Note: This table reports the factor loads for the two factors used in Table A.13. The two factors of teacher value-added measures are created using factor analysis. All nine of the listed value-added measures are variables included in the factor analysis. Factors with an eigenvalue grater than one are included.

Table A.15: Effect of the Elementary School Teacher Value-Added Measures on High School Outcomes without Behavior Value-Added

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Test Score VA	-0.002	-0.001	6.364***	0.003*	0.003***	0.005***	0.023***	0.017***	0.001	-0.000	0.001
	(0.002)	(0.002)	(1.024)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)
Learning Skills VA	-0.002	0.003	0.391	0.004	0.000	-0.001	0.000	0.002	-0.003**	-0.005*	-0.002
	(0.003)	(0.002)	(1.169)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)	(0.003)	(0.001)
Observations	136,236	102,901	60,915	293,777	233,696	233,696	152,803	152,272	316,973	278,053	222,335
R-squared	0.293	0.145	0.617	0.244	0.234	0.239	0.500	0.512	0.040	0.266	0.108

Note: This table reports the effect of a standard deviation increase in test-score and learning-skills elementary school teacher value-added (see Table 4) on students' high school outcomes. Specifically, each column of the table reports the coefficients on the test-score and learning-skills normalized measures of teacher value-added from an OLS regression of the students' high school outcome on the two measures of teacher value-added for the students' teachers in grades 3-5, along with the baseline controls described in section 4.1. The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.16: Long-Term Effect of the Teacher Value-Added Measures Estimated in Percentiles

Grades 3-5 VA	LAUSD Dropout	Took SAT	SAT Score Percentile	GPA Percentile	Effort GPA Percentile	Cooperation GPA Percentile	Math CAHSEE Percentile	English CAHSEE Percentile	Days Suspended	Log Absences	Held Back
Test Score VA Decile	-0.0117	-0.0023	0.0017	-0.0003	0.0001	0.0000	0.0247***	0.0102***	-0.0204	-0.0374	0.0066
	(0.0143)	(0.0153)	(0.0011)	(0.0002)	(0.0002)	(0.0002)	(0.0053)	(0.0024)	(0.0207)	(0.0298)	(0.0131)
Behavior VA Decile	-0.0312	0.1075***	0.0033	0.0003	0.0002	0.0004	0.0181**	0.0063	-0.0254	-0.1405**	-0.1112***
	(0.0501)	(0.0365)	(0.0025)	(0.0007)	(0.0011)	(0.0011)	(0.0088)	(0.0061)	(0.0324)	(0.0630)	(0.0297)
Learning Skills VA Decile	-0.0078	0.0075	-0.0001	-0.0001	0.0026	0.0026	0.0043**	0.0001	-0.0027	-0.0186	-0.0118
	(0.0149)	(0.0111)	(0.0009)	(0.0001)	(0.0020)	(0.0020)	(0.0021)	(0.0010)	(0.0034)	(0.0134)	(0.0094)
Observations	135,786	102,517	60,694	293,028	233,078	233,078	152,345	151,820	316,123	277,333	221,757
R-squared	0.293	0.145	0.245	0.166	0.937	0.906	0.141	0.135	0.039	0.267	0.108

Note: This table shows the effect of a decile increase in three elementary school teacher value-added indices on high school outcomes where teacher value-added, GPA, and test scores are measured in ordinal units, deciles in the case of teacher-value added and percentiles in the case of test scores and GPA. Each column of the table reports the coefficients on each deciles increase in teacher value-added from an OLS regression of the students' high school outcome on the three measures of teacher value-added for the students' teachers in grades 3-5, along with the baseline controls described in section 4.1. The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. The factor loads are reported in Table A.12. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.17: Relationship between Elementary Subject GPA and High School Outcomes

Grades 3-5 GPA	LAUSD Dropout	Took SAT	SAT Score	GPA	Effort GPA	Cooperation GPA	Math CAHSEE	English CAHSEE	Days Suspended	Log Absences	Held Back
Math GPA	-0.038***	0.054***	56.902***	0.156***	0.075***	0.060***	0.337***	0.096***	-0.010***	-0.105***	-0.046***
	(0.003)	(0.003)	(1.983)	(0.004)	(0.002)	(0.002)	(0.005)	(0.004)	(0.003)	(0.005)	(0.002)
Reading GPA	0.004	0.011***	53.384***	-0.017***	-0.015***	-0.009***	0.019***	0.169***	0.012***	-0.009	0.010***
	(0.004)	(0.004)	(2.185)	(0.005)	(0.003)	(0.003)	(0.005)	(0.005)	(0.003)	(0.006)	(0.003)
Writing GPA	-0.029***	0.057***	-0.449	0.151***	0.082***	0.059***	0.022***	0.107***	-0.019***	-0.016**	-0.038***
	(0.004)	(0.004)	(2.053)	(0.005)	(0.003)	(0.002)	(0.005)	(0.005)	(0.003)	(0.006)	(0.003)
Listening GPA	-0.057***	0.064***	-11.538***	0.209***	0.131***	0.134***	0.048***	0.025***	-0.070***	-0.069***	-0.050***
	(0.004)	(0.004)	(2.069)	(0.004)	(0.003)	(0.002)	(0.005)	(0.005)	(0.004)	(0.006)	(0.003)
Speaking GPA	0.024***	-0.037***	0.697	-0.130***	-0.081***	-0.088***	-0.046***	-0.022***	0.040***	0.063***	0.026***
	(0.004)	(0.004)	(2.215)	(0.005)	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)	(0.006)	(0.003)
History/Social GPA	-0.016***	0.018***	-3.133	0.070***	0.040***	0.031***	-0.019***	0.001	-0.013***	-0.030***	-0.022***
	(0.004)	(0.004)	(2.182)	(0.005)	(0.003)	(0.003)	(0.006)	(0.005)	(0.004)	(0.006)	(0.003)
Science GPA	-0.001	0.007*	21.426***	0.019***	0.008***	0.010***	0.020***	0.038***	-0.009**	-0.030***	-0.008***
	(0.004)	(0.004)	(2.204)	(0.005)	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)	(0.006)	(0.003)
Health GPA	0.008*	0.001	1.451	-0.001	0.003	0.011***	0.006	0.003	-0.004	0.007	0.004
	(0.005)	(0.005)	(2.861)	(0.006)	(0.003)	(0.003)	(0.007)	(0.006)	(0.005)	(0.007)	(0.003)
PE GPA	0.021***	-0.032***	-30.433***	-0.099***	-0.065***	-0.071***	-0.046***	-0.115***	0.042***	-0.006	0.024***
	(0.004)	(0.004)	(2.391)	(0.005)	(0.003)	(0.003)	(0.006)	(0.005)	(0.004)	(0.007)	(0.003)
Arts GPA	-0.016***	0.012***	-0.000	0.063***	0.037***	0.037***	-0.009	0.046***	-0.032***	0.004	-0.011***
	(0.004)	(0.004)	(2.410)	(0.005)	(0.003)	(0.003)	(0.006)	(0.005)	(0.004)	(0.007)	(0.003)
Observations	206,598	156,359	93,417	400,540	328,035	328,035	222,916	222,158	433,604	380,614	311,585
R-squared	0.288	0.167	0.645	0.286	0.276	0.282	0.537	0.546	0.043	0.265	0.121

Note: This table reports the predictive effect of a one-point increase in a students' GPA in each of the 10 elementary school subjects on students' high school outcomes. Specifically, each column of the table reports the coefficients on each of the students' 10 subject GPAs from an OLS regression of the students' high school outcome on students' 10 subject GPAs in grades 3-5. The sample includes students in grades 3-5 who attended high school in the LAUSD. Standard errors clustered at the school-cohort level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.