



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



*Performance Estimates of
Teach For America Teachers
in Atlanta Metropolitan
Area School Districts*

Michael Hansen
Tim R. Sass

Performance Estimates of Teach For America Teachers in Atlanta Metropolitan Area School Districts

Michael Hansen
American Institutes for Research

Tim R. Sass
Georgia State University

Contents

Acknowledgements	2
Abstract.....	3
Introduction.....	4
TFA in the Atlanta Region.....	6
Data	8
Sample.....	12
Methods.....	13
Results	15
Filtering out suspect observations to remove the influence of cheating	17
Discussion	21
Conclusion	23
References.....	25
Tables	27
Figures.....	35

Acknowledgements

This evaluation was funded in part by Teach For America, 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.

This report benefited from the cooperation and input of TFA personnel, including Raegen Miller, Jason Atwood, and Rachel Perera; however, TFA had no editorial privilege over the final written report. We acknowledge the cooperation of district personnel from Atlanta Public Schools, Clayton County Public Schools, and Gwinnett County Public Schools in providing access to the longitudinal data for this evaluation. We also acknowledge excellent research assistance from Jeffrey Nelson, Jarod Apperson, and Carycruz Bueno; Dan Goldhaber and Ben Backes provided helpful comments. Any remaining errors are the responsibility of the authors.

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.

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

Performance Estimates of Teach For America Teachers in Atlanta Metropolitan Area School Districts

Michael Hansen & Tim R. Sass
CALDER Working Paper No. 145
October 2015

Abstract

Teach For America (TFA) is an alternative certification program that intensively recruits and selects recent college graduates and midcareer professionals to teach in schools serving high-need students. Prior rigorous evaluations of the program have generally found positive effects of TFA teachers on students' learning in math and science and no significant differences in reading or language arts, compared with non-TFA teachers' effects in the same schools. No prior studies, however, have specifically focused on TFA effects in the Atlanta region.

This report examines the efficacy of TFA teachers in the Atlanta region spanning the 2005-06 through 2013-14 school years. Using longitudinal administrative data from three major school districts with significant numbers of recent TFA placements, we generate TFA effect estimates based on two series of Georgia's standardized tests—the end-of-grade Criterion-Referenced Competency Tests (CRCTs) and end-of-course tests (EOCTs).

We find evidence of a positive effect in student learning due to the hiring of TFA teachers in these three districts, compared with the performance of non-TFA colleagues in the same schools. Estimated TFA effects are positive and statistically significant in social studies and science on the state's CRCTs, and in American literature on the state's EOCTs. We find no significant differences in performance between TFA and non-TFA teachers in the other subjects we analyzed. Supplementary analyses show these results are not sensitive to the inclusion of data from a period of well-documented test score manipulation in Atlanta Public Schools.

Introduction

Teach For America (TFA) is an alternative certification program that intensively recruits and selects recent college graduates and mid-career professionals to teach in schools serving high-need students. Corps members commit to working in these schools for a period of two years, and during this period TFA supports corps members with instructional coaching and professional development.

Prior evaluations of the program have compared the performance of its corps members and alumni on students' learning gains (measured by standardized tests) relative to non-TFA teachers' student's learning gains in the same schools. Using both randomized experiments (Clark et al., 2013; Clark, Isenberg, Liu, Makowsky, & Zukiewicz, 2015; Glazerman, Mayer, & Decker, 2006) and rigorous quasi-experimental methods ((Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006; Hansen, Backers, Brady, & Xu, 2015; Kane, Rockoff, & Staiger, 2008; Xu, Hannaway, & Taylor, 2011), these studies find TFA teachers as a group commonly outperform non-TFA teachers in the same schools in math¹ (and science, where tests are available). In general, these studies find TFA teachers do not show significantly different performance in reading or language arts.^{2, 3}

This report examines the efficacy of TFA teachers (referring to both active corps members and TFA alumni) in the Atlanta region spanning the 2005-06 through 2013-14 school years. No prior studies

¹ Two of these studies do not find statistically significant gains in math attributable to TFA teachers: Boyd et al. (2006) and Clark et al. (2015) both estimate positive coefficients for TFA teachers in math, but the full-sample estimates are not statistically significant. Both studies, however, produce sub-sample estimates that do find statistically significant gains attributed to TFA teachers. A sub-sample of first-year middle school TFA math teachers relative to other beginning teachers in the Boyd et al. study show gains of about 0.05 standard deviations higher for TFA teachers in math. In the Clark et al. study, a sub-sample of lower elementary school students (PK to grade 2) estimates TFA effect sizes of 0.12 standard deviations in reading.

² An exception to the general finding of null effects in reading or language arts is that from Hansen et al. (2015), which estimates small, though statistically significant, TFA effects of 0.02 standard deviations in reading. This study uses data from Miami-Dade County Public Schools, covering grades 4 through 12 over a period of six school years.

³ Though rigorous evidence generally supports the relative efficacy of TFA teachers in math and science, other authors have been critical of TFA for various reasons, including insufficient preparatory training and low retention of corps members in schools after the two-year commitment period ends (Darling-Hammond et al., 2005; Miner, 2010).

have specifically focused on TFA impacts in the Atlanta region—the prior TFA studies referenced above were conducted either in different locales or have focused on multi-site, national samples. Hence, it is unclear whether TFA will show similar impacts in Atlanta-area schools. Using longitudinal administrative data from three major school districts in the Atlanta region that hired TFA corps members during this period, we generate estimates of TFA teachers’ classroom performance relative to other teachers in the same schools. We generate these estimates using two sets of state standardized tests as outcomes: the Criterion-Referenced Competency Tests (CRCTs), administered annually to students in Grades 3 through 8 assessing performance on five subjects; and End-of-Course Tests (EOCTs), administered to all students completing particular courses typically taken in high-school grades (six subject tests are used in this analysis).

In summary of our overall findings, we find evidence of either a positive or neutral impact on student learning across subjects due to the hiring of TFA teachers in these three districts. The positive TFA effects are clearest in social studies and science on the state’s CRCTs, estimated respectively to be 0.095 and 0.115 standard deviations of student achievement using the preferred school fixed effects specification. Although there is some evidence of TFA teachers performing favorably on several of the EOCT subject tests analyzed here, we only find statistically significant differences in teacher performance on American Literature of 0.130 standard deviations of achievement when comparing between TFA and their non-TFA colleagues in the same school. Supplementary analyses show the magnitude of these estimates are not sensitive to the inclusion of data from a period of well-documented test score manipulation in Atlanta Public Schools.

TFA in the Atlanta Region

TFA has been placing corps members in schools in the Atlanta metropolitan area since 2000. Since then, a cumulative total of over 3,000 corps members have been placed in schools in the region. TFA corps members and program alumni work in a variety of schools in the metropolitan area, including both traditional public and public charter schools, as well as serving a variety of student populations in terms of poverty levels and demographic composition.⁴ Active TFA corps members, however, are placed to teach primarily in the region's highest poverty, lowest performing schools.⁵

Historically, the placement of TFA corps members in the metropolitan area has been most heavily concentrated in Atlanta Public Schools (APS), although placements have become more diffuse across other districts in recent years. Table 1 reports regional placement data spanning the 2008-09 to 2013-14 school years, made available from TFA. Table 1 shows 145 or more corps members teaching in APS (constituting 67 to 75 percent of the active corps) for each of the first three school years in this period. In more recent years, APS placements account for 30-35 percent of the active corps in the region, but this drop in representation is due more to an expansion of the corps in other districts in the region than it is to a decline in APS placements. For example, in the 2011-12 school year, TFA placed a large cohort of 64 corps members in Gwinnett County Public Schools (GCPS) where there had been none previously, and Clayton County Public Schools (CCPS) nearly tripled its active corps of TFA in the district from the year prior (going from 18 to 52 corps members). The size of the regional corps has fluctuated during this period, starting at 247 corps members, peaked at 429 in 2012-13, and numbered 292 corps members as of 2013-14.⁶

⁴ In some exploratory investigations, we estimated the TFA impact among active corps members only (flagging alumni as non-TFA) and did not find any qualitatively different results from those reported here. These results are discussed in more detail in the Discussion section below.

⁵ Comparisons of TFA-led versus non-TFA-led classrooms are presented in the Sample section (Tables 4 and 5).

⁶ The surge in corps members in the region during these recent years was part of a broader program-wide surge in the national TFA corps, induced by an Investing in Innovation (i3) funding award from the U.S. Department of

One concern about studies that utilize data from Atlanta is the test-score manipulation scandal under the leadership of APS Superintendent Dr. Beverly Hall (Georgia Bureau of Investigation, 2011). Indeed, several of the early school years that we analyze for this study coincide with the period of widespread testing improprieties during the mid- to late-2000's.⁷ A recent report from Sass, et al. (2015) investigates the patterns of cheating in APS based on erasure analyses conducted by the state's test vendor. The authors find evidence of discrete declines in the distribution of student scores between the 2009 and 2010 test administrations (when monitors were present in suspect schools) for those who were in classrooms flagged for high levels of wrong-to-right (WTR) erasures. These patterns are suggestive that widespread test-score manipulation was occurring in the district in 2009 (and in the years leading up to 2009, according to Georgia Bureau of Investigation, 2011), though suspect behavior was largely neutralized in 2010 and beyond. Sass, et al. (2015) also investigate whether there are any long-term effects for students whose test scores were manipulated in the scandal, and find deleterious effects in reading and English Language Arts (ELA) but mixed results in math; no significant differences were detected on non-test student outcomes including attendance and behavioral infractions.

The high concentration of TFA teachers in APS during this era may consequently cloud our ability to identify the impact of TFA teachers on students using suspect data. To address concerns that our results may be biased by the presence of test score manipulation, we present a range of TFA estimates

Education to scale-up the program.

⁷ The Atlanta Journal Constitution first published evidence of inexplicably large gains in Georgia schools based on the 2008 results of the state's CRCT, and updated results from the 2009 CRCT showed similar results highlighted that the majority of the schools posting the largest gains were in APS. Subsequently, the Governor's Office of Student Achievement (GOSA) had an erasure analysis performed by CTB McGraw-Hill to flag schools where testing irregularities were highest; over half of those considered most severe were in APS. Following this, the governor deployed a special investigative team, led by the Georgia Bureau of Investigation (GBI), to uncover what was going on in APS. In June 2011, the GBI published the findings in a report (Georgia Bureau of Investigation 2011), which concluded cheating had occurred in 44 of the 56 APS schools that were targeted for investigation. The report identifies 178 educators guilty of cheating, including 38 school principals; it also "suggests there were far more educators involved" (p. 2) but the evidence was insufficient to identify them by name in the report. Test score manipulation is recognized as having occurred over several years in APS peaking in the 2009 test administration, and was reduced considerably on the 2010 tests when state monitors were placed in select schools under scrutiny around cheating allegations. Superintendent Hall resigned in 2010, and criminal charges brought against former APS educators recently resulted in 11 convictions (Blinder, 2015).

that use a variety of strategies (based on the identification strategies laid out in Sass et al., 2015) to circumvent the issue. We find the magnitude of our TFA estimates are generally robust across these strategies, and we are confident that the TFA estimates presented here are not driven by the use of suspect data.

Data

Longitudinal administrative data were obtained from three Atlanta-area districts for this study—APS, CCPS, and GCPS.⁸ The data span the 2004-05 through 2013-14 school years, although the range of years varies across districts based on the number of TFA placements in the district and readily available data. The first year of data from each district is used for pretest scores only. The range of years included for each district is described in further detail below.

The outcome of this study is student achievement on state standardized tests, controlling for past test performance. Student achievement is assessed using two series of tests: the Criterion-Referenced Competency Tests (CRCTs) and End-of-Course Tests (EOCTs). The state of Georgia has administered the CRCT in reading, ELA, math, science and social studies since the spring of 2004. Until the 2010-11 school year the tests were administered annually in Grades 1-8, and since then have only been given in Grades 3-8. To maintain a consistent range of grades throughout the sample, we do not include any test scores from Grades 1-2 and include Grade 3 test scores as pretests only; the final CRCT sample covers Grades 4-8.

Next, EOCTs have been given to students of particular focal courses typically taken during high school (although advanced students may take these courses in earlier grades). Although there are eight different subject-specific EOCTs that have been offered during recent years, some have been changed or

⁸ We also requested data from both Fulton County Public Schools and DeKalb County Public Schools, the other two school systems in the region that have hired significant numbers of TFA corps members in recent years. Both districts declined to participate in the study.

replaced by other course tests. We chose to focus on the EOCTs that have been consistently offered to all students between the 2009-10 and 2013-14 school years in the six following courses: Math 1, Math 2, 9th Grade Literature, American Literature, Biology, and U.S. History.⁹ All raw test scores, both the CRCTs and EOCTs, are converted into z-scores relative to the mean and standard deviation of the test-taking population in the state for the year and subject of test administration (and grade, in the case of CRCTs).

In addition to standardized test scores, the data files record a variety of student characteristics: race/ethnicity, gender, free- or reduced-price lunch (FRL) eligibility, limited English proficiency (LEP) status, and whether a student is flagged as having a mental, physical, or emotional disability. Enrollment files are used to create flags for students participating in special education courses and those who move across schools during the year since the pretest (for moves unrelated to typical grade promotion). Many of these variables are included as explanatory variables in the estimating equation used in this analysis.

Teacher personnel files in the districts' data contain information on teachers' experience (or hire dates in the district, from which experience values are calculated). These values are likewise used as covariates in the analysis that follows. An important element is the necessity to identify TFA teachers appearing in the administrative data. TFA created a master list of known TFA-affiliated teachers in the region to provide to districts to enable them to create a TFA indicator variable in the data. This list included two groups of teachers: 1) all corps members ever placed in the Atlanta region; and 2) TFA alumni who were placed in a different region of the country as corps members but started teaching in the Atlanta metropolitan area after their time in the corps.¹⁰ Data analysts in the district offices created the TFA flag based on matching teachers' identifying information from these lists.

⁹ The Math 1 EOCT was not administered in 2013-14, though we only have data from one district (APS) for this school year. All other subject EOCTs are available for this year.

¹⁰ TFA administers an annual Alumni Survey to track the career paths of former corps members. Among other things, this survey captures whether alumni continue to teach and where they move. Information from this survey is used to identify alumni of other regions who report teaching in the Atlanta region.

An analysis of the impact of teachers on student performance requires linking students to their test scores and to the teachers responsible for their instruction in the relevant subject. All districts provided course-level files that link students to teachers, and students in core subject courses are linked to their teachers. We calculate teacher dosages for students linked to multiple teachers or multiple courses in the same subject based on the number of teachers and time of exposure across the various teachers. These dosage values are used as analytical weights in the regressions that follow below, as described in Hock and Isenberg (2012).

Not all TFA teachers in these three districts are represented in the samples constructed for the analysis. Teachers do not appear in the sample if they are not assigned to a CRCT tested grade or an EOCT tested subject, or if course records linking students and teachers are incomplete. In each of the three districts, TFA teachers included in at least one of our analysis samples account for more than 50% of known TFA teachers in the district, although this varies across districts. In APS, 729 unique TFA corps members and alumni are identified in administrative files spanning the study period and 374 (51%) are included in the analysis files (slightly more than one half of teacher-year observations are from years as active corps members, the remainder are from alumni). In CCPS, only teacher files from tested grades and subjects were provided (thus we cannot obtain a direct count of TFA outside of these grades and subjects); however, according to TFA's placement figures for the included years of data, approximately 80 unique TFA were placed in the district during these years, and 51 (64%) are represented in the analysis samples (more than 90% of teacher-year observations are from active corps members). In GCPS, 44 of 64 placed TFA corps members (69%) are represented (all are active corps members, and no alumni are identified in GCPS).

Table 2 shows the time span of each district's data on the state CRCTs and the count of unique teacher-year observations of linked TFA teachers represented in each of these districts. Two columns

are associated with APS in this table, the first column reports all data provided by the district, the second column reports the data considered clean from the cheating scandal (i.e., data since the 2010-11 school year).¹¹ In all subjects, TFA teachers in APS constitute the large majority of all TFA teachers observed across these districts. This is due not only to the relatively large share of TFA corps members placed in APS each year but also due to the historically heavy concentration of TFA in APS, which implies a greater concentration of TFA alumni in these schools as well.

Table 3 represents the analogous information for the EOCT data files. Unlike in Table 2, APS only has one column in this table representing what is considered to be data clean from cheating because none of the EOCTs were implicated in the APS test manipulation scandal. Moreover, all of the EOCTs used for the analysis postdate the scandal.¹² Also, note that the number of unique TFA teachers represented in these samples is far smaller than the numbers in Table 2; this is because the EOCT files have far fewer student observations overall (about one tenth of the size of the CRCT samples) and teachers in these samples commonly teach multiple sections of the same course. As an example, where a typical teacher in the CRCT sample will teach a single class of only 25 students, a typical teacher in the EOCT sample will teach three or four classes of 25 students. The smaller number of teachers is noteworthy because we have fewer teachers on which to estimate TFA impacts.¹³ As in the CRCT sample, most TFA teachers in the data samples teach in APS.

¹¹ As stated above, widespread cheating is generally considered to have been removed by the 2010 CRCT administration. However, since we wish to include non-tainted pretest scores in our analysis of student growth, we consider the 2011 CRCT administration (using 2010 CRCTs as pretest scores) as the first year of clean scores that can be used for our analysis.

¹² Though not all students' pretests postdate the cheating scandal; we conduct a set of strategies to filter out suspect pretest scores from EOCT observations similar to what we do for the CRCTs.

¹³ The small number of unique TFA teachers in CCPS and GCPS mean district-specific TFA impact estimates on the EOCTs are very noisy, and we do not report them as we do with the CRCT sample.

Sample

The sample for the CRCT analysis is limited to students in Grades 4 through 8 who have valid test scores in the district-year combinations shown in Table 2. Estimation of the value-added model specification we employ requires a pretest score in the same subject for all students, for which we use the CRCT score of the prior grade in the prior year. Prior test scores in other subjects are additionally used as covariates in the regression model; however, if a student is missing a prior test score in another subject (but not the same subject as the outcome variable) the missing value is imputed with multiple imputation based on the values of covariates that are observed. Students who skip or are retained in a grade are dropped from the sample due to the unusual grade progression.¹⁴

The EOCT samples consist of students who are taking the course, regardless of the grade of the student. As above, all students are required to have a pretest score in the same subject to be included in the sample. For EOCTs, the same-subject pretest comes from the student's most recent CRCT scores. Since most EOCT observations are from high-school students, the most recent CRCT score is typically from 8th grade, though students who take EOCTs early have their prior grade CRCT results as pretests (e.g., an 8th grader taking Math 1 uses 7th grade CRCT results as pretests). A variable documenting the number of years since the CRCT assessment is used as an additional control variable, as the time since CRCT pretest varies across students.

Summary descriptive statistics for the classrooms of TFA teachers in the CRCT analysis sample are presented in Table 4. Table 5 presents statistics for the EOCT samples. As these tables show, the classrooms of TFA teachers are very different from typical classroom assignments of non-TFA teachers. TFA teachers' students are significantly more likely to be black and FRL eligible, and show significantly lower levels of prior achievement on standardized tests. By design, TFA corps members are typically

¹⁴ These dropped observations accounted for 1.1%, 0.5% and 0.02% of the unique student-year-subject test score observations in APS, CCPS, and GCPS, respectively.

placed into the lowest performing or highest poverty schools in a district; hence, these differences in classroom composition are expected.

Also we note, based on Tables 4 and 5, TFA teachers are significantly less experienced than non-TFA teachers in the sample. This is expected because active TFA corps members generally enter with no prior teaching experience, and these teachers constitute the majority of TFA observations in all three districts. Some alumni are also included, but even these observations tend to be at the earlier stages of their careers: in APS, the only district with a substantial population of alumni, 80% of alumni observations have five or fewer years of teaching experience; the remaining 20% are from more experienced TFA alumni (5% of alumni have at least 10 years of experience).

Methods

The regression model used to estimate TFA effects in the CRCT samples is represented as follows:

$$A_{ist} = \alpha I^{Grade} A_{it-1} + \beta_1 I^{Grade} X_{it} + \beta_2 I^{Grade} C_{it} + \beta_3 I^{Grade} EXP_{it} + \beta_4 TFA_{it} + I^{Grade} \tau_t + \tau_t \delta_d + \gamma_s + \varepsilon_{ist}$$

This equation predicts student achievement for student i in school s in classroom c at time t on test scores (A_{ist}) as a function of prior student achievement (A_{it-1}), student characteristics (X_{it}), classroom characteristics (C_{it}), a teacher's experience (EXP_{it}) an indicator variable for TFA active corps members and alumni (TFA_{it}), a grade-by-year fixed effect ($I^{Grade} \tau_t$), a district-year fixed effect ($\tau_t \delta_d$), and a school fixed effect (γ_s). The vector of prior student achievement includes a squared expansion of pre-test scores in the prior-year CRCT scores for all five subjects for each student. The vector of student characteristics includes variables for student race/ethnicity, gender, eligibility for free or reduced-price lunch (FRL) status, switching schools within the past year, special education and disability status, and

English Language Learner (ELL) status. The vector of classroom characteristics includes the percentages of the FRL-eligible, black, and Hispanic students. Teacher experience is entered as a series of indicator variables (representing 1 year, 2 years, 3-4 years, 5-9 years, and 10+ years of experience; novice teachers are the omitted category). Note that the pretest, student, classroom, and teacher experience vectors are interacted with grade-specific variables (I^{Grade}), allowing each covariate to have a grade-specific association with test scores.¹⁵ Due to small numbers of TFA teachers in each grade, computing an overall TFA impact that pools effects across all grades is the most feasible strategy to obtain estimates with acceptable power (this is the same strategy used in most other non-experimental TFA evaluations; e.g., Kane et al., 2008).

The inclusion of school fixed effects (γ_s) in the model is preferred here because the contexts in which TFA teachers are assigned are so different from non-TFA teachers in the sample (as demonstrated in Tables 4 and 5). TFA are assigned to high-poverty students who may have unobservable tendencies to perform poorly on standardized tests. Failure to include school fixed effects would compare the performance of students assigned to TFA teachers relative to all other teachers, which would likely bias the TFA effect downwards due to these unobservable factors. Including school fixed effects compares TFA teachers against non-TFA teacher in the same school assigned to students with similar observable characteristics, which should remove any bias due to the context of TFA assignments.¹⁶

The EOCT models largely parallel those described in the regression equation above, with a few small modifications. Because the students in the EOCT files may take the course in a range of grades, we include the number of years since the CRCT pretest scores were earned as an additional control in the

¹⁵ Due to documented differences in effect sizes across tested grades (Hill et al., 2008), interacting grade with each of the covariates is our preferred strategy for this analysis, where the sample spans multiple tested grades. This ensures within-grade (and not across-grade) covariation on explanatory variables is factored into the results.

¹⁶ The inclusion of school fixed effects removes any across-school bias of unobserved variables, though within-school bias could still remain in the data if principals systematically assign students to TFA teachers in a non-random way based on characteristics not observed in the data.

vector of prior student achievement. Also, multiple grades take the same EOCT by construction, and we therefore do not interact grade indicators with the covariates or year fixed effects in these models.¹⁷

Results

Table 6 presents the TFA impact estimates based on the full CRCT samples across all years and districts. Results are presented by subject, with two columns for each subject corresponding to an ordinary least squares (OLS) model and a school fixed effects model. The estimated coefficients are generally not statistically different from zero in math, reading, or language arts, with the exception of a negative point estimate in reading under the OLS model; none of the preferred fixed effects models show statistically significant estimates in these three subjects. On the other hand, the results in science and social studies show very large and significant effects attributable to TFA teachers, equal to 0.095 standard deviations in science and 0.115 standard deviations in social studies.¹⁸ These magnitudes are not significantly different from the performance advantage estimated for teachers with two or more years of experience in these samples (compared against a novice teacher).¹⁹

The null estimates on reading and language arts are not surprising, given the plurality of studies that find similar results in these subjects. The one study that has investigated science tests as an outcome (Xu et al., 2011) similarly finds a large and significant effect, which is consistent with the results here.

¹⁷ By contrast in the CRCT, students in separate grades take separate tests, and the regression model combines all observations across tests by estimating grade-specific covariates (implicitly resulting in a set of covariates corresponding to each tested sample of students). Since the EOCT sample is the full sample of tested students, we do not want to difference out grade-specific differences.

¹⁸ We would like to convert these estimates into month of learning, a common approach that facilitates interpretation of the findings. We cannot do this with the science and social studies CRCTs, however, because the tests are not vertically aligned to represent common units across grades, nor do we have estimates of expected one-year gains in these subjects.

¹⁹ As described previously, we categorize teachers' experience into bins representing 1, 2, 3-4, 5-9 and 10 or more years of prior experience. In both of the school fixed effects specifications for science and social studies, differences in performance are not statistically significant across groups with two or more years of experience. The TFA coefficient estimates also fall within the confidence intervals of these experience estimates.

The two surprises in this table of results are the estimates on math and social studies. First, most (though not all) prior studies of TFA teachers' effectiveness estimate statistically significant gains in math, in the range of 0.02 to 0.15 student standard deviation units depending on the grade span and study. The math point estimates presented here (0.003 standard deviations) fall below this range. Second, no prior studies that we are aware of have generated TFA impact estimates in social studies, and given the common association of TFA's effectiveness in the hard sciences, we are surprised to see point estimates in this subject that are larger than both math and science.

To further explore what may be driving the results we find, we estimated the regression model separately by each district and again separately by school level (elementary vs. middle grades). Figures 1 and 2 present the results of this exploration and illustrate the TFA coefficient estimate in each of the CRCT subjects corresponding to the district- or level-specific models. In these figures, solid bars represent statistically significant estimates (with a p-value < 0.05). Figure 1 shows the TFA impact in math achievement is significantly large at the expected magnitude in GCPS, but estimates are virtually zero in APS and CCPS. Impact estimates vary across districts in science and social studies, but are all positive and statistically significant. Recall that due to the high concentration of TFA teachers in APS, the overall estimates in all subjects are heavily weighted by the APS point estimates, which are near zero in all subjects but science and social studies. One possibility is that these estimates in APS may be related to cheating; we return to this issue in the next section below.

Figure 2 illustrates the TFA coefficient estimates in the five tested CRCT subjects separately by elementary and middle grade levels. Note in contrast to Figure 1, all districts' data are now combined. This figure shows the positive impact estimates in science and social studies are driven overwhelmingly by teachers in middle grades, with elementary grades contributing significantly less to these outcomes.

Looking to the math impacts, the estimated middle school effect is modestly larger here; however, the magnitude of the estimate is still less than 0.02 standard deviations.

The results of the EOCT regression models are presented in Table 7. In Math 2, American Literature, and Biology, the OLS estimates are statistically significant; however, only the American Literature sample shows statistical significance of TFA teachers in the school fixed effects models. The theoretical motivation for the inclusion of school fixed effects in these models is to avoid a downward bias on the TFA estimate due to unobservable tendencies for low performance that may be unique to schools where TFA are placed. In most studies of TFA efficacy, the point estimates on TFA increase with the inclusion of school fixed effects in the regression model.²⁰ Yet, the results in Table 7 indicate smaller point estimates in the school fixed effects for Math 1 and Biology, which runs counter to prior empirical evidence. In other words, the school contexts or comparison teachers in schools with TFA teachers in the Atlanta region appear to be, on average, slightly higher performing than expected among the sample generally. We return to this issue in the Discussion section below.

Filtering out suspect observations to remove the influence of cheating

In light of the known presence of cheating in the data, we test the sensitivity of our findings to the removal of students whose test scores may have been inappropriately manipulated. This involves adjusting our analysis samples for both suspicious pre- and post-test scores during this period. To aid in filtering out suspect scores, we use student-level erasure data provided to APS by the test vendor spanning the 2008-09 through 2012-13 school years.²¹ We employ four approaches to removing suspicious data, each of which is discussed below.

²⁰ For example, see Table 4 of Hansen, et al. (2015), which shows slightly larger TFA point estimates on the fixed effects models compared to the OLS models.

²¹ The erasure data during the first three school years of this time period are only partial erasure samples among schools deemed suspect in the scandal (and the sample of included schools shifts over these years); the last two

The first, and most extreme, approach removes all data from tainted test years in APS. Recall cheating occurred in the years leading up to and including the 2009 CRCT administration, and the years since then are generally considered free of widespread cheating.²² Though the 2010 post-test scores are not suspect, observations from the 2009-10 year are suspect due to the pretest scores. This approach, therefore, drops all APS observations prior to the 2009-10 school year, consistent with the modified analysis sample documented in the second Atlanta column of Table 2. This considerably limits the number of TFA teachers in the sample, but maintains the greatest degree of separation from the stain of cheating in the sample.

The second approach removes all students who had suspiciously high numbers of WTR erasures in the 2009 test administration. Following Sass, et al. (2015), we use the distribution of WTR erasures in the 2013 administration of the CRCT to represent a clean administration of the test, and we documented the 95th percentile of WTR erasures for each grade and subject of the CRCT. Students whose 2009 CRCT showed a higher count of WTR erasures on the answer sheet are flagged as suspect. All suspect students' observations were dropped from all years during the cheating period. This approach appropriately deals with test scores known to be problematic, but likely still includes many suspect scores because it fails to drop observations that may still be problematic (for example, some students' scores may have been manipulated in another year but not 2009 and are therefore not flagged). Another potential drawback to this approach is that if low-achieving students' tests were more likely to be manipulated (as Sass et al., 2015 show), the remaining students in the sample are those who were relatively strong and may therefore inflate the performance of suspect teachers.

years provide erasure data for all students in the district. We primarily use the 2009 erasure data to flag suspect student scores and classrooms.

²² Cheating may still have occurred in some classrooms on the 2010 CRCT administration, as some classrooms had unusually high levels of WTR erasures (though compared against the 2009 administration in the same schools, the proportions of students in flagged classrooms dropped by 65% or more across subjects). However, because the 2010 erasure data is incomplete (only including students from 17 schools in 2010), we do not apply any special treatment to test scores earned in 2010 or later.

The third approach filters out all students who were in a classroom associated with a suspect administrator in the 2009 CRCT administration. All students' CRCT exams have a proctor recorded, which is not necessarily the students' teacher. Again applying the methodology of Sass, et al. (2015), we estimate the average number of WTR erasures among all students sharing the same proctor, which generally correspond to instructional classrooms. Where this classroom average exceeds the 95th percentile of the expected number of WTR erasures expected based on a clean administration, all students in the classroom are flagged as suspect and all of their observations during the cheating period are dropped. This approach differs from the second approach above in that the filter is applied to the classroom's proctor, not the individual student. Hence, it's possible some suspect students in the second approach above may not be considered suspect if most of their classmates' scores are not suspect; however, a whole classroom of students is dropped under this approach even if a particular student's exam did not show a high level of WTR erasures.

The final approach filters out all students ever linked to a cheating teacher. As described above, the GBI conducted a thorough investigation into the APS cheating scandal. In schools where there was sufficient evidence of cheating, GBI named specific staff who confessed to cheating during the investigation. Those teachers identified as cheating based on this report were flagged, and all students linked to these staff in any year during the cheating period were considered suspect. In theory, this should be the preferred method of filtering out suspicious scores because it identifies the staff member as suspect (not the student or classroom). Yet, many of the named cheaters in the GBI report cannot be successfully linked with an administrative teacher ID and of those that are linked, very few of them actually appear in the course files that link students with teachers (i.e., they are not typical classroom teachers, but more commonly school administrators). In the end, roughly 30 percent of named cheaters in the GBI report are successfully identified in the student-teacher links and have all of their linked students dropped from this analysis.

CRCT Results With Filtered Samples

For each of these four approaches, suspect observations are removed during the cheating years in APS, and the school fixed effects regression model described above is re-estimated. In Table 8, we present the results of the models applied to samples that filtered out suspicious observations from the analysis sample, as described in the Methods section above. Panel A reports the baseline estimates from the overall sample (identical to the school fixed effects results in Table 6), and Panels B through E correspond to the four methods for filtering out suspicious data. Note that only the preferred school fixed effects estimates are presented in this table for brevity.

The point estimates change relatively little in each of these filtered samples, although they tend to generally bump slightly upwards. Comparing Panel A (the baseline estimates) with Panel B (the most extreme filtering approach) shows slightly larger TFA effects in all subjects, however, none of the revised estimates change the statistical significance of the estimate, nor are the differences between the point estimates large enough to be statistically significant. This is true of all revised estimates in the remaining panels. Hence, widespread cheating does not appear to qualitatively change the overall results presented in Table 6.

Note that even though the EOCTs were not the focus of the test-score manipulation scandal in APS, our EOCT results could be influenced by the scandal due to the inclusion of suspect CRCTs as pretest scores. In Table 9, we apply the same four methods described above to filter out students with suspect pretest scores. Here, the statistical significance of the TFA estimate in American Literature is sensitive to the method of filtering out observations, although the magnitude of the estimate is not. Comparing across the point estimates for American Literature in the five panels the point estimate exceeds 0.11 standard deviations in all cases, though the standard errors grow large enough such that the estimate is no longer statistically significant (notably in Panels B and E). Also, we highlight that of all

EOCT subjects, the American Literature sample contains the fewest number of unique TFA teacher-year observations (from Table 3); hence, it is not surprising that this estimate and the standard error show some volatility here. In summary, filtering out suspect observations does not appear to qualitatively change the magnitude of the TFA impact in American Literature, although it does affect whether the finding is considered statistically significant. Estimates on the remaining EOCT subjects do not qualitatively differ due to the exclusion of suspect pretest scores.

Discussion

Here we discuss some of the findings from both the CRCT and EOCT to contextualize these in the broader research literature. First, the positive TFA impact on the science CRCT is consistent with those presented in Xu, et al. (2011); however, no significant differences were found on the Biology EOCT with a school fixed effects model (although this result could be due to low power, as the point estimate is 0.09 standard deviations). Next, TFA's impact on social studies achievement has not been previously explored in the literature, but the estimate on the CRCT is notable and surprisingly large.

No statistically significant effects were detected in math, reading, or language arts on the CRCT. The null findings in reading and language arts are consistent with most prior TFA impact studies (with the exception of Hansen et al., 2015). On the other hand, the null estimates in math are consistent with only two prior rigorous studies (Boyd et al., 2006; Clark et al., 2015) where at least five studies report statistically significant gains associated with TFA teachers in math (Clark et al., 2013; Glazerman et al., 2006; Hansen et al., 2015; Kane et al., 2008; Xu et al., 2011). Of the prior studies, perhaps the samples in Clark et al. (2015) and Hansen et al. (2015) are the most contextually similar to the sample here, as both studies analyze TFA impacts in more recent years, a time in which TFA has greatly expanded the size of its corps nationally (and in the samples of both studies), just as the corps has expanded in the Atlanta region. Though Hansen et al. (2015) find significant TFA effects in both reading and math in Miami

schools, Clark et al. (2015) use a randomized control trial and do not find any evidence of differential performance between TFA and non-TFA teachers. If we use these studies as the primary point of comparison for TFA estimates, then the null effects in math are not as surprising as they may otherwise be when taken in the context of all prior studies.

Next, the EOCT results show statistically significant point estimates in Math 2 and Biology under an OLS specification, though these effects do not hold up when we employ a school fixed effects specification. These findings are surprising in that they conflict with prior empirical evidence. This surprising finding, along with the somewhat surprising null effects in math, prompts us to explore potential factors that may help explain what may be going on in the Atlanta region.

We explored two issues that may possibly contribute to these surprising results, though for brevity do not present tables of estimates but discuss the qualitative findings.²³ First, to address the null effects in math, we estimate the TFA teachers' effect among active corps members only for the CRCT and EOCT samples. If there is some type of adverse selection into alumni status, such that the top-performing corps members are not retained in these districts, then high numbers of alumni could attenuate the TFA impact estimates. This could be particularly relevant in Atlanta, which has a significant number of TFA alumni and also showed the lowest point estimates in the CRCT samples (see Figure 1). The estimates resulting from this exploration were not significantly different from those presented in Table 6, across all subjects of both the CRCTs and EOCTs including math. Hence, the null effects reported in the tables do not appear to be driven by non-random selection into alumni status.

The second factor we explore, to address the reversal of significance in the EOCT results, is whether the schools in which TFA teachers are teaching are atypically *more* productive than other

²³ Tables of estimates are available from the authors by request.

schools, instead of *less* productive as is the case in most prior TFA studies.²⁴ We estimated a series of OLS models that flag school-year-grades in which TFA are placed (instead of directly flagging TFA teachers) and test whether these estimates are significantly greater than zero. In the CRCT samples, the point estimates were positive in all subjects (with the exception of reading) and statistically significant in science. In other words, yes, the CRCT data supported the idea that TFA teachers were in school-grades that appear slightly more productive; though the reversal of significance was not a major issue in the CRCT samples. In the EOCT samples, on the other hand, the point estimates on placement school-year-grades was significant in Math 1 and large (though not significant) in US History, but close to zero in other subjects. However, the estimates do not appear to explain the reversal of significance on the EOCT school fixed effects models in Math 2 and Biology. So, though there is a slight tendency for TFA teachers to be in slightly higher-performing school-grade cells, this does not appear to account for the reversals of statistical significance in Table 7. This would suggest some other dynamics are at play in the distribution of students or teachers in these districts that affect the within-school estimates of TFA effectiveness, though further exploration into this issue is beyond the scope of this study.

Conclusion

TFA has placed several thousand teachers in public schools in the Atlanta metropolitan area over the past 15 years, and this is the first report to estimate impacts of TFA teachers specifically in this region. Data from three school districts in the metropolitan area were obtained and analyzed for TFA impacts on the state's standardized tests, the CRCTs and EOCTs. The results reported here indicate TFA

²⁴ A related hypothesis is the placement of TFA into charter schools in these districts: one could speculate the reversal may be due to charter schools out-performing traditional schools if TFA are placed disproportionately into charters. This hypothesis, however, cannot be supported with the data because too few TFA teachers are observed in charters (particularly in the EOCT samples) to make a meaningful influence on the estimates. Of the three districts, APS has the most charter schools; in addition, students in APS charters perform slightly better on the CRCTs (based on OLS regressions). Yet, fewer than 15% of TFA observations in the CRCT samples are from charter schools, and less than 5% of TFA observations in the EOCT samples are from charter schools.

teachers had a positive impact on students' CRCT scores in science and social studies, and positive impacts on the American Literature EOCT. The estimated TFA teachers' effects are similar in magnitude to the performance differential associated with teachers having two or more years of experience in the data (compared to novice teachers). Supplemental investigations into the sensitivity of these findings to the inclusion of data tainted by the APS cheating scandal showed no qualitative change in the point estimates of these results, though the statistical significance of the EOCT estimates in American Literature is not entirely consistent across these specifications.

We find no evidence of significantly lower performance among TFA teachers on any of the tests we analyzed. Summarizing across all subjects, hiring TFA teachers in these school districts have been associated with net increases in student learning.

References

- Blinder, A. (2015, April 1). Atlanta educators convicted of school cheating scandal. *New York Times*. Retrieved from <http://www.nytimes.com/2015/04/02/us/verdict-reached-in-atlanta-school-testing-trial.html>
- 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.
- Clark, M., Chiang, H., McConnell, S., Silva, T., Sonnenfeld, K., Erbe, A., . . . Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the Teaching Fellows Programs* (No. 7890). Princeton, NJ: Mathematica Policy Research.
- Clark, M. A., Isenberg, E., Liu, A. Y., Makowsky, L., & Zukiewicz, M. (2015). *Impacts of the Teach For America investing in innovation scale-up*. Princeton, NJ: Mathematica Policy Research.
- Darling-Hammond, Linda, Deborah J. Holtzman, Su Jin Gatlin, and Julian Vasquez Heilig. “Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach For America, and Teacher Effectiveness.” *Education Policy Analysis Archive* 13 (42). 2005.
- Georgia Bureau of Investigation (2011). Special Report Probing Allegations of Test Tampering and Related Matters in the Atlanta Public School System (APS). Published in three volumes: http://www.11alive.com/data/pdf/CRCTInvestigationReport_Volume_1of3.pdf; http://www.11alive.com/data/pdf/CRCTInvestigationReport_Volume_2of3.pdf; http://www.11alive.com/data/pdf/CRCTInvestigationReport_Volume_3of3.pdf, accessed July 26, 2013.
- 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.
- 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 Working Paper #113. Washington, DC: American Institutes for Research.
- Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives* 2(3), 172–177. (2008).
- Hock, H., & Isenberg, E. (2012). *Methods for accounting for co-teaching in value-added models*. Washington, DC: Mathematica Policy 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.
- Miner, Barbara. “Looking past the spin: Teach For America.” *Rethinking Schools* 24(3). Retrieved from http://www.rethinkingschools.org/archive/24_03/24_03_TFA.shtml. 2010, Spring.
- Sass, T. R., Jarod Apperson, J., & Bueno, C. (2015). *The long-run effects of teacher cheating on student outcomes: A report for the Atlanta Public Schools*. Atlanta, GA: Atlanta Public Schools.
- 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.

Tables

Table 1. Count of Teach For America (TFA) Corps Members, by District and Year

Year	District						Total
	Atlanta	Clayton	Dekalb	Fulton	Gwinnett	All Others	
2008-09	165	-	-	72	-	10	247
2009-10	153	-	-	47	-	11	211
2010-11	145	18	-	3	-	26	192
2011-12	141	52	72	20	64	41	390
2012-13	130	66	64	54	62	53	429
2013-14	102	72	10	82	1	25	292

Source: Data on TFA placement in the Atlanta region.

Table 2. Criterion-Referenced Competency Test Data Inventory and Teach For America (TFA) Teacher Counts

Year	Atlanta		Clayton	Gwinnett
	All data	Since 2011		
2005-06	✓	-	-	-
2006-07	✓	-	-	-
2007-08	✓	-	-	-
2008-09	✓	-	-	-
2009-10	✓	-	✓	-
2010-11	✓	✓	✓	-
2011-12	✓	✓	✓	✓
2012-13	✓	✓	✓	✓
2013-14	✓	✓	-	-
TFA Teacher-Years				
Mathematics	339	180	8	33
Reading	431	208	15	23
ELA	430	207	15	23
Science	318	140	14	25
Social Studies	275	136	4	28

Table 3. End-of-Class Test (EOCT) Data Inventory and Teach For America (TFA) Teacher Counts

Year	Atlanta	Clayton	Gwinnett
2009-10	✓	✓	-
2010-11	✓	✓	-
2011-12	✓	✓	✓
2012-13	✓	✓	✓
2013-14	✓	-	-
TFA Teacher-Years			
9 th -Grade Literature	32	3	1
American Literature	29	3	0
Biology	41	12	9
Mathematics 1	27	16	4
Mathematics 2	34	20	3
U.S. History	42	1	0
<p>Note: Each district includes all EOCT subject tests for each year corresponding to a check mark, with the exception of Mathematics 1, which was not tested in 2013-14.</p>			

Table 4. Descriptive Statistics, Criterion-Referenced Competency Test (CRCT) Samples, by Teacher Type

	Non-TFA	TFA
Student-Level Variables		
Mathematics achievement	0.019	-0.268
Reading achievement	-0.008	-0.253
Language Arts achievement	0.067	-0.144
Science achievement	0.021	-0.269
Social Studies achievement	0.081	-0.189
FRL eligible	64.21%	80.56%
Black	60.39%	79.40%
Hispanic	17.51%	13.23%
English language learners	3.12%	1.98%
Special Education	8.00%	8.28%
Total student-year observations	301,786	27,562
Teacher-Level Variables		
Years of experience	7.144	1.938
Black	48.06%	46.00%
Hispanic	2.33%	2.90%
Total teacher-year observations	19,118	715
Total unique schools	270	107
Notes: FRL, free or reduced-price lunch. TFA, Teach For America teacher. Mean test scores are based on samples of students assigned to teachers in a given CRCT subject. Demographic information is an average for all students and teachers in the CRCT sample, weighted by student-teacher dosages.		

Table 5. Descriptive Statistics, End-of-Class Test (EOCT) Samples, by Type of Teacher

	Non-TFA	TFA
Student-Level Variables		
Prior Mathematics achievement	0.019	-0.297
Prior Reading achievement	-0.033	-0.351
Prior Language Arts achievement	0.032	-0.239
Prior Science achievement	-0.050	-0.403
Prior Social Studies achievement	-0.011	-0.373
FRL eligible	57.61%	82.49%
Black	56.36%	81.35%
Hispanic	17.93%	13.88%
English language learners	2.11%	1.92%
Special Education	7.73%	8.58%
Total student-year observations	121,216	11,086
Teacher-Level Variables		
Years of experience	6.817	0.902
Black	40.32%	28.52%
Hispanic	2.95%	0.28%
Total teacher-year observations	4,814	221
Total unique schools	99	34
Notes: FRL, free or reduced-price lunch; TFA, Teach For America teacher. Mean test scores are based on samples of students assigned to teachers in a given EOCT-tested subject. Demographic information is an average for all students and teachers in the Criterion-Referenced Competency Test sample, weighted by student-teacher dosages.		

Table 6. Estimates of Effects of Teach For America (TFA) Teachers for Criterion-Referenced Competency Tests (CRCTs), by Subject

	Mathematics		Reading		Language Arts		Science		Social Studies	
	1	2	3	4	5	6	7	8	9	10
TFA	0.005 (0.023)	0.003 (0.022)	-0.033 [†] (0.016)	-0.022 (0.016)	-0.005 (0.016)	0.001 (0.016)	0.085 [‡] (0.022)	0.095 [‡] (0.021)	0.118 [‡] (0.034)	0.115 [‡] (0.028)
Observations	362,571	362,571	528,388	528,388	528,038	528,038	360,841	360,841	326,219	326,219
R-squared	0.677	0.682	0.635	0.639	0.662	0.666	0.71	0.716	0.71	0.721
OLS	✓		✓		✓		✓		✓	
School fixed effects		✓		✓		✓		✓		✓

Note: *, †, ‡ represent p value < 0.10, < 0.05, < 0.01, respectively. OLS, ordinary least squares. Control variables include a squared expansion of pretest scores in the prior-year CRCT scores for all five subjects for each student. The vector of student characteristics includes variables for student race/ethnicity, gender, eligibility for free or reduced-price lunch (FRL) status, switching schools within the past year, special education and disability status, and English language learner status. The vector of classroom characteristics includes the percentages of the FRL-eligible, black, and Hispanic students. Teacher experience is controlled as a series of indicator variables (representing 1 year, 2 years, 3-4 years, 5-9 years, and 10+ years of experience; novice teachers are the omitted category). All control variables are interacted with grade. Year-grade and district-year fixed effects are also included.

Table 7. Estimates of Effects of Teach For America (TFA) Teachers for End-of-Class Tests, by Subject

	Mathematics 1		Mathematics 2		9th-Grade Literature		American Literature		Biology		U.S. History	
	1	2	3	4	5	6	7	8	9	10	11	12
TFA	0.037 (0.03)	0.000 (0.03)	0.103† (0.05)	0.049 (0.05)	-0.01 (0.03)	-0.017 (0.04)	0.091† (0.04)	0.130† (0.06)	0.112† (0.05)	0.090 (0.06)	0.178 (0.12)	0.047 (0.09)
Observations	50,232	50,232	67,487	67,487	84,517	84,517	45,219	45,219	83,143	83,143	76,457	76,457
R-squared	0.672	0.689	0.552	0.578	0.743	0.746	0.48	0.496	0.732	0.744	0.311	0.354
OLS	✓		✓		✓		✓		✓		✓	
School fixed effects		✓		✓		✓		✓		✓		✓

Note: *, †, ‡ represent p -value < 0.10, < 0.05, < 0.01, respectively. OLS, ordinary least squares. Control variables include a squared expansion of pretest scores in prior CRCT scores for all five subjects for each student, and a variable on the number of years since the Criterion-Referenced Competency Test (CRCT) was assessed. The vector of student characteristics includes variables for student race/ethnicity, gender, eligibility for free or reduced-price lunch (FRL) status, switching schools within the past year, special education and disability status, and English language learner status. The vector of classroom characteristics includes the percentages of the FRL-eligible, Black, and Hispanic students. Teacher experience is controlled as a series of indicator variables (representing 1 year, 2 years, 3-4 years, 5-9 years, and 10+ years of experience; novice teachers are the omitted category). Year fixed effects are also included.

Table 8. Estimates of Effects of Teach For America (TFA) Teachers for Criterion-Referenced Competency Tests (CRCTs), by Subject, When Filtering Suspect Student Observations

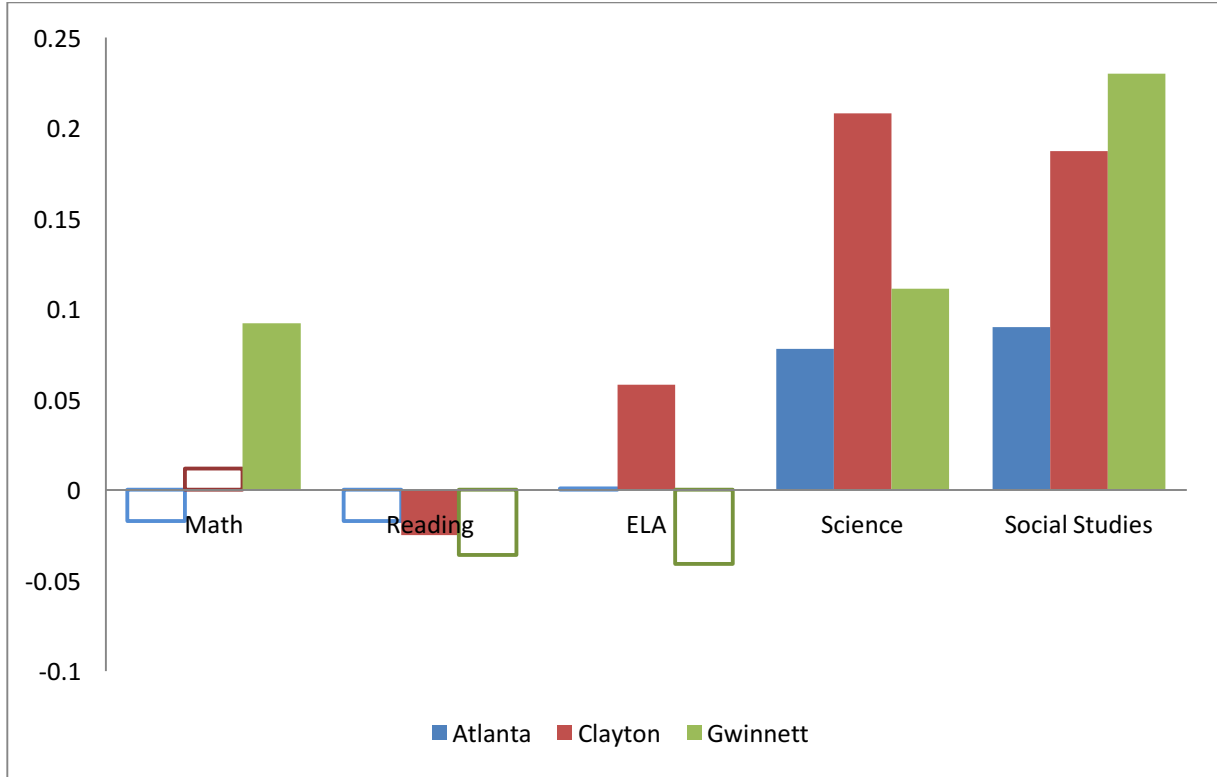
Subject	Mathematics	Reading	Language Arts	Science	Social Studies
Panel A. Baseline TFA Estimates					
TFA	0.003 (0.022)	-0.022 (0.016)	0.001 (0.016)	0.095‡ (0.021)	0.115‡ (0.028)
Observations	362,571	528,388	528,038	360,841	326,219
R-squared	0.682	0.639	0.666	0.716	0.721
Panel B. Dropped all observations that existed in Atlanta in cheating years					
TFA	0.004 (0.02)	0.001 (0.01)	0.012 (0.01)	0.102‡ (0.02)	0.128‡ (0.03)
Observations	275,260	395,334	394,942	271,317	266,116
R-squared	0.716	0.662	0.689	0.743	0.751
Panel C. Dropped all observations with wrong-to-right (WTR) erasure above 95th percentile					
TFA	0.007 (0.02)	-0.013 (0.01)	0.002 (0.02)	0.092‡ (0.02)	0.122‡ (0.03)
Observations	350,361	508,780	508,385	348,536	317,262
R-squared	0.698	0.651	0.677	0.728	0.731
Panel D. Dropped all students in a given classroom led by a suspect proctor					
TFA	0.003 (0.02)	-0.024 (0.02)	-0.003 (0.02)	0.099‡ (0.02)	0.100‡ (0.03)
Observations	358,025	522,189	521,817	356,042	322,498
R-squared	0.687	0.642	0.669	0.719	0.724
Panel E. Dropped all student observations associated with a cheating teacher in the 2009-10 year					
TFA	0.009 (0.02)	-0.026* (0.02)	0.000 (0.01)	0.125‡ (0.02)	0.127‡ (0.03)
Observations	437,990	613,797	613,219	434,697	395,236
R-squared	0.688	0.641	0.67	0.715	0.725
<p>Note: *, †, ‡ represent p value < 0.10, < 0.05, < 0.01, respectively. Control variables include a squared expansion of pretest scores in the prior-year CRCT scores for all five subjects for each student. The vector of student characteristics includes variables for student race/ethnicity, gender, eligibility for free or reduced-price lunch (FRL) status, switching schools within the past year, special education and disability status, and English language learner status. The vector of classroom characteristics includes the percentages of the FRL-eligible, black, and Hispanic students. Teacher experience is controlled as a series of indicator variables (representing 1 year, 2 years, 3-4 years, 5-9 years, and 10+ years of experience; novice teachers are the omitted category). All control variables are interacted with grade. Year-grade and district-year fixed effects are also included.</p>					

Table 9. Estimates of Effects of Teach For America (TFA) Teachers on End-of-Class Tests (EOTCs), by Subject, When Filtering Suspect Student Observations

Subject	Mathematics 1	Mathematics 2	9th-Grade English	American Literature	Biology	U.S. History
Panel A. Baseline TFA estimates						
TFA	0.000 (0.03)	0.049 (0.05)	-0.017 (0.04)	0.130† (0.06)	0.09 (0.06)	0.047 (0.09)
Observations	50,232	67,487	84,517	45,219	83,143	76,457
R-squared	0.689	0.578	0.746	0.496	0.744	0.354
Panel B. Dropped all observations that existed in Atlanta in cheating years						
TFA	-0.013 (0.03)	-0.002 (0.04)	-0.06 (0.05)	0.222 (0.18)	0.037 (0.06)	-0.031 (0.14)
Observations	44,473	59,834	78,282	37,975	76,745	68,103
R-squared	0.689	0.555	0.741	0.438	0.748	0.312
Panel C. Dropped all observations with wrong-to-right (WTR) erasure above 95th percentile						
TFA	0.001 (0.031)	0.054 (0.048)	-0.023 (0.039)	0.128† (0.058)	0.079 (0.061)	0.056 (0.091)
Observations	48,877	66,172	81,354	44,310	80,262	75,443
R-squared	0.691	0.576	0.744	0.492	0.746	0.35
Panel D. Dropped all students in a given classroom led by a suspect proctor						
TFA	0.009 (0.031)	0.057 (0.047)	-0.022 (0.037)	0.126* (0.063)	0.082 (0.061)	0.066 (0.087)
Observations	48,541	65,899	80,740	44,178	79,743	75,295
R-squared	0.691	0.575	0.742	0.489	0.746	0.348
Panel E. Dropped all student observations associated with a cheating teacher in the 2009-10 year						
TFA	0.006 (0.031)	0.061 (0.049)	-0.027 (0.035)	0.116 (0.075)	0.091 (0.062)	0.052 (0.089)
Observations	48,564	65,532	81,497	43,640	80,295	74,678
R-squared	0.691	0.574	0.744	0.487	0.745	0.347
<p>Note: *, †, ‡ represent p value < 0.10, < 0.05, < 0.01, respectively. Control variables include a squared expansion of pretest scores in prior CRCT scores for all five subjects for each student, and a variable on the number of years since the Criterion-Referenced Competency Test (CRCT) was assessed. The vector of student characteristics includes variables for student race/ethnicity, gender, eligibility for free or reduced-price lunch status, switching schools within the past year, special education and disability status, and English language learner status. The vector of classroom characteristics includes the percentages of the FRL-eligible, Black, and Hispanic students. Teacher experience is controlled as a series of indicator variables (representing 1 year, 2 years, 3-4 years, 5-9 years, and 10+ years of experience; novice teachers are the omitted category). Year fixed effects are also included.</p>						

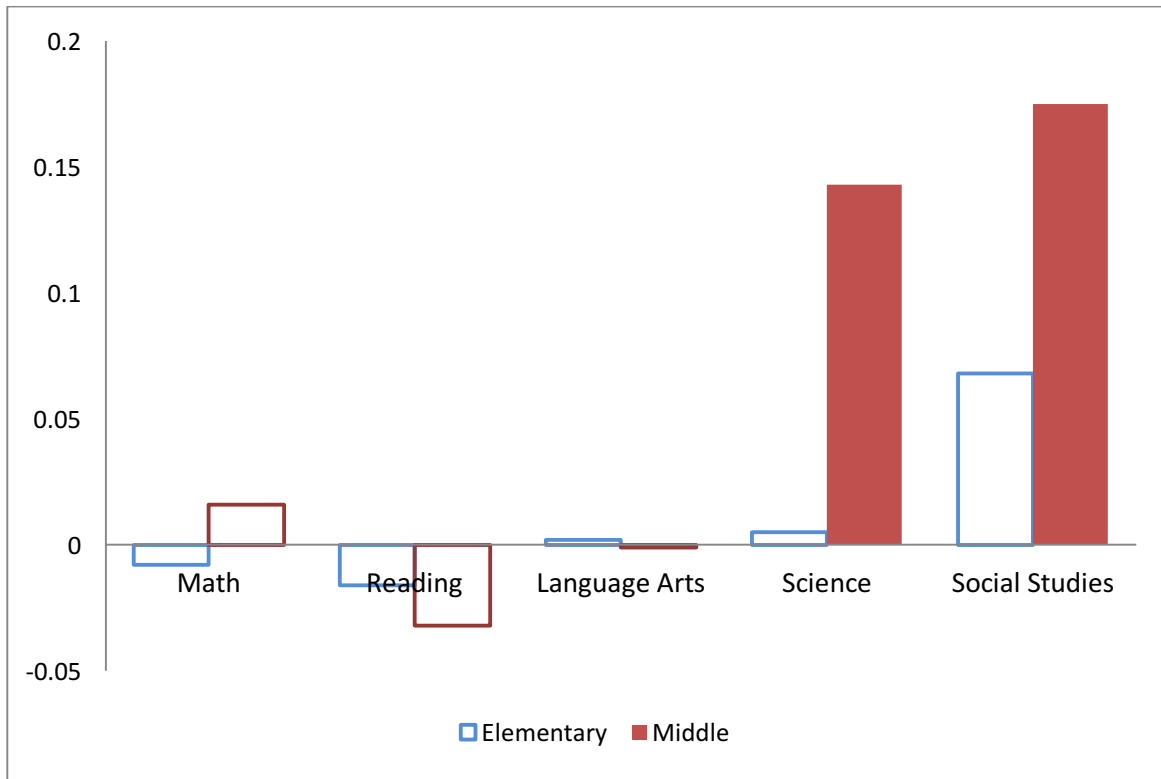
Figures

Figure 1. Estimates of Teach For America (TFA) Teachers on Criterion-Referenced Competency Tests by Subject and District



Notes: ELA, English language arts. Bar heights represent estimates of TFA teachers' effects in units of student standard deviations of achievement. Solid bars represent statistically significant estimates at p values < 0.05 .

Figure 2. Estimates of Effects of Teach For America (TFA) Teachers on Criterion-Referenced Competency Tests, by Subject and School Level



Note: Bar heights represent estimates of TFA teachers' effects in units of student standard deviations of achievement. Solid bars represent statistically significant estimates at p values < 0.05 .