CESifo
AREA CONFERENCES 2019

Economics of Education
Munich, 30–31 August 2019

Getting Effective Educators in Hard-to-Staff Schools

Andrew Morgan, Minh Nguyen, Eric A. Hanushek, Ben Ost, and Steven G. Rivkin

Getting Effective Educators in Hard-to-Staff Schools

Andrew Morgan*, Minh Nguyen*, Eric Hanushek***, Ben Ost*, and Steven Rivkin**

August 2019

PRELIMINARY AND INCOMPLETE
PLEASE DO NOT CITE

Abstract

Financial incentives to attract and retain educators in difficult to staff schools have typically not differentiated by educator performance, and this substantially weakens efforts to elevate the quality of instruction and raise achievement. The Dallas Independent School District developed a program that provides educators substantial additional compensation to teachers and principals willing to work in very low-achievement schools. Crucially, the amount of the additional pay depends upon a teacher or principal's effectiveness in previous years as measured by comprehensive systems of evaluation. Moreover, the district committed to staffing these schools with effective educators. Difference-in-differences estimates reveal dramatic achievement increases in reading and especially mathematics that substantially narrow the gap between students in these schools and the district average, highlighting the potential for targeted compensation programs linked with educator effectiveness to reduce the gap in school quality and achievement. Changes in the composition of teachers accounts for a portion of the improvement, though more effective leadership, stronger performance incentives, data-driven instruction and enhanced professional development may have also contributed to the increase.

*University of Illinois at Chicago; ***Stanford University, University of Texas at Dallas, and NBER, ** University of Illinois at Chicago, University of Texas at Dallas, and NBER

1. Introduction

Attracting and retaining effective educators constitutes a primary challenge to raising the quality of instruction, achievement and social mobility in schools that are chronically low achieving. Evidence suggests that additional pay can increase retention, but supplemental pay for working in disadvantaged schools does not typically depend upon effectiveness and therefore is highly unlikely to have a major impact on the quality of instruction.[1] Clotfelter et al. (2011) argues that these types of pay premia are unlikely to equalize teacher quality across advantaged and disadvantaged schools because they differentially attract teachers with worse credentials. Conversely, a recent randomized controlled trial provides evidence that a program that paid effective educators $10,000 per year for two years succeeded in attracting small numbers of high value-added teachers to designated schools and modestly raising achievement (Glazerman et al, 2013). Whether financial inducements to qualified educators could be implemented at scale and transform low-performing schools is a fundamental question for those seeking to raise the quality of instruction in disadvantaged urban and rural schools. The identification of effective teachers and school leaders requires a system that accurately measures educator quality, a challenging problem given the difficulties of measuring educator effectiveness.

In an effort to raise the quality of teaching and school leadership the Dallas Independent School District (DISD) introduced comprehensive evaluation and compensation reform that rated

[1] Research that investigates the effects of programs designed to attract educators to hard-to-staff schools includes Clotfelter, et al. (2008); Clotfelter, Ladd and Vigdor (2011); Steele et al. (2010); Cowan and Goldhaber, (2015); Springer et al. (2010); Springer, Swain and Rodriguez (2016); and Glazerman et al (2013)

educators on the basis of supervisor observations, student achievement, and student or family surveys and established compensation systems based on these evaluations. Concerns about persistent low performance and the possibility that the reforms could exacerbate staffing difficulties in low-performing schools, DISD implemented a compensatory pay policy, Accelerating Campus Excellence (ACE), that offered bonuses and substantial salary increases to educators willing to work in seven of the persistently lowest-achievement schools in the district. Importantly, DISD structured the program such that the pay premia depend upon the teacher's evaluation during the previous year. They also made a commitment to transform these schools immediately; over 60 percent of the teachers and all seven principals in ACE schools were replaced prior to the 2015-16 academic year.

The ratings produced by the comprehensive evaluation reform provided the necessary information on educator quality, and a sizeable pool of educators applied to work in an ACE school. Note among ACE teachers, roughly one third were in their first three years and thus not eligible for a proficient rating, another 28 percent were rated proficient, and less than 40 percent were in the higher, distinguished category.

Our empirical analysis uses a difference-in-differences design based on the fact that the central administration in Dallas identified a set of 25 schools that would potentially receive the intervention, but only 7 schools were initially selected into ACE; the remaining were referred to as ISN, (we refer to them as "near-ACE"), and provide a comparison group. ACE grew by six additional schools three years later in the second wave of the program, three coming from the "near-ACE" group and the remainder from other Dallas ISD schools. The first wave of ACE includes four primary schools and three middle schools, while the second wave of ACE includes five primary schools and one middle school. [2]

We show that at baseline, schools in the first wave of ACE adoption are systematically more disadvantaged than near-ACE schools, but the two groups are on quite similar achievement trends prior to policy implementation. These trends diverge sharply following ACE adoption, and ACE schools catch up with and overtake near-ACE schools within one year. Schools in the second wave of ACE are on a strong downward trend in the years leading up to ACE adoption, reflective of district efforts to target struggling schools. This trend reverses immediately following ACE adoption in 2017-2018. Because of the absence of parallel trends in Wave 2, we focus on Wave 1.

By embedding salary inducements to teach in low-performing schools into comprehensive systems of evaluation linked with compensation, DISD was remarkably effective at their goal of transforming the district's worst performing schools. In a single year, the test-score gap between the targeted, low-performing schools, and the average DISD school closed by more than 50 percent, and these gains continue in subsequent years. This dramatic success is consistent with the notion that a strong personnel system combined with the commitment of resources necessary to attract effective educators can dramatically improve school quality. Importantly, average mathematics achievement in Dallas ISD increased substantially during this period, suggesting positive direct benefits of educator evaluation and compensation reform in addition to their use in the compensatory pay program.

Though the central analysis estimates the overall effect of ACE, we also explore potential mechanisms through which the reform affects achievement. ACE not only overhauls the

[2] For the purposes of forming a control group, we define "near-ACE" as ISN schools that are never selected for ACE.

composition of teachers and replaces the principal, but it also amplifies the incentives for teachers and principals to elevate their performance and achievement. A more effective principal would be expected to enhance teacher development, and the evidence suggests that teachers benefit from working with highly effective peers.[3] The structure of the teacher pay premia also strengthened incentives because the supplement amount depends on the evaluation score. Each ACE educator receives a $2,000 annual bonus, and teacher salary supplements increase from $6,000 for those in their first three years or not rated proficient, $8,000 for those rated proficient, and $10,000 for high-scoring teachers who successfully complete the distinguished teacher process. Therefore, the supplements increase the salary differentials under the teacher compensation reform, elevating the financial incentives to raise the quality of instruction.[4] The pay-for-performance literature has mixed findings with some studies finding positive effects and many studies finding null effects.[5]

Although ACE focuses on educator quality and roughly 85 percent of the program cost goes toward pay supplements, it also incorporates other components that could affect the quality of instruction. These include the extension of the school day by one hour, additional professional development including an average increase of roughly one extra coach per ACE school, the requirement to adopt data driven instruction and a tool for monitoring assessment, and mandated provision of after-school programs until 6 pm that included dinner and transportation home. This complicates efforts to identify the effects of the personnel components, and we therefore focus

---

[3] Jackson and Bruegmann (2009)

[4] Principals and educators in other positions received fixed bonuses that did not vary by prior rating.

[5] See (Fryer, 2013; Goldhaber and Walch, 2012; Balch and Springer, 2015; Figlio and Kenny, 2007; Fryer, Levitt, List and Sadoff, 2012; Dee and Wyckoff, 2015; Brehm, Imberman and Lovenheim, 2015; and Glazerman and Seifullah, 2012). A recent recent meta-analysis (Pham, Nguyen and Springer 2018) finds that on average, pay-for-performance improves outcomes slightly.

on changes in the composition of teachers based on value added and the performance measures included in the DISD reforms.

The dramatic and immediate changes in the composition of educators has elements of a school turnaround intervention. The literature on school turnarounds reports mixed results, with several studies finding null or negative effects (Heissel and Ladd, 2016); Dougherty and Weiner, 2017) and other studies finding large gains from turnarounds (Dee, 2012; Strunk et al., 2016; Papay, 2015; Schueler et al (2016); Zimmer, Henry, & Kho, 2016). Schueler et al (2016) investigated a state-led turnaround in Lawrence, Massachusetts that generated mathematics and reading achievement increases that approach those produced by ACE. This intervention involved less teacher turnover and an intensive tutoring program for struggling students during school vacations that appears to be a primary driver of the improvement.

The next section describes the administrative and program data, and Section 3 discusses the Dallas ISD evaluation and pay reforms. Section 4 presents the differences-in-differences empirical model and estimates, highlighting potential threats to identification and relevant evidence. Section 5 explores the contributions of potential channels to the achievement increase, and Section 6 uses synthetic control methods to compare achievement changes in Dallas following educator reform with comparison districts. The final section summarizes the findings and considers implications for personnel policies.

2. Data

The data for the analyses come from several sources. Data on student and staff characteristics come from Dallas ISD administrative data as submitted to the Texas Education Agency. We use student standardized math and reading test scores from the State of Texas

Assessments of Academic Readiness (STAAR), as measures of student achievement, where the tests are standardized within Dallas ISD separately by subject and year. Currently we are assembling data with other tests used by the district, and this will permit us to include students and teachers in earlier grades in the estimation of overall ACE effects and teacher value added. Other student information includes courses, grades and schools attended, race, gender, indicator for students qualifying for programs such as free or reduced lunch, gifted, special education, and limited English proficiency. Staff information also contains demographics and courses, grades and schools taught.

We also have access to unique data that include scores and sub-metrics for all the components used in the evaluation rating process, such as teacher performance as measured by rubric-based observations, and student or family perception as measured by surveys from students and parents. We construct a panel that links teachers, students, and schools together from the 2011-2012 to 2017-2018 school years.

Table 1 shows descriptive statistics for ACE schools, near ACE schools, and other Dallas ISD schools before the policy. ACE and Near ACE schools are lower performing and have a substantially higher percentage of African American students and much lower percentages of Hispanic and LEP students than other Dallas ISD schools. There are also gaps between ACE and near-ACE schools in achievement and the black enrollment share; the achievement deficits for ACE schools approach 0.15 standard deviations in both math and reading. ACE teachers also have substantially lower math and reading value-added, and average teacher experience is twice as high in near-ACE schools as in ACE schools.

3. Institutional Background

Dallas is a large urban school district in north Texas comprised of roughly 160,000 students and 230 schools. Several years prior to the adoption of ACE, DISD undertook a process that fundamentally alters the evaluation and compensation of district educators. Prior to describing the ACE reform, we outline some of the main features of the principal and teacher evaluation and compensation reforms.

2.a. TEI and PEI

The district introduced the Principal Excellence Initiative (PEI) during the 2012-2013 academic year and the Teacher Excellence Initiative (TEI) during the 2014-2015 academic year. Though they differ in many details, the two reforms share a similar structure. Each contains an achievement component based on standardized assessments, a performance component based largely on supervisor observations and judgements, and a survey component based on feedback from students or families. There are target distributions for ratings categories and the components of TEI and PEI to limit evaluation inflation and retain control over the personnel budget. The current-year composite evaluation score determines the evaluation rating category, and the two-year average score determines the salary bin (referred to as effectiveness rating) with some qualifications. PEI and TEI delineate in great deal the requirements of the initiatives, points awarded for each criteria, and educator responsibilities for carrying them out. We now highlight some main features of each and relevant implementation details.

The PEI evaluation component is determined by both overall achievement and success at reducing the achievement gap. The district developed numerous assessments to measure achievement in subjects and grades lacking a state-standardized test. Initially three separate achievement scores were calculated, and the number of points assigned was the highest from three alternatives: Status (percentage of tests with scores at a specified standard); a value-added measure; and achievement score relative to the scores of a designated peer group of schools based on prior achievement. Subsequently, the status alternative was dropped. The number of achievement points depends on success at reducing achievement gaps by race and ethnicity. This codifies the objective of equity and support for students in demographic groups that have lower average achievement in the district and state.

PEI places substantial weight on effectiveness as an instructional leader. Almost 20 percent of the performance component focuses directly on improving teacher effectiveness and congruence between teacher performance and student achievement. Thus, the principal is rated on their work in support of teachers and the alignment between the subjective teacher evaluation and teacher effectiveness at raising achievement. The congruence component of the evaluation is designed to mitigate the tendency to inflate more subjective evaluations and to deter arbitrary judgements of teachers based on factors other than the quality of teaching. Unlike the case for TEI, attendance and enrollment also contribute to the performance score for principals.

TEI has a similar structure as PEI, but naturally there are important differences between teacher and principal evaluation systems. Supervisor classroom observations constitute the primary source of evidence for the performance score. TEI specifies ten, 10- to 15-minute spot observations of each teacher and one 45-minute extended observation per year by the designated supervisor, typically the principal or assistant principal. The supervisor is required to provide written feedback following all observations and conference with the teacher following the extended observation.

Student Perception is measured by surveys conducted in the second week of April. Most students in grades 3-12 complete two surveys, one online and one in paper. Results from the surveys will be summarized by a statistic for teachers with sufficient number of responses. Points are assigned based on the target distribution at grade-level to assure equity because early grade-level students tend to provide more positive responses.

The achievement score is based on the results for a teacher's students (when available) and the outcomes for the entire school. This is intended to foster collaboration and a common mission, but it likely also handicaps teachers who work in schools with a high fraction of ineffective educators. This may exacerbate difficulties of attracting and retaining teachers in low-performing schools, the problem ACE was designed to remedy.

Differences in grade, subject, and role lead to cases in which a teacher may not have a measure of achievement for her own students or student survey results. TEI divides teachers into four categories and assigns different weights to the performance, student perception and achievement components depending upon the availability and type of assessment and survey data collected. We focus on teachers whose students take a state standardized mathematics or reading test. Though teacher pay is determined largely by the evaluation score, there are other consideration including, teachers must apply to and pass a distinguished teacher review process in order to place into the highest evaluation and pay categories; experience and education determine the salary for teachers new to Dallas ISD; teachers with fewer than three years of experience are limited to the maximum rating and compensation they can receive; and district teachers who taught in Dallas ISD prior to the TEI reform cannot have their nominal pay lowered below their pre-reform level.

2.b. Accelerating Campus Excellence

In academic year 2015-2016, one year after TEI adoption, Dallas ISD implemented the Accelerating Campus Excellence (ACE) program to raise the quality of instruction and achievement in Dallas ISD's chronically low-performing schools. This intervention incorporates several components including enhanced professional development, tools and commitment to data driven instruction and ongoing assessment, an extra hour in the school day, and after school enrichment programs, but the cornerstone of ACE is the dedication of substantial resources to attract and retain highly effective teachers and leadership teams. Educators who apply and are selected to work at ACE campuses receive signing bonuses of $2,000 and stipends that depend upon position, and, in the case of teachers, on TEI effectiveness ratings for the prior year. Stipend amounts equal $13,000 for a principal, $11,500 for an assistant principal, $8,000 for a counselor, $6,000 for an instructional coach, and between $6,000 and $10,000 for teachers. Note that classroom teachers and specialists were eligible for the ACE payments. The ACE program had a total budget of $4,720,200 for the 2015-16 academic year which came out of general operating funds from Dallas ISD. The signing bonuses and stipends constituted roughly 85 percent of the budget, with the remainder divided among professional development ($350,000), transportation ($246,000), and uniforms ($125,000) for schools that decided to require them.[6]

Based on the target distribution of ratings, approximately 20% of Dallas ISD teachers qualify for the $10,000 pay premium by having passed distinguished teacher review, 40% of teachers qualify for an $8,000 pay premium by obtaining a proficient rating, and 37% qualify for a $6,000 premium by receiving a progressive rating due either to being inexperienced or to failing to reach proficiency. In the first year of the ACE program, 40 percent of ACE teachers qualified for a $10,000 stipend, 28 percent for an $8,000 stipend and 32 percent for a $6,000 stipend. In

[6] Information on costs and programs comes from EA16-601-2, "Updated Evaluation of Accelerating Campus Excellence (ACE) 2015-16," produced by the Dallas ISD Department of Evaluation and Assessment.

addition to raising the level of compensation, the structure of these stipends amplifies the TEI pay for performance incentive for teachers in an ACE school by increasing the differential between ratings categories. For example, at a non-ACE school, moving from the level just below distinguished up to the first rung in the distinguished category raises salary by $5,000.[7] At an ACE school, the same rating change raises salary by $7,000.

Dallas identified a total of 25 low-performing elementary and middle schools that they considered for the ACE intervention, but ultimately designated seven as ACE schools in 2015-2016 based on persistent low-achievement; the remaining 18 schools were designated as ISN. Another six schools were selected for the second wave of ACE in 2017-2018. All potential ACE teachers (including those who were effective) were required to interview and/or were evaluated to stay at an ACE campus. Some teachers decided to leave even if offered the opportunity to stay, perhaps in response to the requirement to contribute three hours per week to the after-school program. To the extent possible, campuses were reconstituted with teachers who had earned high evaluation ratings.

Over 60 percent of teachers and all principals in schools newly designated as ACE were different from the teachers and principals who had been in the school the previous year. Such turnover would be expected to affect adversely the quality of instruction, as teachers adjust to different schools and in many cases different grades. Consequently, benefits of ACE would be expected to increase over time as the teaching force stabilized.

Figure 1 illustrates the dramatic increases in educator quality in Wave 1 ACE schools as measured by PEI and TEI ratings for the three years the schools were designated as ACE. The lower panel shows that the share of principals rated effective in ACE schools rose from 0 in the

[7] In the terminology used by the district, this is a movement from proficient 1 to proficient 2.

year prior to ACE to more than 70 percent in the first year of the program. This share also increased more modestly in near-ACE schools but fell by almost 10 percentage points in the remaining schools that constitute the vast majority in Dallas. This decline results largely from the entry of many educators who had not previously been principals into the principal position.

The upper panel of Figure 1 illustrates a similar pattern for teachers. The share of teachers in Wave 1 ACE schools that had at least a proficient rating in the 2014-15 academic year rose from 29 percent in 2014-15 to 72 percent in 2015-16; the share declined a small amount in near-ACE schools and only slightly in the rest of the district. In results not shown, we find that Wave 2 schools show a similar pattern, with an almost 50 percent increase in proficient share, far greater than the small increases in the near-ACE and other Dallas ISD schools. Clearly ACE substantially increased the shares of educators rated effective in ACE relative to near-ACE schools, and this likely understates the improvement in educator effectiveness since early career teachers selected to work in an ACE school were not eligible to earn a proficient rating.

4. Estimation of ACE effects

4.a. Empirical approach

Identification of the effects of ACE on mathematics and reading achievement requires a valid counterfactual, and a difference-in-differences specification that uses the 18 schools designated as near-ACE as the control group seems like a promising approach. This approach requires the satisfaction of the common trends assumption, and we describe primary threats to this assumption and relevant evidence on each. Following this discussion, we present the basic difference-in-differences results using the near-ACE schools as the comparison group and illustrate the annual achievement changes by school classification: ACE, near-ACE, and the remaining schools. Comparisons of trends prior to the policy period provide information on the validity of the common trends assumption, and achievement changes following policy enactment provides suggestive evidence on the importance of channels other than fixed changes in educator composition such as greater stability and teacher growth.

Our primary analysis is based on a simple difference-in-differences design that includes demographic controls in some specifications. This design is facilitated by the fact that DISD identified both ACE schools (that received the intervention) and "near-ACE" schools that were considered for, but did not receive the ACE designation in either 2016 or 2018. Our baseline estimating equation restricts the analysis to just ACE and near-ACE schools. Because there are relatively few treated schools, clustering at the school level has the potential to over or under state standard errors. For now, we cluster standard errors at the school-by-year level but will utilize permutation tests in the subsequent draft.

There are several potential threats to credibly identifying the effect of the ACE program on student and school outcomes. First, as with any difference-in-differences approach, it is possible that the near-ACE schools would have trended differently in the post-period even without program implementation. Based on discussion with administrators, there were no policies targeted towards near-ACE schools that did not also apply to ACE schools. Combined with the fact that the achievement for Wave 1 ACE schools trends similarly to that for near-ACE schools performance prior to the policy, we believe that the available evidence supports the common trends assumption for Wave 1. In contrast, schools assigned to Wave 2 trend far more negatively in the two years prior to their designation. This raises the possibility that a temporary negative shock precipitated assignment to Wave 2. However, given the extensive information Dallas ISD has on each school, we believe a more likely explanation is that classification resulted from indications of serious problems in these schools. Nonetheless, the divergent Wave 2 achievement trends raises questions about the interpretation of DID estimates for Wave 2, and so we focus the analysis on the Wave 1 schools where we have a more credible counterfactual.

A second threat to identification is the possibility that students (or their parents) respond to ACE designation by transferring their students into or out of the ACE schools. This could alter the composition of ACE schools and thereby improve school average test scores without raising the quality of instruction. These schools did experience large annual enrollment fluctuations, and we investigate the possibility of improvement in student composition by comparing changes in demographic characteristics around the time of program inception.

A third threat to identification is the possibility that the near-ACE schools are also affected by the policy. For example, ACE may adversely affect the quality of educators in near-ACE schools through the loss of teachers to ACE schools or greater difficulty attracting and retaining effective teachers and principals. This concern is mitigated by the fact that Wave 1 ACE schools represent only 7 out of 234 Dallas schools and thus spillover effects in the teacher market are likely to be relatively small. We show that in practice, fewer than 2% of near-ACE teachers moved to ACE schools. In fact, among all schools in the district, the modal number of teachers lost to Wave 1 ACE schools is one.

An issue separate from identification considerations but relevant for policy interpretation is the effect of ACE on the distribution of teacher quality. Although ACE schools are too small a share to substantially affect the aggregate teacher market, it remains the case that the policy caused many effective teachers and principals to move to ACE schools. If teachers who left ACE schools following the adoption of the policy are simply redistributed among other schools, it is possible that the policy has no aggregate effect on student achievement in Dallas.[8] If this is not the case, then the effect depends upon the quality difference between teachers who left an ACE school and exited the district and entrants into non-ACE schools. Importantly, the long-run equilibrium of a system that pays substantial stipends for educators in low-performing schools is likely to differ substantially from the short-term changes following policy adoption. We intend to explore these issues in future work.

4.b. Baseline results

Our primary goal is to assess how the ACE designation affects achievement. As a first step towards this goal, we plot math and reading scores in ACE Wave 1, ACE Wave 2, near-ACE and other Dallas schools from 2011-2012 to 2017-2018. Academic years are indicated based on the spring, so the first ACE wave is treated in 2016 and the second wave is treated in 2018. Figure 1 shows that both waves of ACE and the near ACE schools were on downward trajectories in the four years before the first wave of ACE (2012 to 2015). Schools that were selected for ACE during the first wave were lower performing than near-ACE schools, but the trends for the two groups are roughly parallel. When the first wave of ACE adoption occurs in 2016, average math achievement in ACE schools increases by approximately 0.3 standard deviations, surpassing near-ACE schools and closing the gap between ACE and the other Dallas ISD schools by over 50%. Near-ACE schools also make improvements in 2016, though these are far smaller than those made in the ACE schools.

Schools that were selected for ACE in 2018 (Wave 2) were similar to near-ACE schools from 2012 to 2015 but experienced sharp declines in math and reading achievement between 2015 and 2017. This is consistent with the district selecting schools that were trending

[8] Even without mean improvement, reallocation may be a desirable outcome as it reduces inequality with no apparent efficiency loss.

downward for the second wave of ACE treatment. Schools in the second wave of ACE increased math scores by approximately 0.4 standard deviations in 2018.

In addition to previewing our main results, Figure 1 highlights the magnitude of the baseline deficiency of ACE schools. In the year just before adoption, average achievement in both waves of ACE schools was approximately 0.5 standard deviations below that in the other Dallas ISD school category. By 2018, both waves of ACE schools were within 0.1 standard deviations of the other Dallas ISD schools, while near-ACE schools still lagged by approximately 0.3 standard deviations.

In Figure 2, we show the analogous results for reading scores. The pattern is similar to that of math scores, but the ACE improvement is less dramatic. As with math scores, Wave 1 ACE schools and near-ACE schools trended very similarly up until 2015. In 2016, both Wave 1 ACE and near-ACE schools improve, but the improvement is twice as large among ACE schools. ACE Wave 2 schools trend downward until 2017 and then improve substantially in 2018.

We view Figure 1 and 2 as indicating a large ACE effect, but it is worth noting that the Wave 2 ACE schools trend quite differently than near-ACE schools leading up to ACE adoption in 2018. If this divergence continued, Wave 2 ACE schools would have fallen even further behind near-ACE schools suggesting that DID estimates may be downward biased. That said, it is also possible that Wave 2 ACE schools would have recovered in the absence of the intervention. More generally, the divergence in pre-trends introduces uncertainty into the estimates of program effects on Wave 2 schools. Therefore, we focus our analysis of the channels through which ACE affected achievement on Wave 1.

Table 2 presents difference-in-difference estimates for the effects of ACE on achievement for Wave 1 and Wave 2 separately. All coefficients are highly significant based on clustering the standard errors by school-by-year, though we acknowledge that these standard errors may understate true uncertainty. Clustering at the school-level would also be inappropriate, however because of the small number of treated schools. In future drafts we plan to test for statistical significance using a permutation test. Estimated effects on math achievement are larger than those for reading in both periods, equaling 0.24 standard deviations in Wave 1 and 0.30 standard deviations in Wave 2. Reading coefficients from specifications with student demographic controls equal 0.11 in Wave 1 and 0.20 in Wave 2. The fact that the estimates are largely insensitive to the inclusion of the controls provides additional support for the difference-in-differences approach.

We now examine changes in student composition that could potentially contribute to achievement increases in ACE schools. Table 3 reports changes in demographic and program characteristics for Wave 1 ACE schools, near-ACE schools and remaining Dallas ISD schools following ACE implementation and reveals only very small changes in any of the characteristics. Share low income fell by 1 percent in ACE and near-ACE schools, and share black fell by 3 percent in near-ACE schools while remaining constant in ACE schools. Thus, there is little or no evidence that student composition drives or even contributes to the achievement increases in ACE schools.

Although we observe little change in demographic characteristics, it is possible that ACE students improve along unobservable dimensions. We take a standard approach to account for unobserved heterogeneity by including a cubic in lagged test score in the specifications. This substantially reduces the sample size by throwing out all students in third grade and those without lagged scores. In addition, it alters the comparison from differences in average achievement to differences in average annual achievement growth. The ACE effect on math achievement growth is roughly 33 percent smaller than that on achievement, equaling 0.16 with a standard error of (0.065). This is a large and highly significant effect of ACE on annual math achievement growth. The effect on reading achievement growth equals 0.07, which is slightly more than one third smaller than the effect on achievement. It has a standard error of 0.049 and is thus not significant at conventional levels. Nonetheless, the smaller effect on reading achievement growth lines up with the smaller effect on reading achievement level.

Table 4 provides information on the second threat to internal validity, the possibility that the labor market for teachers is sufficiently thin so that the incentive for high quality teachers to move to ACE schools and departure of many who previously taught in an ACE school adversely affected the quality of instruction in near-ACE schools. The transition matrix for the years surrounding the two ACE waves shows that only a small numbers of teachers transition between ACE and near-ACE schools, and this mitigate concerns about negative spillovers to near-ACE schools. For Wave 1, only 13 out of 633 near-ACE teachers in 2016 just arrived from an ACE school, and only 8 teachers transitioned from a near-ACE to an ACE school during this time. The numbers are somewhat higher for Wave 2 but still small: 13 out of 545 near-ACE teachers just arrived from an ACE school and 16 ACE teachers came from a near-ACE school. Moreover, ACE schools hired a total of 156 teachers in their first year of implementation, less than 2% of the total Dallas teacher labor market. We cannot rule out the possibility that ACE schools hired some high-quality teachers who otherwise would have ended up at near-ACE schools, but the negligible direct movement suggests that ACE is unlikely to substantially affect the stock of teachers at near-ACE schools.

5. Contributions of various channels

Although salary supplements and bonuses for educators at the seven ACE schools accounted for 85 percent of the program cost, the intervention involved other components including a longer school day, required after school programs, mandated and supported data-driven instruction and enhanced professional development. Lavy (2015) and Rivkin and Schiman (2015) find a positive effect of instruction time on achievement, and Fryer (2014) shows that a set of interventions including increased instruction time, data-driven instruction, and educator replacement substantially increased mathematics achievement. Although evidence on professional development and coaching is less convincing, it is possible that these components contributed to the achievement increase. Because we cannot identify the magnitude of their combined effect, we focus more narrowly on the effects of changes in teacher composition.

Specifically, we describe the efficacy of teachers who enter, exit and remain in ACE and near-ACE schools in 2016 and difference between ACE entrants and exits and changes over time in the effectiveness of teachers who remained in ACE schools following the introduction of the program. Estimates of teacher value added capture not only fixed teacher productivity but also influences of peer teachers, principal effects on school quality, and other factors potentially including the other components of ACE. Comparisons of new entrants and teachers who exited ACE schools prior to 2016 provide suggestive evidence of differences in effectiveness, but the less favorable working conditions for those who exited the ACE schools prior to the treatment suggests the difference overstates the fixed differences in productivity. The difference between those remaining in ACE schools and those exiting provides information on the productivity deficit of leavers, and the increase in value added for teachers who remain in ACE schools provides information on the contributions of the other factors to the achievement growth.

Teachers who remain in ACE schools likely have increased value-added following ACE adoption for a variety of reasons. These teachers experience dramatic changes in the composition of peer teachers and increases in their own experience, both of which raise the quality of instruction.[9] Furthermore, more effective school leaders also raise teacher value-added as do the amplified incentives under TEI. Because working conditions for these incumbent teachers unambiguously improve, we expect that the growth in their average value added net of the increase due to the acquisition of early experience and higher teacher peer quality provides an upper bound estimate on the contributions of other factors such as professional development, increased instruction time. Notice that we have not attempted to quantify the contribution of the principal, as it is quite difficult to separate from other factors.

We consider two metrics of teacher quality. First, in Tables 5 and 6 we present average performance points (based primarily on supervisor observations). Despite the detailed rubric, there remains a subjective element to these evaluations, and differences among principals and school circumstances almost certainly contribute to the score variation. Many of these teachers also receive points for a third evaluation component based on student perception, and teachers who enter ACE schools tend to be regarded more favorably than those who exit. However, because of differences among students and school environments and substantial evaluation inflation we do not focus on this metric of the evaluation system. Tables 5 and 6 use the same teacher quality metric, but Table 5 includes a broader set of teachers because in Table 6 we restrict the analysis to only teachers for whom we can calculate value-added.

The second metric of teacher quality is estimated value-added in math or reading. Although performance points based largely on supervisor observations reflects the degree to

[9] Jackson and Bruegmann (2009) identify the positive effects of higher value-added peers, and there is a consensus that early experience raises the quality of instruction.

which teaching practice meets the prescribed standards, value added to achievement and other outcomes measures the contribution of a teacher to learning. Following established practice, we estimate value added to mathematics and reading achievement for students in grades four to eight using teacher fixed effect models with cubic polynomials in lagged mathematics and reading achievement, demographic characteristics, and program variables as controls. We estimate the models separately by subject, grade and year. If a student reports having two teachers for a subject (either two math teachers or two reading/ELA teachers), we weight each teacher by 0.5 in producing averages by subject, year, sector and teacher transition status.

Table 5 shows the contributions of composition and improvement to the substantial increase in ACE teacher quality measured by standardized performance points. Teachers who left an ACE school after 2015 had an average score of -0.51 in 2015, while entrants with scores in 2015 had an average of 0.56. This difference exceeds one standard deviation and is consistent with a substantial contribution of composition to quality improvement, though the fact that the entrants and leavers had different supervisors introduces some uncertainty. Notice, however, that the large group of entrants in 2016 without a performance score from category A or B had an average score in 2016 that was quite similar to the average for ACE leavers who remained in the district after leaving an ACE school. Consequently, the average gain from changes in composition would appear to be closer to 0.85 standard deviations of performance points. An important caveat that we will discuss below is the fact that many of these new teachers had no prior teaching experience in the Texas public schools and were selected based on their district applications and interviews. It may well be that inexperience dampens the performance score even in cases of effective teaching. Finally, the 32 A & B teachers who remain in the district experience an average improvement of 0.14 standard deviations. All these teachers had different principals in 2016 than in 2015, and this introduces some uncertainty into the interpretation of the increase. By comparison, entrants to the Near-ACE schools in 2016 with performance points in 2015 lagged entrants to ACE schools by more than one standard deviation. ACE attracted many teachers from other Dallas ISD schools, while a much higher fraction of entrants to Near-ACE schools were not previously teaching in the district.

Table 6 reports average performance points for the subset of teachers with value-added scores, meaning that they taught mathematics or reading/ELA in grade 4 or higher. Currently the value-added sample based on the STAAR tests includes only a fraction of teachers for whom value added can be estimated once we integrate the other district assessments are into the estimation. The expanded sample is particularly important in looking at teachers who taught in ACE schools prior to the intervention and remained in those schools under the ACE program.

There are similarities but also differences between the patterns in Tables 5 and 6. On the one hand, entrants to ACE have much higher 2015 averages than exits from ACE or entrants into Near-ACE schools. On the other hand, the ACE stayers with VA scores in both years have a slightly lower average score in 2015 than the stayers with VA scores in both years in Near-ACE schools. The differences are not large, and the much more favorable compositional shift in ACE schools is the dominant change. A final similarity between the tables is the negative selection out of both sectors among teachers who do not have performance points calculated in 2016. Many of these teachers left Dallas ISD, consistent with the findings of negative selection of existing teachers in existing work on large Texas districts (Hanushek et al, 2016).

Table 7 reports average VA scores by sector, year and teacher transition status, and the estimates are consistent with the belief that both changes in composition and increased value added of incumbent teachers contribute to the achievement gain in ACE schools. Those leaving an ACE school have an average VA of -0.16 in 2015, while entrants with a VA score in 2015 have an average of 0.12 in 2015 and 0.19 in 2016. Entrants without a score in 2015 have an average of -0.04 in 2016, still far higher than the leavers. The differences between entrants and leavers suggests that changes in teacher composition made an important contribution to the gains in ACE schools, but the fact that the leavers had more difficult working conditions than the stayers muddies the comparison. In results not shown, we find that, the group of entrants who had taught in Dallas ISD in 2015 had higher value-added peer teachers and more highly rated principals in their prior schools than did the leavers from the ACE schools. And the change in principal, data driven instruction and coaching complicate comparisons between the average value added of all entrants to ACE schools in 2016 and all teachers who left an ACE school after 2015. Thus the difference between the average VA of entrants in 2016 following their arrival to an ACE school and leavers in 2015 prior to their departure from an ACE school provides an upper bound on the contribution of changes in composition to the ACE program effect.

Similarly, the value-added improvement of teachers who remain in an ACE school which are equal to 0.14 in mathematics and 0.18 in reading provide estimates of upper bounds on the contribution of factors other than fixed differences in teacher quality, and we make use of information on changes in peer-teacher quality and experience to estimate the contribution of those factors. This leaves the contributions of data-driven instruction, strengthened performance incentives, additional instructional time, enhanced professional development, an extended after school program, uniforms and a more effective principal as ACE components that could raise value added.[10] Existing evidence suggests that more intensive coaching and professional development are unlikely to increase value added substantially, but it is not possible to separate

[10] The absence of subject-specific instruction time information precludes the estimation of the effect of additional instructional time.

the contributions of these ACE components.[11] Therefore, our approach is to approximate the effects of higher peer-teacher quality and experience, remove these from the average improvements of teachers who remain in an ACE school, and consider remainder as an estimate of the contributions of factors other than fixed differences in teacher effectiveness. The low numbers of teachers responsible for mathematics and reading instruction certainly limit the value of this exercise, but we intend to make use of district assessments that permit the estimation of value added for additional teachers. At this point, the primary purpose of this exercise is to illustrate the approach to the estimation of the contribution of fixed differences in teacher quality to the ACE effects on mathematics and reading achievement.

We turn now to estimates of the contributions of experience and peer teacher quality to the growth in value added. Table 8 shows that two (40 percent of) math and reading teachers who remained in an ACE school had zero years of prior experience in 2015 and one reading teacher had only one prior year of experience. Estimates suggest that value added increases by roughly 0.08 standard deviations in math and slightly less in reading between the first and second year of teaching and by somewhat less between the second and third year. This suggests that the additional experience would tend to increase average VA by roughly 0.03 standard deviations in both subjects. Note that Table 8 also shows that a sizeable fraction of entrants into ACE schools had no prior teaching experience, consistent with achievement increases in the second and third years of ACE. The uncertainties in the estimation of the changes in peer-average value added lead us to use a conservative estimate of 0.2 standard deviations, and the estimates in Jackson and Bruegmann (2009) suggest that a change of this magnitude would be expected to increase value added by roughly 0.01 standard deviations.

[11] Footnote Mathematica studies on PD and coaching

Taken together, this suggests that other factors including principal quality account for no more than 0.1 standard deviations of the value-added growth out of the total of 0.24 standard deviations in math and 0.17 standard deviations in reading (top row of Table 7). An alternative and more direct approach is to compare the average 2015 value-added of leavers and entrants, imputing values for the entrants without value added estimates in 2015. Because roughly three quarters had no prior experience in 2016, it seems reasonable to lower their value-added scores only 0.02 standard deviations for experience, 0.01 standard deviations for peer quality, and an approximation of 0.06 standard deviations for all other changes. This produces a similar average improvement due to the replacement of leavers with entrants of approximately 0.16 standard deviations, but it relies upon strong assumptions regarding the comparability of value-added estimates in different school environments.

Although we do not attempt to quantify the contribution of principals to the increase in value added, Table 9 illustrates the difference in principal performance points and family survey responses in the seven ACE schools following program implementation and the arrival of a new leader. Total principal performance points increase by more than 2 standard deviations, and the shares of families that strongly agree that the learning is appropriate and there is a safe learning environment also increase substantially.

6. Dallas ISD Achievement Trends Following Reform

The ACE analysis compares ACE with Near-ACE and other Dallas ISD schools, but the performance of Dallas ISD schools relative to other Texas districts following the evaluation and compensation reform provides context within which to consider the findings. A finding that ACE schools simply tread water while the other Dallas ISD schools decline paints a different picture than a finding that ACE schools outperform other Dallas ISD schools that exhibit substantial improvement relative to other Texas districts.

Of course, all districts experience changes in policy, student composition, resources and personnel during this period, and therefore the counterfactual to achievement under PEI and TEI cannot be an estimate of how Dallas ISD would have fared in the absence of the policy. Rather we compare achievement trends in Dallas ISD following the reform with those of a comparison district defined by having quite similar outcomes in the pre-policy period. Because selection of any single comparison district or set of districts would be arbitrary, we use synthetic control methods to construct a counterfactual district based on mathematics achievement trends prior to the reform. (cite and describe the method) The transition from the more basic TAKS to the more challenging STAAR state-standardized test in the year prior to the reform adds an important source of variation in the pre-period, as Dallas ISD experienced a substantial decline in test performance during this transition.

A substantial change in demographic composition following the reform would potentially compromise the synthetic control approach, and Figure 3 illustrates trends in share black, share eligible for a subsidized lunch, share classified as limited English proficient (LEP), and share receiving special education. The figure shows small increase post-2012 in share LEP in Dallas ISD vis-à-vis the synthetic control and a slight decline in share black, but these would be expected to cause negligible changes in the achievement differential. Trends in the rates of special education classification and eligibility for a subsidized lunch are quite similar.

Figure 4 illustrates the mean achievement trends in the pre- and post- reform periods (left panel) and their differences (right panel), highlighting both the virtually identical trends for Dallas ISD and the synthetic control in the pre-period and substantial improvement in Dallas ISD following the implementation of TEI. The fairly flat line between 2013 and 2015 suggests that PEI on its own had little effect, while the decline between 2015 and 2016 is consistent with some turbulence in the first year of TEI. Between 2015 and 2018 average math achievement increased by roughly 0.2 standard deviations in Dallas ISD relative to the synthetic control, similar to the increase following the initial adoption of PEI.

We use permutation test to assess whether Dallas ISD achievement in the post-reform years significantly exceeds synthetic control achievement. Essentially a separate synthetic control is constructed for each district in the control donor pool, and we compare the treatment effects for Dallas ISD with the distribution of placebo treatment effects for the other districts. Figure 5 reproduces the average achievement differences between Dallas ISD and its synthetic control (solid black line) and the corresponding differences when each of the other districts is the treatment. The trend for in the estimated treatment for Dallas ISD is much more positive than that for the majority for placebo treatments, and by 2018 the difference for Dallas ISD is near the top of the distribution.

Nevertheless, at conventional significance levels we fail to reject the hypothesis of no difference between average achievement in Dallas ISD and the synthetic control in any year. Figure 6 shows the raw P values (two times the share of schools with estimated achievement differences higher than Dallas ISD in the right panel) and P Values adjusted for the precision of the synthetic control pre-period matching (left panel), and even in 2018 the P values do not fall below 0.2, where Dallas ISD falls at roughly the 90th percentile in terms of achievement difference with its synthetic control.

7. Summary and Policy Considerations

We document the remarkable success of a program implemented by Dallas ISD that embeds increased compensation for working in a low-performing school into a comprehensive evaluation and pay structure in which educator effectiveness serves as the primary determinant of the level of compensation. The evaluation systems generate information that can be used to identify effective educators and support the growth of all teachers, and the district salary structure establishes the practice of differentiated pay by the quality of instruction and leadership. ACE amplifies salary differences by level of effectiveness thereby strengthening performance incentives and encouraging effective educators to accept the risk of working in a previously low-performing school.

Several factors make our empirical findings credible. First, there is a well-defined control group that was considered for but did not ultimately receive the ACE designation. This group of schools was on a very similar trend to Wave 1 ACE schools in the years prior to the ACE intervention. Second, we see little change in student composition at ACE schools relative to near-ACE schools suggesting that altered student composition does not explain the test score improvement. Finally, there are two waves of ACE, and the timing of test score improvement exactly lines up with the timing of ACE implementation.

Because ACE is a multi-dimensional intervention there are substantial challenges to the identification of the effects of specific components of the program. Our approach is to focus on within and between teacher variation in mathematics and reading value added and estimate an upper bound on the effects of all factors other than fixed differences in teacher effectiveness. The initial sample size is quite small, and we will expand the sample by incorporating district assessments into the analysis. Nevertheless, the pattern of results is consistent with the notion that changes in the composition of teachers accounts for a substantial share of the ACE effect. Evaluations of principals suggest that the replacement of principals led to a pronounced improvement in school leadership, though we are not able to identify the aggregate leadership effect or channels through which higher principal quality raised achievement.

The results demonstrate the potential to elevate the quality of instruction in even the lowest-performing urban schools, and an important policy question concerns the optimal structure for such a targeted compensation program. This depends crucially on the dynamic forces that lead some schools to be very low achieving. One approach would be to target the highest poverty or most racially isolated schools, but evidence illuminates substantial variation in school effectiveness among schools with similar predicted achievement on the basis of student demographic characteristics. Figure 7 plots school-average achievement taken over two different two-year periods against predicted achievement based on student demographic characteristics for the top and bottom quartiles of the distribution based on predicted achievement. The figure shows that there is substantial variation around the regression line. In fact, predicted achievement explains roughly half the variation in actual achievement. Of course, this is a limited set of variables, and random shocks and test-measurement error reduce the R^2. Yet these are large samples of tested students. Figure 8 plots the difference between actual and predicted achievement in the later period against the difference in the early period, and there is substantial movement between periods, particularly in the lowest predicted achievement quartile.

Table 10 quantifies the changes over time in school performance by the joint distribution of initial quartile of predicted achievement and initial quartile of the difference between actual and predicted achievement for two different time periods. The top panel compares school performance in 2010-2011 and 2004-2005, while the bottom panel compares performance in

2006-2007 and 2004-2005. If the year-to-year differences result largely from random shocks the differences in the two panels should be similar. However, if more persistent factors lead to differences over time the differences in the top panel should exceed those in the bottom panel.

Entries in the top panel reveal substantial mean reversion that is particularly pronounced for schools in the lowest quartile of predicted achievement. On average mathematics achievement decreases by almost one quarter of a standard deviation between 2004-2005 and 2010-2011 for schools in the highest performing, lowest predicted achievement schools in 2004-2005. Note that the magnitude of the decline decreases as predicted achievement quartile increases, and the magnitudes in all four predicted achievement quartiles are more than twice as large as the declines between 2004-2005 and 2006-2007. Similar, though smaller changes emerge for the third achievement quartile in terms of performance in 2004-2005. Finally, there is little or no change for schools in the second performance quartile regardless of predicted achievement quartile and large increases for those in the bottom performance quartile that vary far less by predicted achievement quartile. Although sample sizes are somewhat smaller on average for schools in the first and fourth performance quartiles, the differences are not that large.

The extensive information on educator effectiveness produced by TEI and PEI can facilitate a more in-depth understanding of the dynamics of school quality fluctuations and the forces that contribute to concentrations of ineffective educators in some schools. This would provide information that could contribute to the design of compensation policies designed to foster equity and successful interventions in cases of deteriorating or persistently low school quality.

Bibliography

Balch, Ryan, and Matthew G. Springer. 2015. "Performance pay, test scores, and student learning objectives." *Economics of Education Review 44* (2015): 114-125.

Brehm, Margaret, Imberman, Scott A., and Lovenheim, Michael F., 2015. "Achievement Effects of Individual Performance Incentives in a Teacher Merit Pay Tournament." NBER *Working Paper No. 21598*, National Bureau of Economic Research, Inc.

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor. 2008. "Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina." *Journal of Public Economics 92*: 1352–1370.

Clotfelter, Charles, Helen G. Ladd, and Jacob L. Vigdor. 2011. "Teacher Mobility, School Segregation, and Pay-Based Policies to Level the Playing Field." E*ducation Finance and Policy 6* (3): 399-438.

Cowan, James and Dan Goldhaber. 2015. "National Board Certification and Teacher Effectiveness: Evidence from Washington." *CEDR Working Paper 2015-3*. University of Washington, Seattle, WA.

Dee, Thomas S. and James Wyckoff. 2015. "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." *Journal of Policy Analysis and Management*. https://doi.org/10.1002/pam.21818

Dee, Thomas. 2012. "School turnarounds: Evidence from the 2009 stimulus." *NBER Working Paper No. 17990*, National Bureau of Economic Research, Inc.

Dougherty, Shaun M. and Jennie M. Weiner. 2017. "The Rhode to Turnaround: The Impact of Waivers to No Child Left Behind on School Performance." *Educational Policy.* https://doi.org/10.1177/0895904817719520

Figlio, David N. and Lawrence W. Kenny. 2007. "Individual teacher incentives and student performance." *Journal of Public Economics 91*: 901–914.

Fryer Roland G. 2014. "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments." *Quarterly Journal of Economics*. 129 (3) :1355-1407.

Fryer, Roland G. 2013. "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." *Journal of Labor Economics, 31* (2): 373-427, Fryer, Roland G., Steven D. Levitt, John List and Sally Sadoff, 2012. "Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment," *NBER Working Papers 18237*, National Bureau of Economic Research, Inc.

Glazerman, Steven and Allison Seifullah. 2012. "An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after Four Years. Final Report." *Mathematica Policy Research Report*, Mathematica Policy Research.

Glazerman, Steven, Ali Protik, Bing-ru The, Julie Bruch, and Jeffrey Max. 2013. *Transfer Incentives for High-Performing Teachers: Final Results from a Multisite Experiment* (NCEE 2014-4003). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goldhaber, Dan and Joe Walch. 2013. *Rhetoric Versus Reality: Is the Academic Caliber of the Teacher Workforce Changing? CEDR Working Paper 2013-4*. University of Washington, Seattle, WA.

Hanushek, Eric A., Steven G. Rivkin, and Jeffrey C. Schiman. 2016. "Dynamic Effects of Teacher Turnover on the Quality of Instruction." Economics of Education Review (55): 132-148.

Heissel, Jennifer A. and Helen F. Ladd. 2016. "School Turnaround in North Carolina: A Regression Discontinuity Analysis." *CALDER Working Paper No. 156*.

Jackson, Kirabo C. and Elias Bruegmann. 2009. "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers." *American Economic Journal: Applied Economics 1(4): 85-108.*

Lavy, Victor. 2015. "Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries." *The Economic Journal*, 125 (588). F397-F42

Ost, Ben. 2014. How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal: Applied Economics 6* (2): 127-151.

Papay, John P. 2015. *The effects of school turnaround strategies in Massachusetts*.

Pham Lam D., Tuan D. Nguyen, Matthew G. Springer. 2017. "Teacher Merit Pay and Student Test Scores: A Meta-Analysis." *Unpublished Research*. Accessed at: https://s3.amazonaws.com/vu-my/wp-content/uploads/sites/868/2013/02/05145950/Pham-Nguyen-Springer-2017.pdf

Rivkin, Steven and Jeffrey Schiman*. 2015. "*Instruction Time, Classroom Quality, and Academic Achievement.*" The Economic Journal, 125 (588). F425-F448*

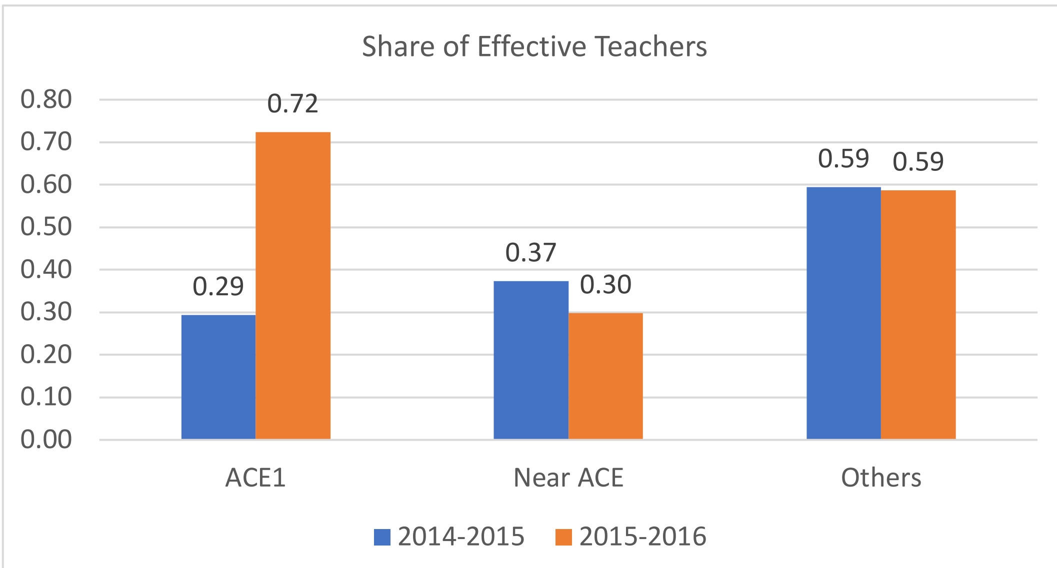
Ronfeld, Matthew, Susanna Loeb and James Wyckoff. 2013. How teacher turnover harms student achievement. *American Educational Research Journal 50* (1): 4-36.

Schueler, Beth E., Joshua S. Goodman, and David J. Deming. 2017. "Can States Take Over and Turn Around School Districts? Evidence From Lawrence, Massachusetts." *Educational Evaluation and Policy Analysis 39* (2): 311-332.

Springer Matthew G., Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher. 2010. *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives at Vanderbilt University.

Springer Matthew G., Walker A. Swain, and Luis A. Rodriguez. 2015. "Effective Teacher Retention Bonuses: Evidence From Tennessee." *Educational Evaluation and Policy Analysis 38* (2): 199-221.

Figure 1. Shares of principals and teachers rated effective in the first year of ACE and the previous year, by school type



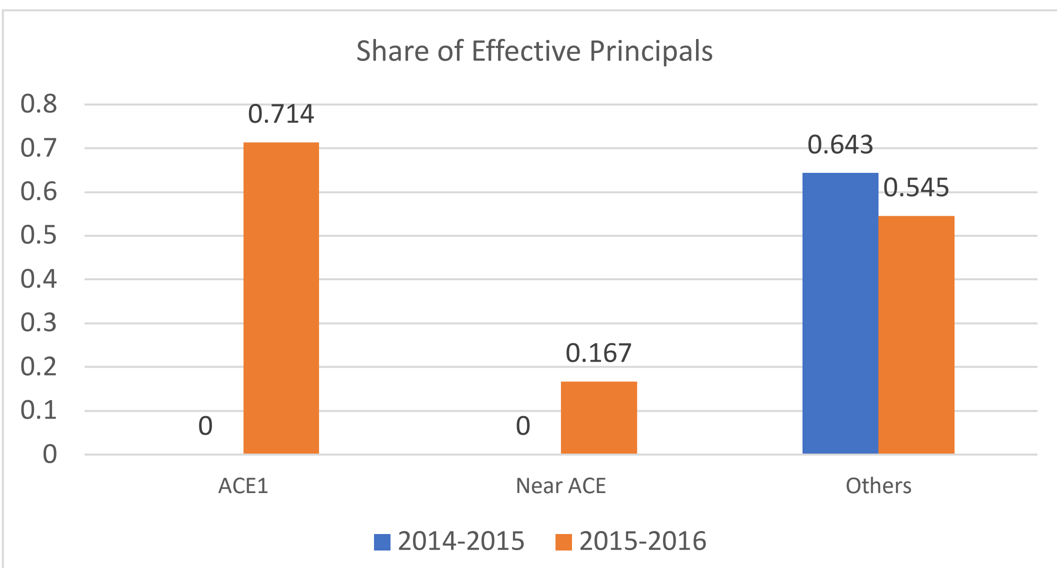


Figure 2. Trends in mathematics and reading achievement, by school category: 2012-2018

Note that Non-ACE refers to all schools other than ACE and near-ACE

Figure 3. Trends in District Share Limited English Proficient (LEP), Share in Special Education, Share Black and Share Low-Income in Dallas ISD and the Synthetic Control: 2004 to 2018

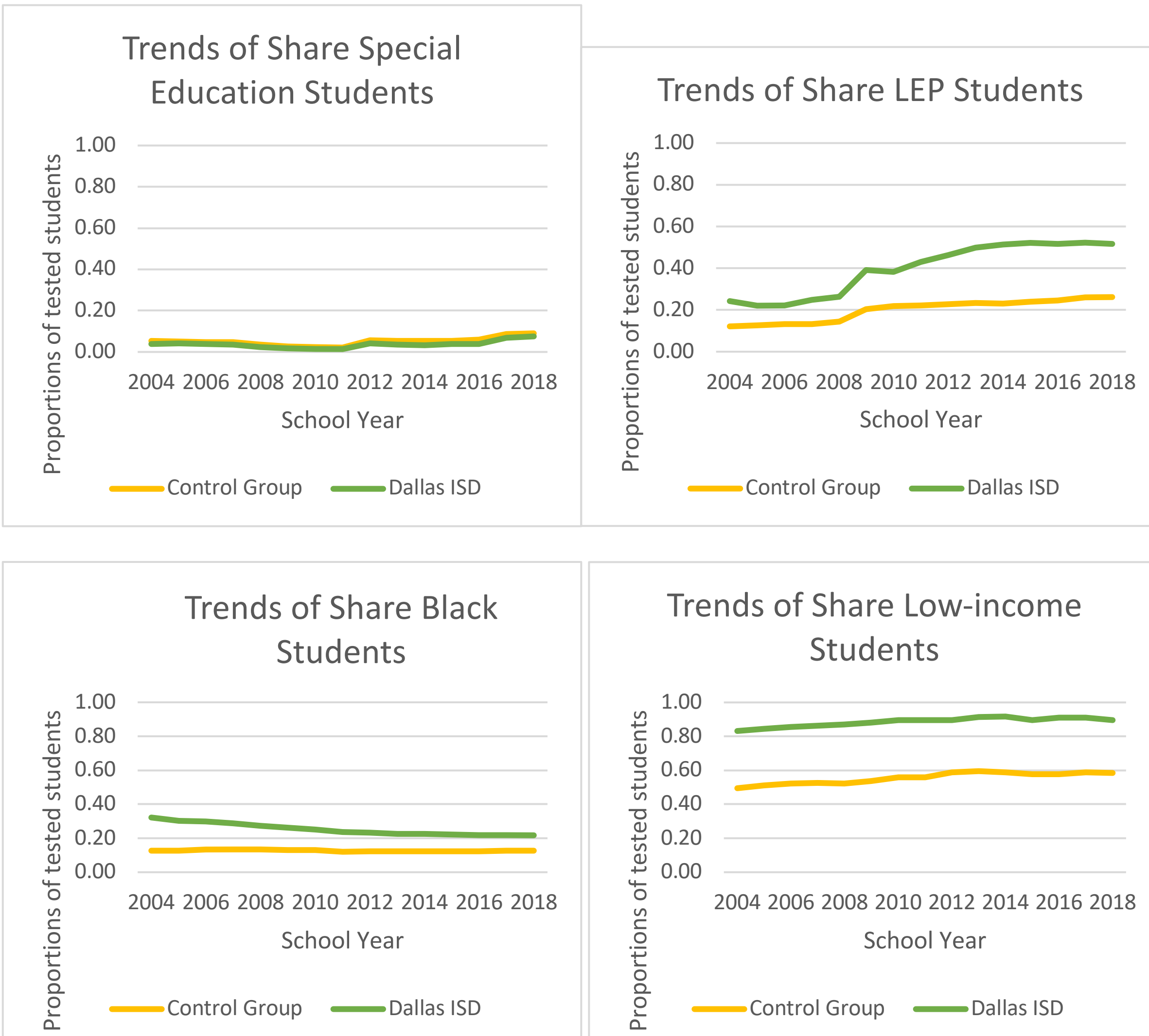
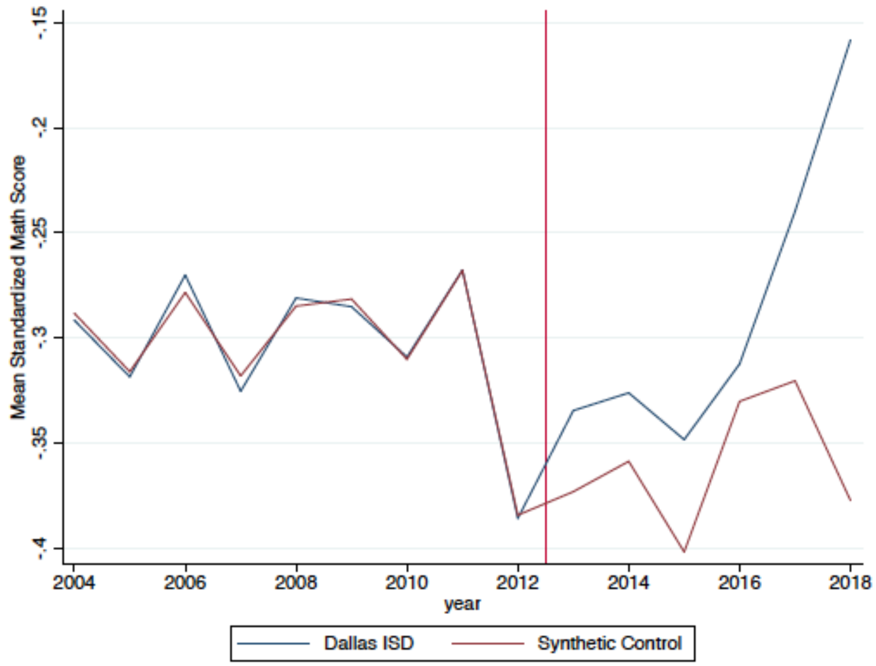


Figure 4. Average Mathematics Achievement in Dallas ISD and the Synthetic Control Before and Following the Adoption of the Principal Excellence Initiative (2013) and Teacher Excellence Initiative (2015): 2004 to 2018



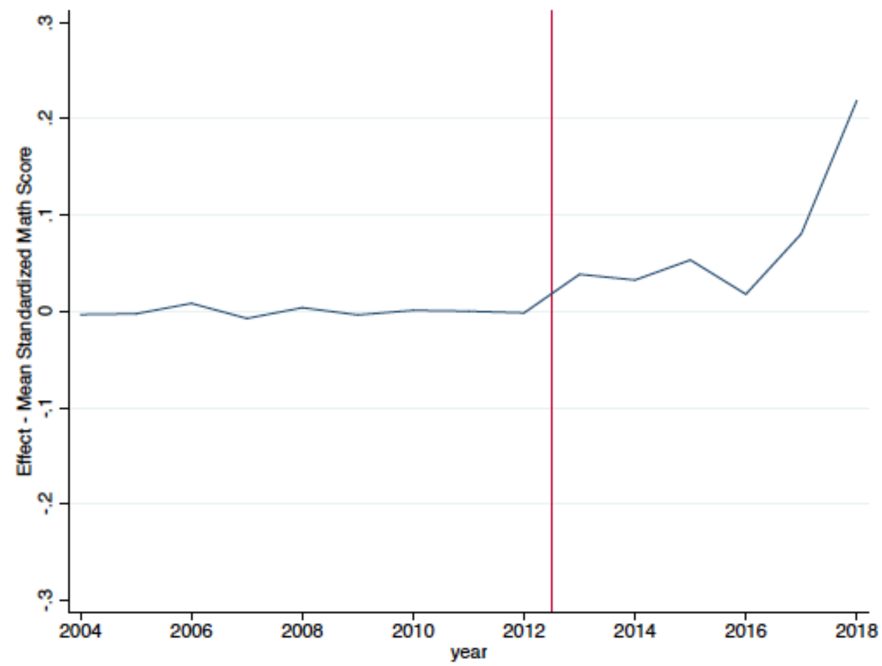


Figure 5. Synthetic Control Estimated Treatment Effects for Dallas ISD and Placebo Synthetic-Control Estimates for the Other Districts in the Synthetic Control Donor Pool

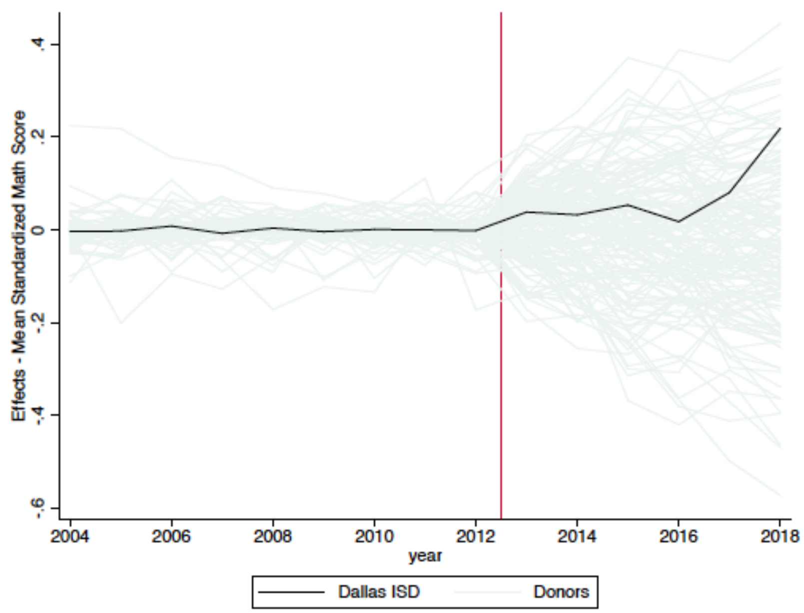
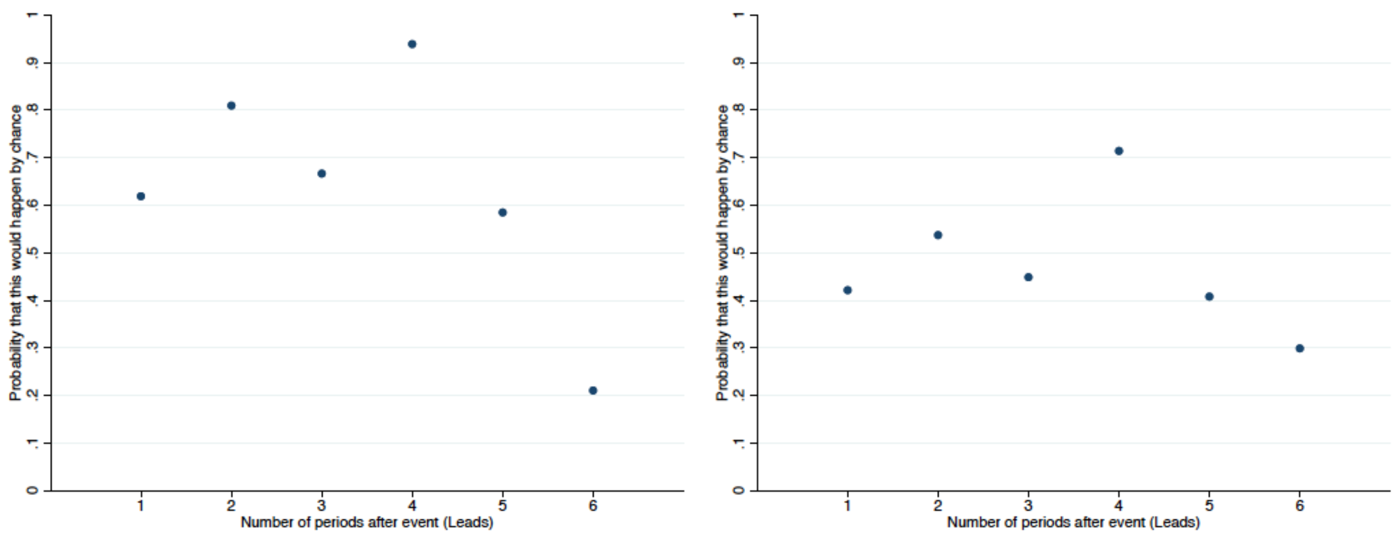


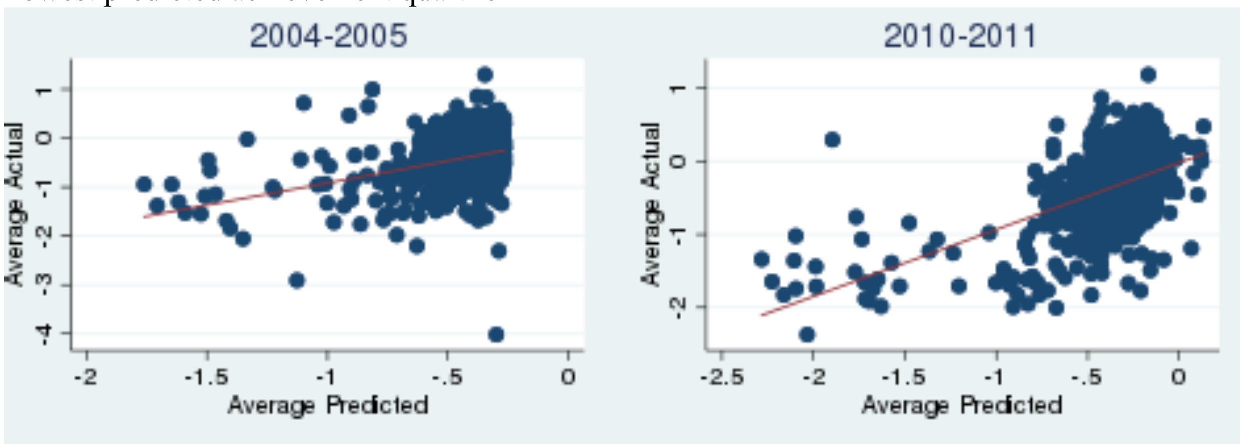
Figure 6. P Values and Adjusted P Values for the Hypothesis of no Difference Between Average Achievement in Dallas ISD and the Synthetic Control Following the Introduction of the Principal Excellence Initiative, by Years Since Policy Adoption



Note: P Values equal two times the fraction of placebo estimates in Figure 4 in a given period that lie above the estimate for Dallas ISD for that period. Adjusted P values equal these P values divided by ???

Figure 7. School-average achievement over a two year period plotted against predicted achievement based on school demographics for schools in the lowest and highest predicted achievement quartiles: 2004/05 and 2010/11

Lowest predicted achievement quartile



Highest predicted achievement quartile

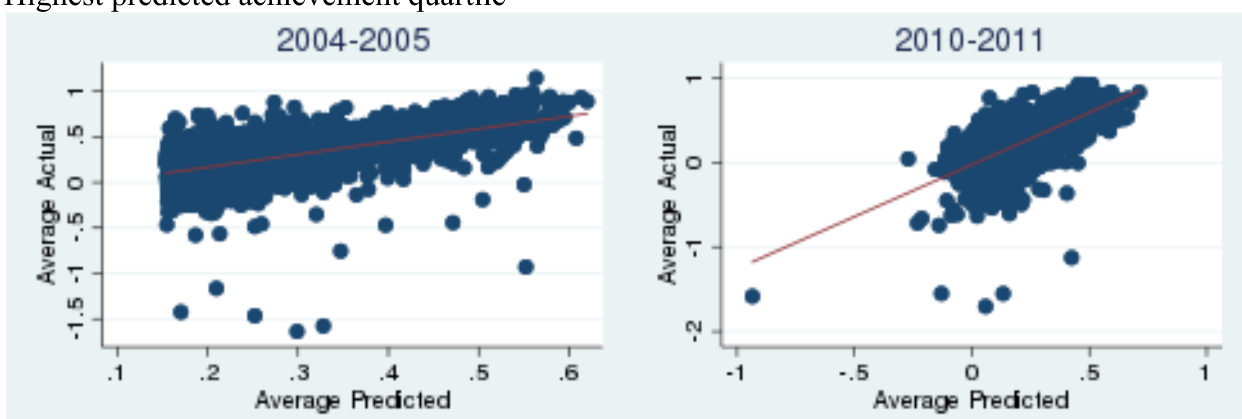
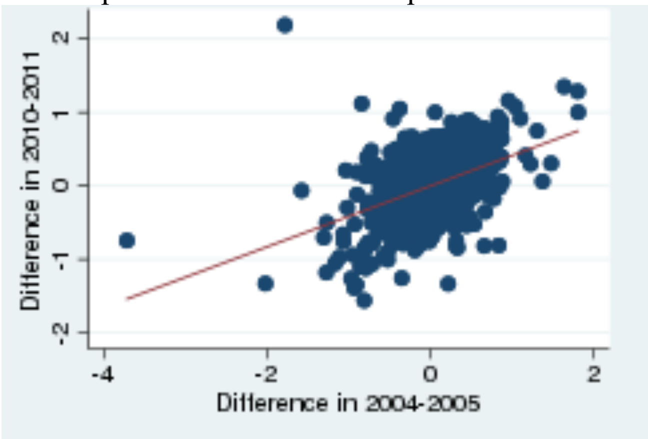


Figure 8. Relationship between the difference between actual and predicted achievement in the two periods, by poverty quartile

Lowest predicted achievement quartile in 2004/05



Highest predicted achievement quartile in 2004/05

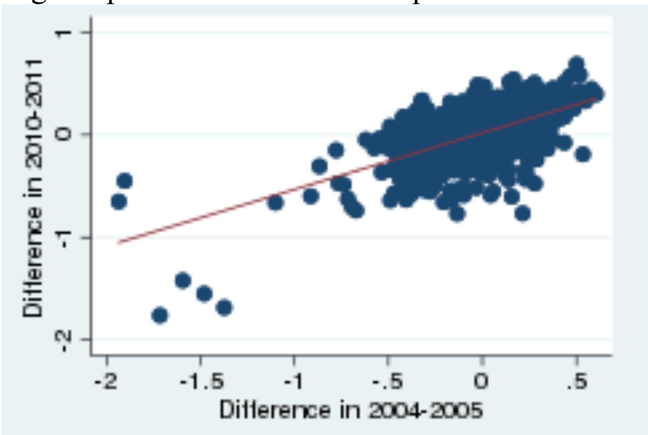


Table 1. Student demographic and program characteristics and teacher evaluations and characteristics in 2014-2015, by school type

school type	ACE Wave 1	ACE Wave 2	Near ACE	Other Schools
Student variables				
Math Score	-0.516	-0.359	-0.373	0.0722
Reading Score	-0.495	-0.364	-0.368	0.0329
Free or Red. Price Lunch	0.965	0.981	0.967	0.926
White	0.00684	0.0123	0.0106	0.0458
Afr. American	0.594	0.474	0.461	0.180
Hispanic	0.384	0.485	0.508	0.744
Native American	0.00190	0.00130	0.00252	0.00377
Asian	0.000760	0.00973	0.00162	0.0129
Special Educ.	0.125	0.0857	0.0955	0.0831
LEP	0.283	0.385	0.392	0.553
Male	0.543	0.517	0.523	0.516
Teacher variables				
student perception score	0.657	0.798	0.667	0.706
standardized observation score	-0.290	-0.233	-0.191	0.0775
reading value added	-0.155	0.0317	-0.0288	0.0340
math value added	-0.0888	-0.0329	0.00470	0.0783
years of experience	4.875	8.688	9.811	10.77

Table 2. Difference in Differences Estimates of ACE Effects on Math and Reading Achievement, by Wave of ACE

	ACE Wave 1				ACE Wave 2			
	Math		Reading		Math		Reading	
ACEXPost	0.250***	0.246***	0.120**	0.114*	0.284***	0.310***	0.182*	0.203**
	(0.0754)	(0.0666)	(0.0567)	(0.0575)	(0.107)	(0.0923)	(0.100)	(0.0819)
ACE	-0.109***	-0.0634	-0.0991***	-0.0654*	-0.0812	-0.0874**	-0.0583*	-0.0618**
	(0.0408)	(0.0387)	(0.0305)	(0.0335)	(0.0532)	(0.0408)	(0.0334)	(0.0281)
Post	0.000704	-0.00752	0.00895	0.00101	-0.00463	-0.0201	-0.0389	-0.0563
	(0.0377)	(0.0358)	(0.0302)	(0.0312)	(0.0653)	(0.0592)	(0.0451)	(0.0456)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	54,751	54,751	55,267	55,267	48,639	48,639	48,984	48,984

Notes: Robust standard errors clustered by school-year are in parentheses; * p<0.1; ** p<0.05; *** p<0.01; The sample is limited to ACE and near-ACE schools. Controls include indicators for student race-ethnicity, gender, special education status, free or reduced price lunch eligibility, and classification as limited English proficient.

Table 3. Changes in selected student demographic and program characteristics following the implementation of ACE

Comparison periods	2016 to 2018 minus 2012 to 2015		2018 minus 2017	
	ACE Wave 1	near-ACE	ACE Wave 2	Near-ACE
Share low income	0.01	-0.01	0.04	-0.02
Share white	0.00	0.00	0.00	0.00
Share black	0.00	-0.03	0.03	0.00
Share Hispanic	0.01	0.03	-0.03	0.00
Special special education	-0.02	-0.01	0.01	0.00
Share limited English proficient (LEP)	0.03	0.04	-0.02	0.01
Observations	3,424	7,890	1,596	4,579

Table 4. Teacher transitions by origin and destination school type and year

Origin	Destination			
	ACE	near-ACE	other Dallas ISD schools	Out of district
2015 to 2016				
Wave 1 ACE	58	13	98	121
near-ACE	8	407	46	190
other Dallas ISD schools	156	33	7,212	1,903
Out of district	66	180	1,944	n.a.

Table 5. Average Performance Points for Wave 1 ACE and Near-ACE Category A & B Teachers, by Transition Status: 2015 and 2016

	7 ACE schools		15 Near-ACE schools	
Transition status				
	2015	2016	2015	2016
Remain in sector				
A or B in both years	0.07	0.22	-0.11	-0.18
	32		263	
Leave sector				
A or B in 2015 only	-0.72	N.A.	-0.70	N.A.
	74		125	
A or B in both years	-0.31	-0.14	-0.50	-0.14
	74		37	
Enter sector				
A or B in 2016 only	N.A.	-0.16	N.A.	-0.8
	52		126	
A or B in both years	0.56	0.49	-0.72	-0.77
	130		42	

Table 6. Average Performance Points for Wave 1 ACE and Near-ACE Teachers with Value-Added Estimates, by Transition Status: 2015 and 2016				
	7 ACE schools		15 Near-ACE schools	
	2015	2016	2015	2016
Remain in sector				
VA in both years	-0.18	-0.09	-0.06	-0.26
	11		79	
VA in 2015	0.00	0.89	-0.26	-0.17
	5		19	
VA in 2016	1.81	0.64	0.21	0.07
	1		26	
Leave sector				
VA in 2015	-0.74	N.A.	-0.84	N.A.
	38		64	
VA in both years	-0.42	-0.23	0.38	0.15
	10		8	
Enter sector				
VA in 2016	N.A.	0.05	N.A.	-0.95
	25		54	
VA in both years	0.51	0.56	-.94	-0.90
	48		15	

Table 7. Average Value Added for Teachers in Wave 1 ACE and Near-ACE Schools with Value Added Estimates, by Transition Status: 2015 and 2016								
	Mathematics				Reading			
	7 ACE schools		15 Near-ACE schools		7 ACE schools		15 Near-ACE schools	
	2015	2016	2015	2016	2015	2016	2015	2016
all teachers	-0.15	0.09	-0.05	-0.07	-0.17	0.00	-0.09	-0.02
Remain in sector								
	-0.01	0.13	-0.10	-0.02	-0.08	0.10	-0.08	-0.04
	5		31		5		39	
Leave sector								
VA in 2015	-0.16	n/a	-0.02	n/a	-0.18	n/a	-0.08	n/a
	20		34		28		37	
VA in both years	-0.16	0.02	0.17	0.19	-0.19	0.00	0.42	-0.04
	6		5		12		1	
Enter sector								
VA in 2016	n/a	-0.04	n/a	-0.09	n/a	0.07	n/a	-0.01
	15		29		15		39	
VA in both years	0.12	0.19	-0.12	-0.02	0.04	-0.05	-0.22	0.03
	19		3		20		6	

Table 8. Percentage of Teachers with Value-added Estimates in Wave 1 ACE and Near-ACE Schools with 0 or 1 Year of Prior Experience, by Transition Status: 2015 and 2016

	Mathematics				Reading			
	7 ACE schools		15 Near-ACE schools		7 ACE schools		15 Near-ACE schools	
	2015	2016	2015	2016	2015	2016	2015	2016
all teachers								
% 0 years	32.3	28.2	15.7	10.9	33.3	4.9	19.5	9.5
% 1 year	9.7	5.1	12.9	17.2	28.9	17.1	16.9	17.9
Remain in sector								
% 0 years	40.0	0.0	12.9	0.0	40.0	0.0	28.2	0.0
% 1 year	0.0	40.0	19.4	12.9	20.0	40.0	7.7	28.2
Leave sector								
VA in 2015								
% 0 years	35.0	n/a	20.6	n/a	17.9	n/a	10.8	n/a
% 1 year	15.0	n/a	8.8	n/a	35.7	n/a	27.0	n/a
VA in both years								
% 0 years	16.7	n/a	0.0	0.0	66.7	0.0	0.0	0.0
% 1 year	0.0	16.67	0.0	0.0	16.7	66.7	0.0	0.0
Enter sector								
VA in 2016								
% 0 years	n/a	73.3	n/a	24.1	n/a	13.3	n/a	20.5
% 1 year	n/a	0.0	n/a	20.7	n/a	13.3	n/a	2.6
VA in both years								
% 0 years	n/a	0.0	33.3	0.0	15.0	0.0	50.0	0.0
% 1 year	10.5	0.0	n/a	33.3	25.0	15.0	0.0	50.0

Table 9. Average performance score and family perception of principals in Wave 1 ACE Schools, by timing of exit and entry

Principal transition status	exit school immediately prior to ACE implementation			enter ACE school following program implementation		
		Shares of parents who strongly agree that			Shares of parents who strongly agree that	
	performance score	what my child learned this year is what he or she needed to learn to be ready or the next grade	My child's school has a safe learning environment	performance score	what my child learned this year is what he or she needed to learn to be ready or the next grade	My child's school has a safe learning environment
Before ACE	-1.317	0.74	0.68			
During ACE	n.a.	n.a.	n.a.	0.88	0.85	0.80

Table 10. Mean difference between the average difference between actual and predicted achievement in different academic years and the average difference between actual and predicted achievement in 2004-2005, by quartile of predicted achievement in 2004-2005 and quartile of the difference between predicted and actual achievement in 2004-2005: 2010-2011 and 2006-2007

		Quartile based on difference in 2004-2005 within each row			
		1st	2nd	3rd	4th
AY 2010-2011 minus AY 2004-2005					
Quartile based on predicted achievements in 2004-2005	1st	0.23	0.03	-0.09	-0.25
	2nd	0.19	0.01	-0.07	-0.18
	3rd	0.20	0.03	-0.04	-0.13
	4th	0.15	0.05	-0.04	-0.11
AY 2006-2007 minus AY 2004-2005					
Quartile based on predicted achievements in 2004-2005	1st	0.12	0.01	-0.02	-0.10
	2nd	0.08	0.00	-0.04	-0.08
	3rd	0.09	0.02	-0.03	-0.06
	4th	0.08	0.02	-0.01	-0.05

Appendix Table A1. Average Staffing by Role and Subject in Ace and Near-Ace Schools: 2015 and 2016

	ACE		Near ACE	
	2015	2016	2015	2016
number of instructional coaches	1.7	3.1	1.9	2.0
ratio of students/counselor	440	349	515	492
ratio of students/classroom teachers	14.4	13.5	15.0	14.7
ratio of students/math teachers	70.4	67.7	73.6	71.4
ratio of students/reading teachers	63.0	61.9	64.3	62.7
ratio of students/regular teachers	17.6	16.9	19.2	17.9
ratio of students/non-regular teachers	130	223	181	334

Appendix Table A2. Share of Students Who Strongly Agree that their Teacher Supports Learning, by timing of teacher entry and exit

teacher transition status	exit an ACE school immediately prior to ACE implementation	enter an ACE school following program implementation	remain in ACE school following program implementation
2015	0.72	0.83	0.78
2016	n.a.	0.82	0.72