National Center for Analysis of Longitudinal Data in Education Research

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

Teach For America Impact Estimates on Nontested Student Outcomes

Ben Backes

Michael Hansen

WORKING PAPER 146 • OCTOBER 2015

Teach For America Impact Estimates on Nontested Student Outcomes

Ben Backes
American Institutes for Research

Michael Hansen
American Institutes for Research

Contents

Acknowledgements

This work was produced under financial support from the John S. and James L. Knight Foundation. The Knight Foundation funded both the TFA clustering strategy and this external evaluation of that strategy. The Knight Foundation, Teach For America, and Miami-Dade County Public Schools were cooperative in facilitating this evaluation and received early drafts of these findings, but none had editorial privilege over the actual contents or findings presented here.

CALDER has a current contract with Teach For America to conduct a separate evaluation of corps member impacts in a different region of the country, which Teach For America procured through a competitive bid process; these evaluations are independent of each other, and TFA has no editorial influence in the presentation of findings in either the current study or this separate contract.

Our research is also funded by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER) and an anonymous foundation. Research supported by CALDER is funded through Grant R305C120008 to the American Institutes for Research from the Institute of Education Sciences, U.S. Department of Education.

We gratefully acknowledge the cooperation of Miami-Dade County Public Schools, especially Gisela Feild, in providing data access and research support and allowing us to interview district and school personnel regarding Teach For America. We acknowledge the cooperation of Teach For America's national and Miami regional offices in conducting this research. We thank Victoria Brady and Jeff Nelson for excellent research assistance, and appreciate the helpful comments from Cory Koedel, Raegen Miller, Jason Atwood, Rachel Perera, James Cowan, and participants in presentations of findings at the 2015 AEFP Conference.

CALDER working papers have not undergone final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. Any and all remaining errors are our own.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

Teach For America Impact Estimates on Nontested Student Outcomes

Ben Backes & Michael Hansen

CALDER Working Paper No. 146
October 2015

Abstract

Recent evidence on teacher productivity suggests teachers meaningfully influence noncognitive student outcomes that are commonly overlooked by narrowly focusing on student test scores. These effects may show similar levels of variation across the teacher workforce and are not significantly correlated with value-added test score gains. Despite a large number of studies investigating the TFA effect on math and English achievement, little is known about nontested outcomes. Using administrative data from Miami-Dade County Public Schools, we investigate the relationship between being in a TFA classroom and non-test student outcomes. We validate our use of nontest student outcomes to assess differences in teacher productivity using the quasi-experimental teacher switching methods of Chetty, Friedman, and Rockoff (2014) and find multiple cases in which these tests reject the validity of candidate nontest outcomes. Among the cases deemed valid, we find suggestive evidence that students taught by TFA teachers in elementary and middle school were less likely to miss school due to unexcused absences and suspensions (compared to non-TFA teachers in the same school), although point estimates are very small. Other nontest outcomes were found to be valid but showed no evidence of a TFA effect.

Introduction

Teach For America (TFA) is an alternative certification program that selects, trains, and places recent college graduates or other young professionals into high-need schools with a two-year commitment to teach.[1] Much of the prior empirical research on TFA has primarily focused on the impacts of TFA corps members on students' standardized test scores. In general, these studies have shown a positive TFA effect in math (and science, where available) relative to comparison teachers in the same schools teaching similar students, but no significant effect is generally detected in reading (e.g., Clark et al., 2013; Glazerman et al., 2006; Hansen et al., 2015; Xu et al., 2011).

Focusing on test score gains alone, however, is a narrow view of TFA's effects on students and the schools where they teach. TFA has a service-oriented mission that describes corps members as leaders and mentors to disadvantaged youth who help students set high expectations and change their learning and life trajectories in meaningful ways to overcome the challenges of generational poverty. Given the careful selection of TFA corps members based on dimensions of leadership and character (e.g., Dobbie, 2011), one might hypothesize that TFA corps members could have broader impacts on their students beyond simply test scores by influencing other student behaviors in a meaningful way. Recent evidence on teacher productivity suggests teachers meaningfully influence nontest student outcomes that are commonly overlooked by narrowly focusing on student test scores (Jackson, 2014).

This paper uses longitudinal data from recent years in the Miami-Dade County Public Schools, spanning 2008–2009 through 2013–2014, to examine the impact of TFA corps members on student nontested outcomes. We do this by constructing estimates of TFA effectiveness in these nontested outcomes (we refer to these value-added estimates of teacher effectiveness in nontested outcomes as

[1] In most regions where it operates, TFA is an alternative certification program and is widely recognized as such. In Miami, however, TFA does not set teachers up for permanent certification and cannot be technically considered an alternative certification program.

N-VAMs). During the time span of the longitudinal data, TFA experimented with a new placement strategy where exceptionally low-performing schools in the district were specifically targeted for intensive TFA placements. At the same time, the size of the TFA corps in the district more than tripled. The effective result of these two changes in TFA placements resulted in large clusters of TFA in these targeted schools, where on average more than 10% of the teacher workforce was comprised of TFA corps members. The large infusion of the TFA corps in these schools results in large sample sizes and relatively precise estimates, which is critical for our investigation.

This paper presents two main findings. First, we (a) demonstrate that N-VAMs obtained in our data capture persistent differences in teacher effectiveness and (b) using a teacher switching quasi-experiment adopted from Chetty et al. (2014), provide suggestive evidence that these teacher effects represent causal impacts on student outcomes. This first step is important because, in contrast to VAMs based on student test scores, current evidence on whether N-VAMs represent causal effects on student outcomes is, to our knowledge, limited to one paper: Jackson (2014). However, Jackson (2014) is largely devoted to showing that N-VAMs are predictive of long-run outcomes. Thus, Jackson's analysis is narrowly tailored: The sample includes only students in Grade 9; it examines only one composite index of noncognitive outcomes; and it does not examine consistency of N-VAMs within teachers or across outcomes. To our knowledge, no other attempt to validate N-VAMs exists.

Our second main finding links TFA teachers with students' nontest outcomes, where we estimate TFA effects in a value-added framework. We find suggestive evidence that student behavior in elementary and middle school—as measured by days missed due to unexcused absences and suspensions—improves by a small degree when placed in a TFA classroom, compared to non-TFA teachers in the same schools with similar students. We also find a small increase in GPA for elementary school students in TFA classrooms.

This paper proceeds as follows. First, we discuss the background research on attempts to measure teachers' contributions to noncognitive student outcomes. Next, we address TFA placement in M-DCPS and the data we use. Following this, we describe how we construct estimates of TFA effectiveness at improving student nontest outcomes and how we forecast N-VAMs for our validation procedure. We then explore the properties of our forecasted N-VAMs and then estimate TFA effectiveness on these measures. Finally, we conclude.

Teachers' Contributions to Student Nontest Outcomes

Noncognitive skills are defined by Garcia (2014) as "*traits that are not directly represented by cognitive skills or by formal conceptual understanding, but instead by social-emotional or behavioral characteristics that are not fixed traits of the personality, and that are linked to the educational process, either by being nurtured in the school years or by contributing to the development of cognitive skills in those years (or both)*" (p. 6). There is a growing body of research indicating that these noncognitive skills—such as motivation, grit, self-control, and social skills—are important, malleable, and predictive of external outcomes. Grit and self-control, for example, have been found to be predictive of outcomes ranging from persisting at West Point to retention for novice teachers to graduating from high school (Duckworth et al., 2007; Duckworth et al., 2009; Robertson-Kraft & Duckworth, 2014; Eskreis-Winkler et al., 2014). These skills may be as important in determining success in both college and the labor market as cognitive skills, as measured by test scores (Heckman et al., 2006). Furthermore, early in a person's life is the period in which these skills are still malleable (Kautz et al., 2014). Thus, students' time in school is a ripe time to develop these vital soft skills.

Given the importance of these nontest outcomes, it is no surprise that researchers and practitioners have shown a growing interest in measuring these skills and how school factors, especially teachers, contribute to student growth in these skills. For example, studies from Chetty et al. (2011) and

Dee and West (2011) find that students assigned to smaller classrooms have persistent gains in noncognitive outcomes that explain future earnings increases. Some strands of research examining teacher contributions to nontested outcomes (i.e., N-VAMs) show that some teachers are more effective than others at improving noncognitive outcomes (Garcia, 2014; Gershenson, forthcoming; Jackson, 2014) and documenting differential returns to teacher experience for various noncognitive outcomes (Ladd & Sorensen, 2014). These nontest teacher effects may show similar levels of variation across the teacher workforce and are not significantly correlated with value-added test score gains (e.g., Jackson, 2014; Gershenson, 2014). Some prior studies of TFA corps members have explored TFA impacts on these types of nontest student outcomes, such as student absences and grade retention, but have not found any significant differences associated with these outcomes (e.g., Mayer et al., 2004; Clark et al. 2013). We wish to investigate this issue further using our data, which provides information on a broader set of student nontest outcomes across a larger span of grade levels.

TFA Background in M-DCPS

TFA has been placing corps members in M-DCPS since 2003, beginning with 35 initial placements and aiming to place corps members in schools with high levels of student poverty (student bodies exceeding 70% eligibility for free or reduced-price lunch). Beginning with the 2009–2010 school year, TFA began a clustering strategy in which new placements were purposely assigned to schools within designated high-need communities, which contained the district's lowest performing schools.

TFA's clustering placement strategy grew out of an interest in accelerating TFA's impact on student outcomes. The growth of the TFA corps and its density is readily apparent in the placement numbers during the six school years of data used for this analysis. Table 1 presents TFA corps member assignment figures over time. In the 2008–2009 school year, the year immediately preceding the clustering strategy, there was an average of slightly less than two TFA corps members in each school where they were placed. In the years following, the number of schools containing any TFA dropped by about half and the number of active TFA corps members in the district more than tripled, resulting in an average of nearly 10 corps members per school where there was any presence. The net result was a jump in the proportion of TFA corps members in placement schools, going from 2–4% in 2008–2009 to as high as 14–17% in 2012–2013. The concentrations of TFA in placement schools decreased slightly in 2013–2014, due to expanding into more target schools.

Data

We use detailed student-level administrative data that cover M-DCPS students linked to their teachers for six school years (2008–2009 through 2013–2014) from kindergarten through 12th grade. M-DCPS is the largest school district in Florida and the fourth largest in the United States. The district has large minority and disadvantaged student populations, typical of regions TFA has historically targeted; about 60% of its students are Hispanic, 30% Black, and 10% White, and more than 60% of students qualify for free or reduced-price lunch.

Our set of noncognitive outcomes includes six variables: the number of unexcused absences, days absent due to suspension, grade point average, percent of classes failed, grade repetition, and a time-invariant measure of whether a student ever graduates.[2] In addition to these outcomes, we observe a variety of student characteristics: race; gender; free or reduced-price lunch (FRL) eligibility; limited English proficiency (LEP) status; whether a student is flagged as having a mental, physical, or emotional disability; attendance; and disciplinary incidents. In addition, all students are linked to teachers through data files that contain information on course membership.

[2] Grade point average is calculated using course grades from transcripts, where earning an A corresponds to 4 grade points, earning a B equals 3 grade points, etc. For all analysis involving graduation outcomes, we restrict the sample to students that could have plausibly graduated given the grade a student was in when observed and the length of our panel. For a given grade and year, the precise inclusion criteria is 2002 + grade >= year. Thus, for example, sixth graders in 2008 would be included (because progressing on time would have them graduate at the end of the 2014 year, which would be indicated in our data), but fifth graders in 2008 would not.

Teacher personnel files in the M-DCPS data contain information on teachers' experience levels and demographics. These are used as covariates for the analysis that follows. One variable included in the data is a flag on TFA teachers (representing both active corps members and TFA alumni); given the importance of this variable in the analysis, we externally validated this variable with historical corps member lists from TFA.

Empirical Strategy

Conventional VAMs and Derivative N-VAMs

Researchers typically estimate a VAM similar to the following:

$$Y_{ijt} = \beta_0 + Y_{it-1}\beta_1 + X_{it}\beta_2 + \sum_{j\in J} T_j\phi_j + \varepsilon_{ijt}, \tag{1}$$

where Y_{ijt} represents test scores of student *i* taught by teacher *j* in year *t*, Y_{it-1} prior year achievement, X_{it} demographic characteristics, T_j an indicator variable identifying the *jth* teacher of *J* total teachers, and ε_{ijt} an error term. The coefficient on teacher indicator *j*, ϕ_j, is meant to capture teacher *j*'s contribution to growth in student achievement.

Recently developed N-VAMs are close derivatives of the conventional VAM in Equation (1), simply substituting some nontest outcome (notated here as ${Y_{ijt}}^{OUTCOME}$) as the dependent variable, along with its lagged observation as an explanatory variable:[3]

$${Y_{ijt}}^{OUTCOME} = \beta_0 + {Y_{ijt-1}}^{OUTCOME}\beta_1 + X_{it}\beta_2 + \sum_{j\in J} T_j\phi_j + \varepsilon_{ijt}. \tag{2}$$

Models of this type are calculated in Gershenson (forthcoming), Ladd and Sorensen (2014), and others. N-VAMs produced using this conventional methodology are assumed to take on the

[3] For ease of notation, we write ${Y_{ijt}}^{OUTCOME}$ as Y_{ijt} in the remainder of the text, and it is meant to apply to any candidate nontest outcome. The procedure will be performed separately for each outcome.

interpretation and analogous statistical properties of VAMs. For example, estimates are intended to be interpreted as teachers' causal contributions to the corresponding nontest outcome in students; our validity tests described below will help evaluate this claim. N-VAMs may be estimated across various time spans in the data, producing one-year or multi-year teacher estimates as the case may be. The reliability and variability of these estimates over time within teachers can be calculated following methods developed for test-based VAMs in McCaffrey et al. (2009) and Goldhaber and Hansen (2013).

Estimating a TFA Effect on Nontest Outcomes

To address the study's research question regarding the influence of TFA corps members on student nontest outcomes, we replace the teacher fixed effects in Equation (2) above with a TFA indicator and a vector of other teacher characteristics (X_{jt}), school fixed effects (γ_s), and classroom average characteristics (X_{ct}) in order to estimate the average change in student outcomes associated with being in a TFA classroom:

$$Y_{ijt}{}^{OUTCOME} = \beta_0 + Y_{ijt-1}{}^{OUTCOME}\beta_1 + X_{it}\beta_2 + TFA_{it}\beta_3 + X_{jt}\beta_4 + X_{ct}\beta_5 + \gamma_s + \varepsilon_{ijt}. \tag{3}$$

Equation (3) is similar to existing studies of TFA (e.g., Boyd et al., 2006; Clark et al., 2013; Glazerman et al., 2006; Hansen et al., 2014; and Kane et al., 2008). However, where these studies use student test scores as the outcome variable and control for prior year scores, we measure nontest score outcomes: unexcused absences, absences due to suspension, grade point average, percent of classes failed, grade repetition, and graduation. Thus, we estimate Equation (3) for each of these six outcome variables. Regardless of the outcome variable Y_{isct}, we control for a student's lagged value of all of these outcome variables in all regressions (with the exception of graduation, since there is no variation in lagged graduation among students observed in the current year).[4]

[4] The vector of student characteristics includes the following: race; gender; free or reduced-price lunch (FRL) eligibility; limited English proficiency (LEP) status; and mental, physical, or emotional disability status. The vector

Because TFA corps members are placed nonrandomly across schools in the district (at minimum, we know selected schools had at least 70% of students eligible for free or reduced-price lunch; other characteristics may have also played into the selection decision), the point estimates associated with TFA effect may be downward biased because schools chosen to receive TFA corps members were likely targeted precisely because they were likely to have low student performance (both on tests and other nontest outcomes). As a result, we include both school fixed effects (γ_s) and controls for time-varying averages within classrooms such as student demographic characteristics. The inclusion of school fixed effects ensures that TFA teachers are compared to non-TFA teachers within a given school, serving similar classrooms.

Forecasting Nontest Value-Added

To both validate the causal nature of N-VAMs and explore the variability of these measures within and across teachers, we take a different approach from Equation (2) for two reasons. First, Equation (2), when estimated directly on the full sample, assumes that the teacher contribution to the corresponding outcome is fixed over time by estimating a constant $\hat{\phi}_j^{OUTCOME}$ for each teacher. Goldhaber and Hansen (2013) and Chetty et al. (2014), however, provide evidence using test-based VAMs that this is not the case: teacher effectiveness drifts over time. Second, to conduct our validation procedure below, we need to forecast teacher effectiveness in year *t*, not estimate it directly as in Equation (2). As a result, we follow the three-step process used by Chetty et al. (2014).[5]

of classroom characteristics includes class size and classroom-level averages of each of the student characteristics listed above. Teacher controls include teacher race, gender, experience, and whether the race of the teacher matches that of the student. The student characteristic, class average, and teacher demographic controls are interacted with grade indicator variables to allow differences in the influence of these variables across grades. The estimating equation additionally includes indicator variables on grades and years.

[5] The following discussion is meant to provide a conceptual overview for how forecasts of teachers' value-added contributions are constructed. In practice, we use the "vam.ado" program developed for use by Chetty et al. (2014). See Online Appendix A of Chetty et al. (2014) for a step-by-step guide to implementing the methods developed in the paper.

We follow this procedure for each of our six outcomes—absences, suspensions, grade retention, classes failed, grade point average, and graduation—to obtain a forecast of each teacher's value-added effect on each outcome. First, we residualize the outcome variable ${Y_{ijt}}^{OUTCOME}$ by regressing it on the student's prior outcome value (Y_{ijt-1}), student demographics (X_{ijt}), and a vector of teacher fixed effects (T_j):

$$Y_{ijt} = \beta_0 + Y_{ijt-1}\beta_1 + X_{ijt}\beta_2 + T_j\phi + \varepsilon_{ijt}. \tag{4}$$

In all regressions, X_{ijt} includes FRL eligibility, English language learner status, gender, race, special education status, and an indicator of whether the student has an identified learning disability. Y_{ijt-1} includes a cubic function of the lagged outcome variable (unless the outcome is binary, in which case it is simply one indicator variable), and sometimes will include lagged values of other outcomes, as discussed below.[6] Using the estimates of β_0, β_1, and β_2 from the regression in Equation (2), we obtain residualized student outcomes:

$$\hat{Y}_{ijt} = Y_{ijt} - \hat{\beta}_0 - Y_{ijt-1}\hat{\beta}_1 - X_{ijt}\hat{\beta}_2 = T_j\phi + \varepsilon_{ijt}. \tag{5}$$

A teacher's average residuals in year *t* can then be written as $\bar{Y}_{jt} = \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_{ijt}$, where *n* is the number of students taught in year *t* by teacher *j*. Let Y_j^{-t} be the vector of mean residuals in years other than *t*.

Second, we obtain coefficients from the estimation of the best linear predictor of $\bar{Y}_{jt}$ given all other years, both past and future. Specifically, we choose the vector ψ to minimize the mean-squared error of forecasts of test scores across all teacher-year observations in the sample:

[6] For all outcomes other than GPA, for the validation procedure, we first transform each nontest outcome and its lag by adding 1, taking the log, and then standardizing the values to be mean zero and standard deviation 1.

$$\psi = arg\ min_{\{\psi_1,\dots,\psi_{t-1}\}} \sum_j \left(\bar{Y}_{jt} - \sum_{s=1}^{t-1} \psi_s \bar{Y}_{js} \right)^2 . \tag{6}$$

This is equivalent to regressing $\bar{Y}_{jt}$ on the observations contained in the vector Y_j^{-t} and obtaining the coefficients.

Finally, we use the coefficients obtained in Equation (5) to forecast a teacher's value-added contribution for outcome *OUTCOME* in year *t*:

$$\hat{\mu}_{jt}{}^{OUTCOME} = \psi' Y_j^{-t} \tag{7}$$

The result of this process is a forecast of each teacher's value-added contribution in each year for each outcome. In all that follows, we take this forecast to be a teacher's value-added effect in *t*. Thus, for example, when calculating the correlation between teacher effects across two outcome measures *Y1* and *Y2* in a given year, we would calculate the correlation between $\hat{\mu}_{jt}{}^{Y1}$ and $\hat{\mu}_{jt}{}^{Y2}$. Although this does not actually use data from t, $\hat{\mu}_{jt}{}^{Y}$ represents the best linear prediction of teacher performance in *t*, given data from all other years. Chetty et al. (2014) find that these forecasts of teacher performance are predictive of student performance in math and reading. Below, we assess whether forecasts of nontest outcomes exhibit similar properties.

Definition of Bias

To define bias, consider the following regression of outcome residuals on forecasted value-added when students are randomly assigned to teachers:

$$\hat{Y}_{ijt} = \alpha_t + \theta \hat{\mu}_{jt} + \zeta_{ijt} \tag{8}$$

We adopt the Chetty et al. (2014) definition of forecast bias of 1 - θ. Chetty et al. (2014) use two tests for bias to present evidence that the estimates of teacher effectiveness $\hat{\mu}_{jt}$ are forecast unbiased predictors of student test scores in t if the vector of prior achievement Y_{ijt-1} (from Equation (2)) contains lagged test scores. As described below, we will replicate these two tests for bias using the study's four nontest outcomes to determine whether value-added estimates based on these outcomes contain forecast bias.

Assessing Validity

If we take a VAM framework and apply it to nontested outcomes, do the estimates of teacher performance correspond to causal changes in actual student outcomes, as with tested outcomes? The idea behind our test is simple: if we know how students of a given teacher performed in the past, can we make out-of-sample predictions about the outcomes of future students taught by that teacher? Thus, our tests of validity are designed to generate predictive evidence about whether VAMs applied to nontest outcomes truly measure a teacher's ability to improve those outcomes for the students they teach. We believe predictive evidence on N-VAMs is an important first step in assessing whether they should be used in research and policy.

A regression of predicted student outcomes ($\hat{Y}_{ijt}$ from Equation 4) on the forecast value-added estimates ($\hat{\mu}_{jt}$ from Equation 6) is not directly informative about whether $\hat{\mu}_{jt}$ represents a causal relationship between $\hat{Y}_{ijt}$ and $\hat{\mu}_{jt}$because (a) $\hat{\mu}_{jt}$ was constructed to be the best linear predictor of $\hat{Y}_{ijt}$, regardless of the causal relationship, and (b) students are not randomly assigned to teachers. In other words, a coefficient of 1 could either be because of a causal relationship or because of persistent differences in ζ_{ijt} in Equation (7) (e.g., certain teachers being assigned to students with high- or low-income parents). Thus, as in Chetty et al. (2014), we leverage variation in student exposure to teacher N-VAM contributions caused by staffing changes at the grade-school level to test whether student nontest outcomes change corresponding to a 1:1 relationship.

Because we forecast student achievement in multiple years, we can use the predictions to implement the teacher switching research design of Chetty et al. (2014) and Bacher-Hicks et al. (2014). We use variation in student exposure to teacher quality at the school-by-grade level induced by teacher staffing changes to estimate the predictive validity of teacher value-added contributions on student nontest outcomes. Letting ΔY_{sgt} and $\Delta\hat{\mu}_{sgt}$ be first-differenced student outcomes and forecasted teacher value-added contributions aggregated to the school-grade-year level, respectively, we run the following regression:

$$\Delta Y_{sgt} = \mu + \Delta\hat{\mu}_{sgt}\alpha + \epsilon_{sgt} \tag{9}$$

To construct $\Delta\hat{\mu}_{sgt}$, the differenced average teacher value-added contributions at the school-grade-year level, we weight teachers by the number of students they taught. To perform this test, when estimating value-added contributions for teachers in *t*, we omit observations from *t* and *t-1* so as not to introduce bias when using differenced outcomes on the left hand side. In addition, we include year fixed effects and cluster at the school-cohort level.

The key driver of changes in teacher value-added contributions at the school-grade-year level is school staffing changes, where teachers are moving across schools or grades. The crucial identifying assumption is that changes in aggregate student outcomes are driven only by these changes in school staffing decisions and not other factors which influence test scores. Due to the inherent cost to students in switching schools, the assumption appears to be a reasonable one. Importantly, this test avoids the problem of nonrandom student-teacher sorting within schools because it aggregates teacher value-added contributions and student performance to school-grade cells.

In Equation (9), forecast bias is defined as 1-α. If α deviates significantly from unity, it would indicate that the informational content of $\hat{\mu}_{jt}$ does not hold across schools. We test for the presence of forecast bias across the six candidate nontest outcomes, first as a pooled sample and then separately by levels (elementary, middle, and high school grades).

In any given year, students may be exposed to multiple teachers, especially in middle school and high school. When estimating TFA effects, we include each student-teacher link as its own observation and dosage weight each of these records.[7] When conducting the test for forecast bias, we obtain each teacher's school-grade-year weight by dividing the sum of his or her dosages by the total sum of dosages.[8] The mean outcome in the school-grade-year cell is simply the mean across students in that cell.

Results

Basic Properties of Forecasted VA

In order to generate out-of-sample forecasts of VA for each outcome (both tested and nontested outcomes), it has to be the case that the outcomes of students in a teacher's classroom in one year are correlated with students of that teacher in a different year. Otherwise, we would have nothing to base forecasts off of and every teacher would have forecasted VA of exactly zero. Throughout this section, we consider elementary, middle school, and high school teachers as separate groups due to differences in results across school types, though not all outcomes of interest can be estimated at each level. Because students in high school grades are not annually tested, we do not include tested

[7] This method is referred to as the Full Roster Method by Hock and Isenberg (2012).
[8] In the paper where Chetty et al. (2014) develop their test for forecast bias, they keep one observation per student per year, thus there are no instances of multiple teachers per student in their data. In our case, students are linked to multiple teachers in a given year. As pointed out by Isenberg and Walsh (2015), since some students are linked to more teachers than others, they will have a larger contribution in estimating the coefficients on explanatory variables in Equation (4) above. Isenberg and Walsh (2015) recommend the Full Roster-plus Method, which involves creating duplicate replications. However, due to the very large number of student-teacher links in our data, this suggestion is computationally infeasible. Thus, we do not implement this method; however, results are similar when restricting the sample to remove students linked to very few or very many teachers. When the sample is restricted to students linked to between 7 and 12 teachers in a given year (about 80% of the sample), results for the quasi-experimental tests for forecast bias are similar. Finally, Isenberg and Walsh (2015) note that results using FRM and FRM+ are very similar.

outcomes at the high school level due to lack of observations. Due to the relatively short panel, we only consider the graduation outcome for teachers at the high school level. In addition, because very few students repeat grades in middle school, we do not consider grade repetition at that level.

We verify that there is a persistent component of each nontest outcome within teachers by plotting the autocorrelation vectors by outcome and school type, shown in Figure 1. The maximum years of separation is four: of our six years of data, one year must be used as prior controls in the residualization process, leaving five years of data, the maximum of which are four years apart. For teachers of elementary school students, the tested outcomes (math and reading) as well as GPA and unexcused absences have the highest year-to-year correlations, while suspensions and grade repetition tail off quickly. For middle school, all correlations tend to be higher, especially suspensions and the percent of classes failed. In high school, whether students ever graduate declines sharply as the number of years between *t* and *t-j* increases.

We next turn to properties of the VA forecasts themselves. Table 2 displays the standard deviations of the VA forecasts for each outcome for each school level. For reference, the standard deviations of VA forecasts for reading and math from Chetty et al. (2014) are also presented; our comparable measures are slightly larger than theirs. As with Chetty et al. (2014), for tested outcomes, the dispersion of teacher effects is higher in math than English and higher in elementary school than middle school. However, in contrast, many of the nontested outcomes have higher variance in middle school than elementary school. Across all grade levels, the magnitudes of the standard deviations of nontested teacher VA forecasts are slightly larger than the standard deviations of the tested outcomes (both in our data and in Chetty et al., 2014).[9]

[9] The correlations presented here are likely larger than the true correlation in underlying teacher skills due to factors such as unobserved student heterogeneity across outcomes. However, as the outcomes of interest in this paper are the forecasts of teacher effectiveness, estimating these true correlations is beyond the scope of the paper.

Even if we find that N-VAM estimates are valid and reliable, they are only useful to the extent that they provide new information relative to VAMs. As an extreme example, if VAMs for math teachers were perfectly correlated with their N-VAMs, then there would be no reason to include both measures when measuring teacher effectiveness or designing a teacher evaluation system. One of the motivations for measuring N-VAMs is the assumption that they measure, at least in part, something different than VAMs. To explore the relationship between VAMs and N-VAMs, we calculate the correlation between math and reading VAMs and each of our four N-VAMs. The higher the correlation, the larger the estimated overlap between teacher effectiveness on tested and nontested outcomes. If, for example, the correlation between value-added in absences and grade retention were high, it would provide evidence that certain teachers consistently reach their students in a manner that is reflected across multiple measures.

We then turn to the degree to which estimated N-VAMs for a given teacher are correlated. Table 3 displays the correlation across outcomes for a given teacher. Perhaps unsurprisingly, for teachers who taught both math and reading, forecasted teacher VA was highly correlated across the two subjects at nearly 0.60. At the same time, math and reading VA have a notable negative correlation with unexcused absences and suspensions. Looking specifically at the correlations among the N-VAMs, most correlations are relatively low (correlation coefficients with absolute value less than 0.20), with two notable exceptions. Both absence types (unexcused and suspension absences) are modestly correlated (0.24), and unsurprisingly, GPA and the percent of classes failed are highly negatively correlated (-0.71).

Results Using Forecasted VA

To verify that our forecasts were constructed properly, we first regress student-level residuals on the forecasted VA of their teachers:

$$\hat{Y}_{ijt} = \alpha_t + \theta\hat{\mu}_{jt} + \zeta_{ijt}$$

Results are shown in Table 4. Panel A pools all observations across all levels, and Panels B, C, and D are executed separately on elementary, middle, and high school samples, respectively. At the elementary and middle school levels, coefficients are indeed very close to 1. Although not surprising, it is still reassuring that for nontested outcomes, the outcomes of students in a given year can be predicted based on the outcomes of students taught by their teacher in other years. Conversely, the forecasts in high school are not as closely related to current student outcomes, especially for grade repetition and whether students ever graduate. This may be due to the low degree of variation in these outcomes, making them difficult to forecast.

Finally, we implement the Chetty et al. (2014) quasi-experimental estimate of forecast bias described above: changes in student outcomes at the school-grade-subject-year level are regressed on average forecasted teacher performance at that level. Results are shown in Table 5. As above, Panel A presents the results on a pooled sample, and Panels B, C, and D conduct the test by school levels. At the elementary and middle school levels, only for unexcused absences is the coefficient estimate significantly different than 1. However, due to having a short panel, standard errors are generally large and we cannot rule out somewhat large degrees of bias. Thus, rather than viewing this as a definitive test of whether N-VAMs can be thought of as valid, causal estimates of teacher effectiveness, we use this to justify measuring TFA effects on these outcomes, at least in elementary and middle school. For high school, many outcomes are far from 1 and as a result we do not estimate TFA effects for these teachers in our main results section (they are presented separately in Appendix Table 1 for completeness). Even for the high school outcomes which do not fail the test of statistically significant difference from one—classes failed and suspensions—the point estimates for bias are greater than 20%.

Taken together, our results from Tables 4 and 5 suggest that the nontest outcomes of students in elementary school and middle school can be systematically explained by the teachers to which they are assigned. The one exception is unexcused absences in elementary school, which fails the quasi-experimental test of forecast bias with a degree of bias of 50%.

TFA Estimates on Nontest Outcomes

Having validated—at least to the extent possible given our data—most nontest outcomes in elementary and middle school grades, we now move to estimating TFA effects based on a value-added specification with these nontest variables as dependent variables. Results are shown in Table 6, with columns representing separate regressions for each outcome variable. In elementary school, students in classrooms taught by a TFA teacher tended to have fewer unexcused absences, fewer days of suspension, and higher GPAs, with the latter two being at least marginally significant.[10] However, recall that unexcused absences in elementary school failed our test of forecast bias. In middle school, students in TFA classrooms continue to be less likely to have unexcused absences or suspensions (only the former is statistically significant), and the GPA effect is no longer present.[11]

In general, point estimates tend to be modest, with only one coefficient representing more than a 10% change relative to baseline values. For example, a reduction of unexcused absences in middle school by 0.347 per student corresponds to about 7% of the average of unexcused absences across all students in the sample, 4.8. When taking into account that TFA teachers tend to be assigned relatively disadvantaged students (who tend to have more unexcused absences), the percentage change in outcomes becomes even smaller. A reduction of 0.347 absences per year corresponds to only 4% of the

[10] Consistent with Ladd and Sorensen (2014), we find that teacher experience is associated with a reduction in students' unexcused absences, especially in elementary school. We do not find a relationship between teacher experience and student outcomes for the other nontest outcomes we consider.

[11] Many TFA corps members in M-DCPS are placed in schools under the direction of the Educational Transformation Office, which oversees the district's implementation of school turnaround efforts in targeted schools. Results are very similar when adding a year * ETO interaction term, which should not be surprising since in the presence of school fixed effects, estimates are identified from within-school variation.

average number of absences of students taught by TFA teachers. For GPA in elementary school, an increase of 0.05 grade points corresponds to about 1.7% of mean GPA in TFA classrooms (2.96) and about 7% of a standard deviation of GPA.

The one coefficient estimate representing a large percentage change relative to baseline values is suspensions in elementary school, where the reduction of 0.05 days per year corresponds to about one-third of the average number of days suspended for students in TFA classrooms. However, the baseline value is extremely small (the average is about 1/6, meaning that on average, one student in six is suspended one day per year) and the coefficient is only marginally significant due to the relatively large standard error, so it is hard to know whether this particular result is replicable in different data.

Overall, these results provide suggestive evidence that student behavior—as measured by days missed due to unexcused absences and suspensions—improves by a small degree when placed in a TFA classroom. In addition, students in TFA classrooms in elementary school had modest increases in GPA.

Robustness Check: Aggregate Grade-School Effects

In the presence of systematic student-teacher sorting within schools, it could be the case that the tests for forecast bias employed in this paper fail to detect bias even when such bias exists (see, e.g., Rothstein, 2014 and Horvath, 2015). Thus, we supplement our TFA estimates from Equation (3) with a model where we replace the TFA indicator measured at the school level with the share of (dosage-weighted) students in TFA classrooms at the grade-school-year level:

$$Y_{ijt}{}^{OUTCOME} = \beta_0 + Y_{ijt-1}{}^{OUTCOME}\beta_1 + X_{it}\beta_2 + TFA_{gst}\beta_3 + X_{jt}\beta_4 + X_{ct}\beta_5 + \gamma_s + \varepsilon_{ijt}. \quad (10)$$

Equation (10) does not utilize classroom-specific variation in TFA assignment, but rather the intensity of TFA presence in a given school-grade-year cell. Due to the high level of within-school variation over time induced by the clustering strategy, we are able to obtain more precise estimates for these TFA intensity variables than what would otherwise be possible under typical TFA assignment practices. Results are presented in Table 7. In general, where TFA effects were found to be associated with student outcomes in Table 6, these effects are also present when TFA is measured at the grade level. Furthermore, most of the outcomes have signs in the same direction in both Tables 6 and 7, with the exceptions being percent of classes failed (now significantly negative) and grade repetition (negative, but not significant) in elementary school.

Although the direction of the TFA effect is generally consistent across Tables 6 and 7, the magnitudes are not. Taking the estimates at face value, the results from Table 6 indicate that replacing every teacher with a TFA corps member would lower unexcused absences by 0.347 days per student, while the results from Table 7 imply that replacing an entire grade with TFA teachers would lower unexcused absences by 4.3 days per student. However, there are two important differences between Tables 6 and 7 that complicate this comparison. First, a TFA share of 1 is well outside the typical TFA density of schools in the sample: of school-grade cells with any TFA in that year, the median TFA density is about 0.1 and the 90th percentile is about 0.3. Second, Tables 6 and 7 are estimated from two different sources of variation. In Table 6, the TFA coefficient represents the average change in student outcomes associated with being in a TFA classroom relative to a non-TFA classroom in a given school. On the other hand, Table 7 compares school-grade outcomes in years with high dosage to outcomes in the same school-grade in years with low dosage. Thus, while both specifications are intended to measure the TFA contribution to student outcomes, there is little reason to expect them to have results consistent in magnitude.

Another potential explanation for the differences between Tables 6 and 7 is spillover effects, where high concentrations of TFA corps members lead to school improvements beyond their impacts in their own classrooms. In a companion paper (Hansen et al. 2015), we find little evidence of spillover in math or reading. However, one interpretation of Tables 6 and 7 is that the overall school-grade effect in Table 7 is larger than the individual effect in Table 6. In results available from the authors, we find that in a hybrid model controlling for an individual student's teacher *and* the TFA share, the coefficient on TFA share remains similar to that in Table 7, consistent with the spillover hypothesis. Of course, this finding could also be driven by possible unobserved correlates of TFA share, such as other school investments made at the same time that schools added TFA corps members.[12]

Conclusion

The analysis of N-VAMs revealed persistent differences in the influence of teacher effectiveness on nontest outcomes of their students. For all but one outcome in elementary and middle school, we cannot reject changes in student outcomes at the school-grade-subject-year level being fully explained by changes in forecasted teacher effectiveness. Although the short panel and large standard errors prevent us from making strong statements about the causality of estimated N-VAMs, most of the coefficient estimates are close enough to 1 that we feel comfortable estimating TFA effects on nontest outcomes in a value-added framework for elementary school and middle school.

The cases in which our tests reject the validity of N-VAMs also raise some caution for prior studies of teacher effects on nontested outcomes. For example, where Jackson's (2014) analysis presents evidence of teacher effects on an index of nontested student outcomes in ninth grade, we reject many of these same nontest outcomes in high school grades using our data. Although Jackson does conduct validity tests similar to those found in this paper, the tests are noisy enough to be unable to rule out substantial levels of bias (which is consistent with this paper). Another example is Gershenson's (forthcoming) study analyzing teacher effects on unexcused absences of elementary

[12] When estimating these additional specifications, we included interaction terms for year * ETO (see footnote 11) as additional control variables. Hence, any school-level investments that may bias these results would have to be separate from the ETO interventions.

school students; this is the one elementary outcome where the validity tests were rejected in our data. Although our rejection of the validity of these particular outcomes does not invalidate these prior studies' findings (neither of which use the data we use here), it does underscore the need to thoroughly vet new outcomes before using them as dependent variables in a value-added framework. Further research validating new student outcomes across a variety of contexts would be highly valuable before considering policy applications of these N-VAMs.

Returning to our primary focus on TFA effects on these outcomes, we find suggestive evidence that students taught by TFA teachers in elementary and middle school were less likely to miss school due to unexcused absences and suspensions, and that students in elementary school had slightly higher GPAs. Among the outcomes that passed the validity tests, we do not find any evidence that assignment to TFA teachers will have any adverse consequences on students.[13]

Our results stand in contrast to prior studies of TFA corps members, which have not found any significant differences in student absences and suspensions (e.g., Decker et al., 2004; Clark et al. 2013). Decker et al. (2004) report students randomly assigned to TFA classrooms averaged more days absent (by 0.52) and days suspended (by 0.04) than control students, although the result is not statistically significant, and Clark et al. (2013) find very minimal differences in student absences in elementary school and middle school. Although this study does not benefit from random assignment, our results are consistent with prior work in the sense that regardless of the direction of the estimated effect, the scope for TFA corps members affecting nontest outcomes appears to be small.

[13] The exception is the percent of classes failed for high school students, for which we find a statistically significant increase for students in TFA classrooms. However, while we cannot reject forecast-unbiasedness of this outcome due to large standard errors, the point estimate for bias in Table 5 is 22%, and in Table 4 we reject equality with one. Thus, we do not place a great deal of weight on this finding, although it deserves future exploration.

References

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement." *Education Finance and Policy,* 1(2), 176–216.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *Quarterly Journal of Economics*, *126*(4), 1593–1660.

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

Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). The Effectiveness of Secondary Math Teachers from Teach For America and the Teaching Fellows Programs. NCEE 2013-4015. National Center for Education Evaluation and Regional Assistance.

Decker, P. T., Mayer, D. P., & Glazerman, S. (2004). The effects of Teach for America on students: Findings from a national evaluation. University of Wisconsin--Madison, Institute for Research on Poverty.

Dee, T. S., & West, M. R. (2011). The Non-Cognitive Returns to Class Size. *Educational Evaluation and Policy Analysis*, *33*(1), 23–46.

Dobbie, W. (2011). Teacher characteristics and student achievement: Evidence from Teach For America. Unpublished manuscript, Harvard University.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101.

Duckworth, A. L., Quinn, P. D., & Seligman, M.E.P. (2009). Positive predictors of teacher effectiveness. *Journal of Positive Psychology*, *4*(6), 540–547.

Eskreis-Winkler, L., Duckworth, A.L., Shulman, E., & Beale, S. (2014). The grit effect: Predicting retention in the military, the workplace, school and marriage. *Frontiers in Personality Science and Individual Differences*, *5*(36), 1–12.

Garcia, E. (2014). The Need to Address Noncognitive Skills in the Education Policy Agenda (Briefing Paper No. 386). Economic Policy Institute.

Gershenson, S. (Forthcoming). Linking Teacher Quality, Student Attendance, and Student Achievement. *Education Finance and Policy*.

Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative Routes to Teaching: The Impacts of Teach for America on Student Achievement and Other Outcomes. *Journal of Policy Analysis and Management,* 25(1), 75–96.

Goldhaber, D. & Hansen, M. (2013). Is it Just a Bad Class? Assessing the Long-term Stability of Estimated Teacher Performance. *Economica*, *80*, 589–612.

Hansen, M., Backes, B., Brady, V., & Xu, Z. (2015). Examining Spillover Effects From Teach For America Corps Members in Miami-Dade County Public Schools. *CALDER WP #113*.

Heckman, J.J., Stixrud, J. & Urzua, S. The Effects of Cognitive and Noncognitive Abilities On Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 2006, *24*(3), 411–482.

Hock, H., & Isenberg, E. (2012). Methods for Accounting for Co-Teaching in Value-Added Models. Working Paper. Mathematica Policy Research, Inc.

Isenberg, E., & Walsh, E. (2015). Accounting for Co-teaching: A Guide for Policymakers and Developers of Value-Added Models. Journal of Research on Educational Effectiveness, 8(1), 112–119.

Jackson, C. K. (2014). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina (No. w18624). National Bureau of Economic Research.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City. *Economics of Education Review,* 27.6, 615–631.

Kautz, T., Heckman, J.J., Diris, R., Ter Weel, B., & Borghans, L. (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success (Working Paper No. 20749). National Bureau of Economic Research.

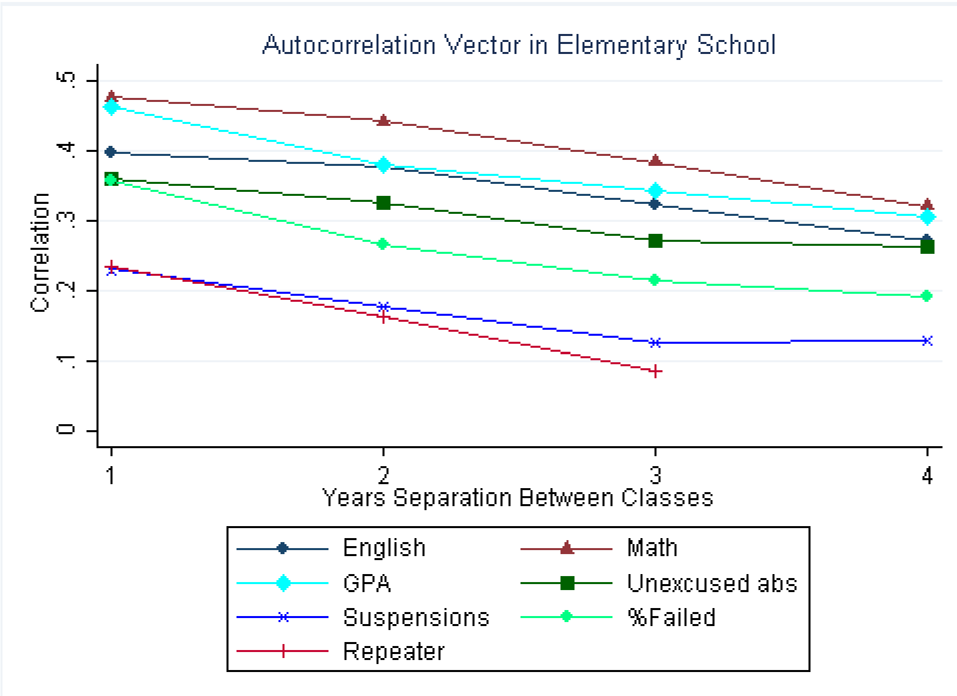
Ladd, H., & Sorensen, L. (2014). Returns to Teacher Experience: Student Achievement and Motivation in Middle School (Working Paper No. 112). National Center for Analysis of Longitudinal Data in Education Research.

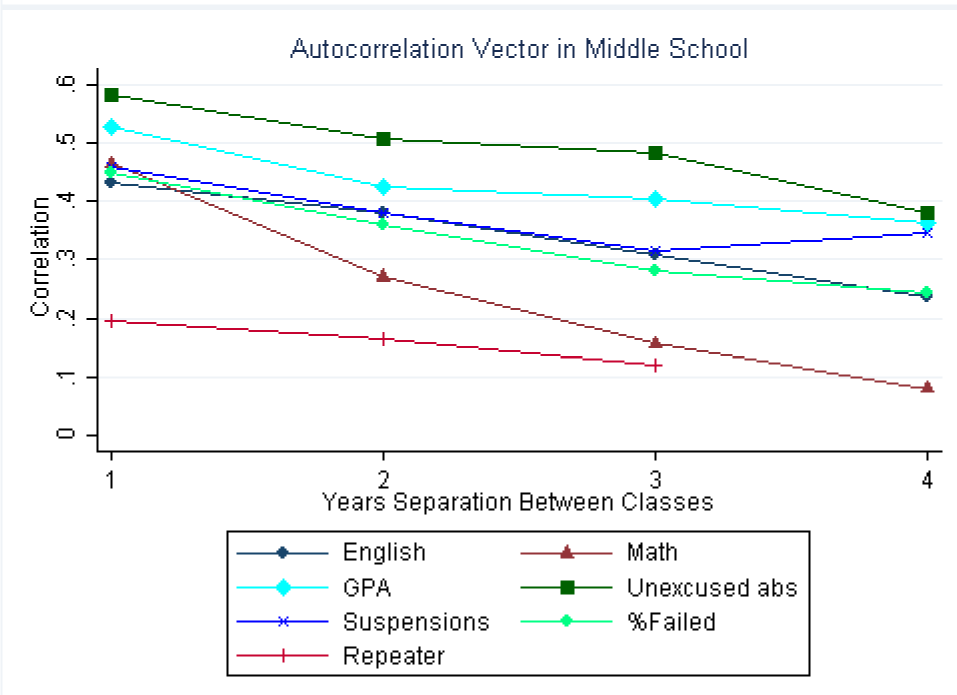
Robertson-Kraft, C., & Duckworth, A.L. (2014). True grit: Trait-level perseverance and passion for long-term goals predicts effectiveness and retention among novice teachers. *Teachers College Record*, *116*(3), 1–27.

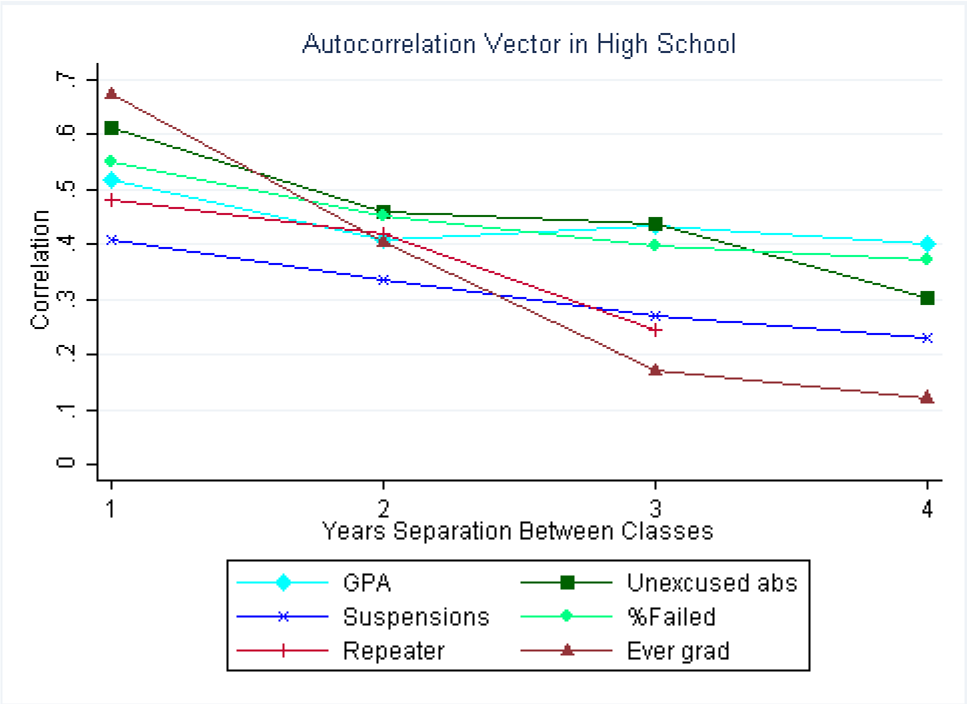
Xu, Z., Hannaway, J., & Taylor, C. (2011). Making a difference? The effects of Teach for America in high school. *Journal of Policy Analysis and Management, 30*(3), 447–469.

Figures

Figure 1. Autocovariance vectors






Autocorrelation Vector in High School
Correlation
0
.1
.2
.3
.4
.5
.6
.7
1
2
3
4
Years Separation Between Classes
GPA
Unexcused abs
Suspensions
%Failed
Repeater
Ever grad

Tables

Table 1. Active TFA Corps Member Assignments

	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14
Total TFA Corps Members	91	91	138	222	271	290
Total schools containing any TFA Corps Members	49	34	23	23	30	37
TFA as proportion of school teachers by school type, conditional on containing TFA						
Elementary	3.6%	4.4%	8.6%	20.4%	13.8%	11.8%
Middle	1.4%	7.6%	8.5%	16.9%	16.9%	13.6%
High	1.7%	4.0%	13.6%	15.9%	14.9%	12.0%

Note: Proportions of schools teachers by school type are calculated among any schools containing any TFA corps members during that school year.

Table 2. Standard Deviation of VA Forecasts by School Type

	Elem	Middle	High
CFR: Math	0.12	0.09	
CFR: English	0.08	0.04	
Math	0.15	0.12	.
ELA	0.10	0.08	.
Abs-Unex	0.11	0.16	0.17
Susp	0.09	0.15	0.11
Pct-Failed	0.14	0.15	0.16
GPA	0.17	0.17	0.16
Repeater	0.10	.	0.26
Grad	.	.	0.37

Notes: CFR values are provided from Chetty et al. (2014) for comparison.

Table 3. Cross-subject Correlation of VA Forecasts by School Type

	Math	ELA	Abs un	Susp	GPA	%failed	Repeat
ELA	0.59						
Abs un	-0.17	-0.32					
Susp	-0.17	-0.24	0.24				
GPA	0.09	0.11	-0.07	-0.20			
%failed	-0.00	0.02	0.07	0.17	-0.71		
Repeat	-0.09	-0.07	0.10	0.03	-0.02	0.08	
Grad	-0.16	0.11	-0.11	-0.06	0.12	-0.11	-0.08

Table 4. Coefficients from Student-level Regression of Outcome on VA Forecast

	Tests	M	ELA	Abs U	Susp	GPA	%Failed	Repeat	Grad
Panel A: All	1.01	1.01	1.02	1.04	1.05	1.00	1.04		
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)		
Reject = 1							x		
Panel B: Elem	1.01	1.00	1.02	1.01	1.05	0.96	0.99	0.95	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.08)	(0.01)	(0.02)	(0.04)	
Reject = 1						x			
Panel C: Middle	1.02	1.01	1.02	1.02	1.04	0.99	1.03		
	(0.03)	(0.04)	(0.02)	(0.03)	(0.04)	(0.02)	(0.03)		
Reject = 1									
Panel D: High				1.06	1.06	1.05	1.08	0.92	0.29
				(0.03)	(0.05)	(0.03)	(0.02)	(0.06)	(0.02)
Reject = 1				x			x		x

Notes: Sample sizes not provided because they are very large since the unit of observation is the student-teacher link.

Table 5. Quasi-Experimental Estimates of Forecast Bias

	Tests	M	ELA	Abs U	Susp	GPA	%Failed	Repeat	Grad
Panel A: All	0.93	0.86	1.20	1.12	0.92	1.23	0.94		
	(0.07)	(0.08)	(0.16)	(0.09)	(0.12)	(0.07)	(0.08)		
	8689	4306	4383	11435	11435	10029	10140		
Reject = 1						x			
Panel B: Elem	0.98	0.89	1.27	0.50	0.88	1.10	1.00	0.88	
	(0.10)	(0.12)	(0.18)	(0.09)	(0.17)	(0.06)	(0.08)	(0.13)	
	4857	2410	2447	7737	7737	6406	6459	7737	
Reject = 1				x					
Panel C: Middle	0.88	0.83	1.10	1.04	0.80	0.83	0.88		
	(0.10)	(0.10)	(0.29)	(0.16)	(0.19)	(0.12)	(0.19)		
	3832	1896	1936	2056	2056	2010	2056		
Reject = 1									
Panel D: High				1.50	1.24	1.88	0.78	2.49	0.07
				(0.17)	(0.36)	(0.25)	(0.16)	(0.41)	(0.05)
				1642	1642	1613	1625	1642	1101
Reject = 1				x		x		x	x

Notes: see text.

Table 6. Relationship Between TFA and Nontest Outcomes

	Unexc abs	Susp abs	GPA	Percent Failed	Repeater
Elementary	-0.044^{χ}	-0.054*	0.050**	0.001	0.001
	(0.139)	(0.032)	(0.023)	(0.003)	(0.005)
Observations	4661344	4661344	4622273	4661344	4661344
R-squared	0.438	0.127	0.645	0.242	0.115
Dep var mean, full sample	4.0	0.07	3.19	0.028	0.026
Dep var mean, students in TFA classrooms	7.2	0.17	2.96	0.041	0.038
Middle	-0.347**	-0.075	-0.009	0.001	
	(0.154)	(0.058)	(0.011)	(0.002)	
Observations	3174990	3174990	3159288	3174990	
R-squared	0.488	0.301	0.659	0.266	
Dep var mean, full sample	4.8	0.79	2.64	0.038	0.012
Dep var mean, students in TFA classrooms	8.4	2.01	2.29	0.067	0.022

Notes: Regression controls for student-level and class average demographics and their interactions with grade. Other controls include class size and teacher race and their interactions with grade.

χ: unexcused absences in elementary school fails the forecast bias test (see Table 5). We display the coefficient here for completeness but urge caution in interpreting this result.

Table 7. TFA at Grade-school-year Level and Nontest Outcomes

	Unexc abs	Susp abs	GPA	Percent Failed	Repeater
Elementary	-0.571^{χ}	-0.179	0.301***	-0.030**	-0.018
	(0.654)	(0.109)	(0.087)	(0.012)	(0.016)
Middle	-4.300**	-0.189	-0.063	-0.005	
	(2.080)	(0.623)	(0.126)	(0.031)	

Notes: Coefficients displayed are on the share of TFA at the grade-school-year level. Regression controls for student-level and class average demographics and their interactions with grade. Other controls include class size and teacher race and their interactions with grade.

χ: unexcused absences in elementary school fails the forecast bias test (see Table 5). We display the coefficient here for completeness but urge caution in interpreting this result.

Appendix Table 1. Relationship Between TFA and Nontest Outcomes in High School

	Unexc abs	Susp abs	GPA	Percent Failed	Repeater	Graduate
High school	0.126	0.007	-0.025	0.008**	0.003***	-0.002
	(0.204)	(0.028)	(0.018)	(0.004)	(0.001)	(0.008)
Observations	4242829	4242829	4226106	4242791	4242829	4242829
R-squared	0.512	0.175	0.577	0.285	0.132	0.622
Dep var mean, full sample	7.59	0.53	2.57	0.08	0.028	0.68
Dep var mean, students in TFA classrooms	12.3	1.14	2.33	0.12	0.012	0.59

Notes: Regression controls for student-level and class average demographics and their interactions with grade. Other controls include class size and teacher race and their interactions with grade.

As all outcomes shown here have an estimate of forecast bias of greater than 20% (see Table 5); **these estimates should not be taken as credible estimates of TFA effectiveness**.