EdWorkingPaper No. 23-772

Attracting and Retaining Highly Effective Educators in Hard-to-Staff Schools

Andrew Morgan
University College London

Minh Nguyen
Ball State University

Eric Hanushek
Stanford University

Ben Ost
University of Illinois at Chicago

Steven Rivkin
University of Illinois at Chicago

Efforts to attract and retain effective educators in high poverty public schools have had limited success. Dallas ISD addressed this challenge by using information produced by its evaluation and compensation reforms as the basis for effectiveness-adjusted payments that provided large compensating differentials to attract and retain effective teachers in its lowest achievement schools. The Accelerating Campus Excellence (ACE) program offers salary supplements to educators with records of high performance who are willing to work in the most educationally disadvantaged schools. We document that ACE resulted in immediate and sustained increases in student achievement, providing strong evidence that the multi-measure evaluation system identifies effective educators who foster the development of cognitive skills. The improvements at ACE schools were dramatic, bringing average achievement in the previously lowest performing schools close to the district average. When ACE stipends are largely eliminated, a substantial fraction of highly effective teachers leaves, and test scores fall. This highlights the central importance of the performance-based incentives to attract and retain effective educators in previously low-achievement schools.

VERSION: May 2023

Suggested citation: Morgan, Andrew J., Minh Nguyen, Eric A. Hanushek, Ben Ost, and Steven G. Rivkin. (2023). Attracting and Retaining Highly Effective Educators in Hard-to-Staff Schools. (EdWorkingPaper: 23-772). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/svep-pd41

Attracting and Retaining Highly Effective Educators in Hard-to-Staff Schools

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

March 2023

Abstract

Efforts to attract and retain effective educators in high-poverty public schools have had limited success. Dallas ISD addressed this challenge with information produced by its evaluation system to offer large, compensating differentials to highly effective teachers willing to work in its lowest-achievement schools. The Accelerating Campus Excellence (ACE) program resulted in immediate and sustained achievement increases. The improvements were dramatic, bringing average achievement in the previously lowest performing schools close to the district average. When ACE stipends are largely eliminated, a substantial fraction of highly effective teachers leaves, and test scores fall. This highlights the central importance of performance-based incentives.

This research was supported by grants from the Arnold Foundation and from the CALDER Research Network. The analysis uses confidential data supplied by the Dallas Independent School Districts (DISD). Individuals wishing to use these data must apply for access to DISD.

*University College London; **Ball State University; ***Stanford University, University of Texas at Dallas, and NBER; ****University of Illinois at Chicago; *****University of Illinois at Chicago, University of Texas at Dallas, and NBER

1. *Introduction*

Elevating the quality of instruction in high-poverty public schools is among the most pressing challenges for public schools, but there are few if any examples of scalable policies with demonstrated success. A key challenge is that most policies designed to increase the supply of teachers to hard-to-staff schools do not increase teacher quality when administrators can neither identify and select high-quality applicants nor provide them with monetary inducements. We investigate a program of highly differentiated incentives designed to attract very effective teachers to the lowest-achieving schools. We show that this intervention not only attracts and retains effective teachers but also significantly increases both current elementary school and future middle school achievement

In 2016, the Dallas Independent School District (DISD) addressed the challenge of attracting and retaining highly effective teachers in schools serving disadvantaged children faced by many districts across the country by providing large compensating differentials to attract and retain effective teachers in its lowest achievement schools. The Accelerating Campus Excellence (ACE) program used information produced by its recently introduced personnel evaluation and compensation system to increase the salaries of high performing teachers who were willing to work in the lowest performing schools.[1] Critically, the financial inducements were based on past performance and were large. Highly effective teachers could increase their base salary by as much as $10,000 per year. Moreover, existing educators who wished to remain in ACE schools had to undergo a rigorous screening process to retain a position. In practice, less than 20 percent of existing teachers were retained following ACE implementation.

This path-breaking intervention emerges directly from consideration of well-known labor-market fundamentals combined with extensive evidence on the variation in teacher effectiveness among schools and districts.[2] Districts have not used large salary inducements to

---

[1] The ACE program included elementary and middle schools, and we focus on elementary schools to enable analyses of immediate effects and effects on achievement at the next schooling level. Beginning in 8th grade, math testing varies by the grade in which a student takes algebra, complicating the analysis of middle school effects on high school achievement.

[2] The issues of rigid salary schedules unrelated to effectiveness have been discussed frequently and over a long period; see Kershaw and McKean (1962) and Hanushek and Rivkin (2004). An extensive literature including Rivkin, Hanushek, and Kain (2005) documents the substantial variation in teacher effectiveness; see also Hanushek and Rivkin (2010, (2012). There is also substantial evidence on the attraction to teachers of both higher pay and higher achieving students (Hanushek, Kain, and Rivkin (2004)).

staff entire schools with highly effective teachers largely because of the lack of comprehensive performance measures, an unwillingness to differentiate pay by teacher effectiveness, and an unwillingness to increase compensation enough to offset labor-market disadvantages of low-achieving, high-poverty schools.[3]

We evaluate the ACE intervention using a difference-in-differences design where other low-performing Dallas schools serve as a control group. ACE schools were specifically selected using test scores two years before implementation. To mimic this selection process, we form a group of control schools based on school-level test scores two years prior to ACE implementation. This selection process for the control group guards against common negative shocks during the period between ACE school selection and program implementation. While our control group is necessarily less negatively selected than the actual ACE schools, these schools follow remarkably similar pre-treatment trends both before program selection and in the run up to program operations.[4] Analytically, decisions about the exact size and composition of the control group have little influence on the results because the improvements at ACE schools prove so dramatic.

We find that the ACE program leads to large immediate improvements in academic achievement at ACE schools, and these program impacts continue into middle school (6th grade) for those with two or more years of treatment. The ACE effects on 6th grade scores increase as the dosage increases from two to three years in an ACE elementary school. The estimated treatment effects on 6th grade scores approaches 0.3 standard deviations in reading and exceeds 0.4 standard deviations in math.

Interestingly, students with only one year of ACE attendance end up with substantially higher 5th grade scores but without a similar increase in subsequent 6th grade scores. The divergence between the time-pattern of contemporaneous and longer-term effects suggests that end-of-year test scores may be more immediately malleable than the accumulation of lasting

[3] The closest policy is IMPACT, which pays larger performance stipends to high quality teachers in disadvantaged Washington DC schools (Dee and Wyckoff (2015)). Unlike ACE, IMPACT did not include the transformational element of rapidly re-assigning a majority of educators in the first year of the program.

[4] Forming the control group based on t-2 test scores has a similar logic to estimating synthetic difference-in-differences, but our approach has two important advantages. First, our selection process is based on institutional knowledge of the actual selection process and is based on a single pre-policy year, limiting concerns of overfitting. Second, because we only use t-2 test scores to identify the control group, we are able to provide more meaningful evidence on pretrends on either side of t-2 than would be possible if we matched on all pre-period outcomes.

skills.[5] The effect of elementary ACE attendance on 6th grade scores are large compared to most contemporaneous interventions and are remarkable given the common observation that most elementary school interventions fade out rapidly.

Two subsequent policy changes provide further evidence on the efficacy of the ACE intervention. First, two years after the initial ACE implementation, the district expanded the program, allowing us to evaluate the effect of ACE for a second wave (dubbed ACE 2). We find very similar effects for this second wave of schools, supporting the notion that the program is scalable. Second, in 2019, the district eliminated most elements of the program for all but one school in the original 2016 ACE cohort because the achievement growth pushed them above the eligibility thresholds for the program. Following the elimination of ACE incentives for the first cohort of ACE schools there was an exodus of highly effective teachers and an immediate reversal of much of the achievement gains. This reversal highlights the central importance of appropriately designed incentives and is further evidence of the efficacy of the ACE program.

We assess the mechanisms for test score improvement by considering whether the ACE gains were plausibly driven by compositional changes of students at ACE schools. Because of limited public discussion prior to ACE implementation and because of the broader history of transitory reforms targeting low-achievement schools, student selection based on ACE is unlikely in the year of adoption, and we find no evidence of changes in student characteristics from 2015 to 2016 at ACE schools. To assess the likely influences of compositional changes in later years, we examine changes in student characteristics following ACE adoption and show that results are robust to intent-to-treat specifications defined by school enrollment in the first year of the program.

Though educator stipends constituted roughly 85 percent of the program budget, the ACE intervention incorporated other components including a small increase in instructional time, a requirement to adopt data-driven instruction, funds for school uniforms and enhanced professional development. We emphasize the central role of teacher quality as this was the focus

[5] An alternative explanation of the one-year increase in scores would be strategic behavior on the part of teachers and administrators. See, for example, the discussion in Figlio and Loeb (2011) that considers the possibility that increases in measured achievement may not increase the stock of human capital. The investigation of 6th grade scores does not generally support purely strategic outcomes. Although 6th grade achievement is incentivized for middle-school teachers and schools through the principal and teacher evaluation systems, the ACE elementary school principals and teachers have no direct incentive to increase 6th grade scores.

of the program, and existing evidence on these other factors suggests that they are unlikely to drive large test score improvements.[6]

2. *Prior evidence on interventions for disadvantaged public schools*

A wide range of programs designed to improve the quality of instruction at traditional public schools for disadvantaged students have been implemented, but the overall record in the United States has not been very successful. 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, Murnane, and Willett (2010); Cowan and Goldhaber (2018); Springer et al. (2010); Springer, Swain, and Rodriguez (2016); and Glazerman et al. (2013).

Clotfelter, Ladd, and Vigdor (2011) found that paying a premium at high-poverty schools successfully improves teacher retention but had little effect on teacher quality because the pay premia differentially attracted teachers with worse credentials. The targeting of academically strong college students through alternative certification programs including Teach for America (TFA) has shown modest success (Boyd et al. (2006), Kane, Rockoff, and Staiger (2008)), and a randomized controlled trial found that $10,000 per year payments to teachers who had high value-added succeeded in attracting a small number of effective teachers to designated high-poverty schools and only modestly raised achievement.[7] Washington, D.C. School District has offered sizeable rewards to teachers who succeeded in generating substantial achievement increases in the most disadvantaged schools through the IMPACT reform. Although this strengthened incentives and made those schools more attractive, it was not implemented in a

[6] Fryer (2014) uses an RCT to investigate the effects of a bundle of best practices including substantially increased instruction time, daily high dosage tutoring of three students per adult, data-driven instruction, replacement of principals, replacement of teachers with low value-added or negative classroom observations, staff evaluation and feedback, and establishment of a culture of high expectations on elementary school achievement including contracts signed by parents and schools in another large, urban district in Texas. The results based on Texas state tests showed that this bundle raised math achievement by roughly 0.15 standard deviations but had little or no significant effect on reading achievement. Importantly, that intervention included the use of information to select teachers for removal and far more extensive and intensive non-educator quality components than those incorporated in the ACE program. Assuming the effect of non-educator components in the ACE program are not substantially more effective than those considered by Fryer (2014), the non-educator quality components of ACE are unlikely to drive a substantial portion of the overall program effect. See also Lavy (2015); Rivkin and Schiman (2015); Hayes and Gershenson (2016); Bietenbeck and Collins (2020); Yeşil Dağlı (2019)).

[7] Glazerman, Mayer, and Decker (2006) find that TFA teachers have higher value-added for math (0.15 student standard deviations) and are equally effective for reading. Glazerman et al. (2013) report on the randomized controlled trial of targeted salary stipends.

manner that enabled the identification of the targeted pay effects on achievement or the quality of instruction (Dee and Wyckoff (2015)).

The literature on school turnarounds is also relevant because the ACE intervention includes the replacement of the existing principal and a large share of the teachers. Existing research reports mixed results of school turnarounds, with some studies finding null or negative effects (Heissel and Ladd (2018); Dougherty and Weiner (2019) and other studies finding significant gains from turnarounds (Dee (2012); Strunk et al. (2016); Papay (2015); Schueler, Goodman, and Deming (2017); Zimmer, Henry, and Kho (2017)). The magnitudes of the achievement effects found by Schueler, Goodman, and Deming (2017) are similar to those in our study, though that intervention involved an intensive tutoring program for struggling students during school vacations that appears to be a primary driver of the improvement.

Importantly, for all the above contexts, there is little or no evidence on whether any observed effects on achievement persist beyond the academic year affected by the policy. For policies that incentivize test scores directly, if achievement gains fail to persist, it raises the possibility that the improvements were the result of teaching to the test.[8] Even in cases where test scores are not incentivized, if achievement improvements fail to persist, evaluations based only on the contemporaneous effect may overstate program efficacy. Our analysis provides the first evidence on whether a transformative teacher quality reform targeting low-achieving schools can generate persistent improvements to the next schooling level.

2.a. Dallas ISD's personnel reforms

Dallas ISD is a large urban school district in north Texas comprised of roughly 160,000 students in 230 schools. After a lengthy development process, the district dramatically changed its personnel system beginning in 2013 by replacing its traditional salary schedule with a compensation system based on performance-based evaluations. After those systems were fully implemented, the district introduced the evaluation-based incentives for working in disadvantaged schools (ACE) that are analyzed here.

The personnel reforms first focused on principals with the introduction of the Principal Excellence Initiative (PEI) in 2013. Two years later, the reforms extended to teachers under the

[8] Gilraine and Pope (2021) contrasts teacher effects on contemporaneous and future test scores to isolate permanent influences, and Dinnerstein and Opper (2022) demonstrate that educators respond to policies that elevate the importance of end-of-year test scores.

Teacher Excellence Initiative (TEI). The changes for teachers are most relevant for our current analysis, though ACE also provided financial incentives to induce effective principals to lead participating ACE schools. DISD moved from a traditional teacher salary schedule based on years of experience and academic degrees to an evaluation and salary system more closely related to performance. The complex set of evaluations incorporated test-based measures of value-added and achievement relative to others teaching comparable students, an extensive observational component using prescribed rubrics for evaluation, supervisor ratings, and student surveys. Specifics varied by subject, grade and other factors, and Appendix A provides details of these evaluation and compensation reforms. Importantly, the separate component scores are aggregated to a single evaluation score that is used to rank teachers (principals) and place the educators into bins that are the primary determinant of salaries.[9] TEI divides teachers into four major groupings with fixed proportions in each category: Exemplary, (1 category), Proficient (3 categories), Progressing (2 categories), and Unsatisfactory (1 category). PEI uses a similar method to allocate principals. After an intensive informational campaign and extensive administrator training, these reforms were implemented uniformly across the entire district. Hanushek et al. (2023) investigate the reform effects on math and reading achievement using synthetic control methods and find that average achievement in Dallas ISD increased by roughly 0.2 standard deviations in math and 0.1 standard deviations in reading following implementation of the reforms.

2.b. Accelerating Campus Excellence (ACE)

The TEI and PEI evaluation information was central to implementation of the incentive program that offered financial inducements for effective educators to work in schools serving disadvantaged students. In academic year 2015-2016, one year following TEI adoption, Dallas ISD introduced the Accelerating Campus Excellence (ACE) program that was designed to raise the quality of instruction and achievement in chronically low-performing schools. Our description below focuses on ACE implementation in elementary schools.

Although this intervention incorporates several additional components, attracting and retaining effective educators is the centerpiece of the program. Educators who applied and were

[9] The linkage to salaries followed a hold-harmless period used to move from the traditional salary system to the new system. In addition, placement in the highest rating categories requires participation in the distinguished teacher review process. Finally, early career teachers are not eligible for the higher ratings categories.

selected to work at ACE campuses received signing bonuses of $2,000 and annual stipends that depended upon position and, for teachers, on TEI effectiveness rating for the prior year. Stipend amounts equaled $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.

Based on the distribution of ratings underlying the TEI reforms, approximately 20 percent of Dallas ISD teachers would qualify for the $10,000 pay premium by having passed a distinguished teacher review, 40 percent of teachers would qualify for an $8,000 pay premium by obtaining a proficient rating, and 37 percent would qualify for a $6,000 premium by receiving a progressing rating due either to being new to the district, to having only one or two years of prior experience, or to failing to reach proficiency. By comparison, 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 the first year of the ACE program, illustrating the district's success at recruiting teachers from the upper part of the effectiveness distribution.

The program required all existing teachers in ACE schools to re-apply for positions and to be interviewed along with the other applicants. Approximately 80 percent of teachers and all principals in a school newly designated as ACE were different from the teachers and principals who had been in the school the previous year. Such turnover by itself would be expected to adversely affect the quality of instruction, as teachers adjust to different schools and in many cases different grades.[10] Thus, any estimates from the program initiation likely place a lower bound on the impacts of an on-going program that is fully imbedded in district operations.

The ACE program was initially rolled out to four elementary schools in 2016 and then an additional five elementary schools in 2018. We refer to the first set of schools as ACE 1 and the second set of schools as ACE 2.[11] In one of the ACE 1 schools (Umphrey Lee), the district found evidence of cheating on the 2013 state standardized test, meaning that the school does not have reliable test records for 2012 and 2013. Because this complicates the examination of pretreatment trends, our baseline analysis excludes Umphrey Lee. However, we show that

[10]Hanushek, Rivkin, and Schiman (2016) find that much of the negative effect of turnover can be explained by lost general and grade-specific human capital. Ost (2014) identifies the substantial returns to recent experience teaching the same grade.

[11]Two of the five ACE 2 schools, Onesimo Hernandez Elementary School and J. W. Ray Learning Center, were consolidated/closed in 2019 complicating outcome analysis for these schools. Our analysis of the second wave of ACE is thus focused on the three schools that exist in both 2018 and 2019. If we instead include the consolidated schools, we cannot examine 2019 outcomes since they no longer exist in 2019, but we find similar 2018 effects.

including Umphrey Lee yields very similar estimated effects.[12] Our preferred analysis includes three ACE elementary schools in the first wave and three in the second wave, raising the possibility that we might have insufficient power to detect moderate effects. In practice as we show below, the ACE effects are large enough that we can statistically reject zero effects at conventional significance levels using permutation test p-values that account for the small number of treated schools.

In 2019, Dallas ISD scaled back the intervention for three of the four ACE 1 elementary schools; the fourth was assigned to a new ACE cohort.[13] The reduced program eliminated the after-school program component, the additional instruction time, and importantly, salary stipends for most teachers. The teachers who continued to receive stipends took on the role of education leaders and worked additional hours in support of professional development for teachers in the school. In other words, the program was no longer pay-for-effectiveness but instead reverted to the more common extra-pay-for-extra-work model. Teacher transitions following 2018 therefore provide evidence on the role of effectiveness-based stipends in teacher retention.

The first wave of ACE schools was selected in 2015 based primarily on 2014 test scores and was treated in 2016. This implementation poses a challenge for a standard difference-in-differences identification strategy, because either idiosyncratic 2014 shocks or downward trends in performance could lead to an ACE designation. A purely transitory negative shock in 2014 could lead to these schools making unusually large recovery-induced gains in 2015 that continue into 2016, the first year of treatment. On the other hand, a downward trend in performance might both lead to the initial ACE designation and continue into 2016. To address these concerns, we construct a control group based explicitly on low performance in 2014. Since our control group is also selected based on low 2014 performance, it also may have experienced negative 2014 shocks or persistent negative trends. Our main analysis defines the control as the lowest 15 percent of non-ACE 1 schools in terms of 2014 average test scores, but we show robustness to varying this definition of low performance. ACE 2 schools were selected primarily based on

[12] The exception is that if we include Umphrey Lee in the pretrend assessment, the ACE schools show a large test score drop from 2013 to 2014 that is not observed in the control. This is because 2013 was the last year that Umphrey Lee's test scores were deceptively high through cheating. Importantly, even with Umphrey Lee included, the treatment and control groups show very similar trends from 2014 to 2015, (t-2 to t-1).
[13] The 2019 ACE cohort was given a much less intensive version of the ACE intervention with more limited stipends. All four ACE 1 schools no longer receive the original ACE intervention in 2019.

low-achievement in 2016 (t-2), and the control group is similarly selected based on 2016 test scores.

3. *Data and Descriptive Statistics*

The analytical database is constructed from several sources. Data on student and staff characteristics come from Dallas ISD administrative data as submitted to the Texas Education Agency. We use the math and reading test scores from the State of Texas Assessments of Academic Readiness (STAAR), which we standardize within Dallas ISD separately by subject, grade, and year to have mean zero and standard deviation one.[14] Other student information includes race, gender, and indicators for students qualifying for programs such as free or reduced price lunch, gifted, special education, and limited English proficiency. Staff information includes role, experience, subject, grade, and school.

We also have access to unique data that include scores and sub-metrics for all the inputs into the TEI evaluations and rating categories along with salaries and stipend amounts. The evaluation data include scores for teacher performance as measured by rubric-based observations, student perceptions reported in surveys of students above the second grade, and achievement. As we describe below, the information used to produce teacher evaluation scores differs by grade and subject taught. The educator stipend data contain information for all ACE-school educators in each year a school participated in the program. We combine the data sets and construct a panel that links teachers, students, and schools together from the 2011-2012 to 2018-2019 school years. Henceforth we use the spring year (e.g. 2019) to reference an academic year.

Table 1 shows descriptive statistics for the two waves of ACE schools and their respective control groups for the year prior to program implementation for each wave. Compared to the controls, ACE schools are lower performing and have a higher percentage of Black students and lower percentages of Hispanic and LEP students. Most students in Dallas are eligible for subsidized lunch, and there is little difference in this between treatment and control.

The immediate objective of the ACE program was to encourage teachers with high evaluation scores to work in ACE schools. The requirement that existing teachers re-apply to continue working at an ACE school led to extensive teacher turnover. In the top and bottom

[14] Test scores used in the teacher value-added analysis are standardized by grade, subject and year at the state rather than the district level.

panels of Figure 1 we show the turnover rate by year for ACE 1, ACE 2, and their respective control groups. Turnover is 80% or higher during program implementation for both waves of ACE, reflecting the fact that relatively few teachers were rehired.

Teacher evaluation scores may increase either because ACE attracts higher quality teachers or because the program raises scores through its effects on achievement or teacher performance based on supervisor observations. To isolate the role of teacher composition, we describe all teachers in terms of their t-1 evaluation rating. Figures 2 and 3 illustrate the rightward shifts in the ratings distributions for ACE schools and the absence of such changes in control schools. For both ACE 1 (Figure 2) and ACE 2 (Figure 3), the shift in teacher evaluation ratings is transformational: Before ACE, the vast majority of teachers were rated in the bottom 3 categories whereas after ACE, far fewer teachers fall in this range. The increase in the share of ACE 1 teachers rated Proficient II or above exceeded 50 percentage points. For ACE 2, a 25 percentage-point increase in the share rated Proficient 1 accompanied a 35 percentage-point increase in the share rated Proficient II. Remarkably, 0 percent of the teachers in these ACE 2 schools had a rating above Proficient I prior to program implementation.

Large reductions in the share of ACE 2 and particularly ACE 1 teachers with no prior experience contribute to their rightward shifts in the ratings distributions. This share dropped by 10 percentage points following the implementation of ACE 2 and a whopping 30 percentage points following the implementation of ACE 1. In contrast, the share of teachers with no prior experience fell by less than 4 percentage points in the ACE 1 and ACE 2 controls. Evidence documents the significantly lower effectiveness of novice teachers, especially those in their first year, and their concentration in high-poverty schools.[15] By attracting effective teachers from the existing workforce the ACE intervention removed an important impediment to raising achievement.

[15] Much research presents evidence on the concentration of novice teachers in schools serving high fractions of low-income and nonwhite students and on returns to early experience. For example, Lankford, Loeb and Wyckoff (2002) provide evidence on the distribution of novice teachers, and Papay and Kraft (2015) present evidence on teacher experience effects.

4. *Effects of ACE on achievement*

We investigate the efficacy of ACE by comparing test score growth at ACE and control schools. We first present the raw test-score data for ACE and control schools for both contemporaneous and future achievement. Subsequently we describe the empirical models used to identify the program effects and to address potential confounding factors. Finally, we present regression results including extensive sensitivity analysis.

a. A simple description of achievement trends

The top and bottom panels of Figure 4 offer strong evidence that the ACE intervention led to dramatic growth in math and reading achievement. Not surprisingly, compared to both the district average and the control schools, average math achievement in ACE 1 schools began very low. Performance in 2014 dipped, consistent with a transitory negative 2014 shock at the already low achieving ACE 1 schools, and we observe a small recovery from 2014 to 2015. Importantly, although we use only the 2014 test scores to select the control schools, they follow a remarkably similar trend to the ACE schools over the entire pretreatment period. Control schools, which by design are not quite as negatively selected, also exhibit low performance in 2014 and a 2014 to 2015 recovery. For ACE 2, the control and treatment groups do not match quite as closely, but it remains the case that they are on very similar overall trends: the gap between ACE 2 and the control is virtually identical in 2012 and 2017 (t-6 and t-1).

The ACE schools show an immediate and very large increase in achievement upon program implementation. The math score increases for both ACE waves exceeded 0.4 standard deviations. Achievement barely changed in either control group. Student outcomes after the termination of the intervention in 2019 at all but one ACE school reinforce the interpretation of the achievement pattern as being driven by the teacher incentives. Scores in ACE 1 schools fall below those seen in the prior three years. The bottom panel of Figure 4 shows a quite similar pattern for reading scores. As generally found, schools and programs have smaller effects on reading than math, and the impacts of ACE are not an exception. Nonetheless, the improvement pattern following ACE implementation is that same as that for math, including the retrenchment after ACE is discontinued in 2019.

There are no obvious alternative explanations for the time pattern of performance after the implementation of ACE. Nevertheless, the possibility remains that other factors including changes in student characteristics contribute to the achievement growth. We therefore undertake a more comprehensive analysis of ACE effects on both contemporaneous and future achievement.

b. Analytical framework

Our test-score analysis uses a student-level event study framework to identify the effects of the ACE program. Throughout, we use the control groups discussed above and exclude other Dallas schools from the analysis. To avoid concerns regarding staggered difference-in-differences designs, we analyze the two waves of ACE separately (Goodman-Bacon (2021)). We first examine how ACE students perform over time, and then move to a dosage analysis that distinguishes students according to the length of time that they are treated (either one, two, or three years of ACE exposure).

The baseline model is a standard event study examining how test scores of ACE students evolve compared to the control. Specifically, we separately estimate for ACE 1 and ACE 2:

$$A_{it} = \alpha_0 + \gamma_0 ACE_{it} + \rho_t + \sum_{k \neq -1} \delta_k D_{it}^k + \epsilon_{it} \tag{1}$$

where A_{it} denotes math or reading test score for student i in year t, ACE_{it} is an indicator for schools in the relevant ACE wave, ρ_t is a year fixed effect and D_{it}^k are event-time indicators that are 1 if the school is k years from treatment and zero otherwise. The event-time indicators omit -1 and span from -4 to +3 for ACE 1 and from -6 to +1 for ACE 2. The parameters of interest are δ_k, which capture the divergence between ACE and the control group in year t+k relative to the omitted year. The time effects span from 2012 to 2019, with the year before treatment taken as the omitted group (2015 for ACE 1 and 2017 for ACE 2). Throughout, we report standard errors clustered at the school level, but because there are relatively few treated schools, we also report p-values based on a permutation test using 500 iterations.

To assess dosage effects, we investigate the effect of n years of treatment (one, two or three years for ACE 1 and one or two years for ACE 2), running separate regressions for each dosage. The estimating equation is identical as equation (1), but we restrict the sample to students who have been at an ACE or control school for at least n years by the end of year t. In the absence of the sample restriction, some students in an ACE 1 school in 2017 had two years of treatment while others who moved to the school prior to 2017 had only one year of treatment. This means that estimates of δ_2, for example, do not identify the effects of three years of treatment in the absence of sample restrictions. Below we discuss steps taken to account for any endogenous selection in school moves.

Importantly, only some of the event-time coefficients are informative of the relevant dosage effect. For example, for n=3, the δ_0 coefficient captures the effect of attending ACE for two pre-treatment years and one treatment year and is consequently not informative of the effect of 3 years of exposure. As such, for the dosage analysis, we are primarily interested in the coefficient δ_{n-1}, which corresponds to the student having exactly n years of exposure following program start. This is measured relative to the omitted relative-time indicator, $k = -1$, corresponding to the last fully untreated cohort with n years of exposure. Importantly, though only some event-time indicators are of interest, we include the entire set of indicators in the model to ensure that only $k = -1$ is the reference year. For the dosage analysis, we are still able to assess pretrends by examining estimates of δ_k for $k < 0$.[16]

Since tests are only available for grades 3-5, and the schools we study are K-5, the one-year of exposure specification include students in grades 3-5 in year t=0, the two-year specification samples include students who were in grades 2-4 in year t=0, and the three-year specification samples include students who were in grades 1-3 in year t=0.[17]

To assess the effect of ACE 1 elementary attendance on test scores at the next schooling level (6th grade), we modify equation (1) so that the outcome is 6th grade rather than contemporaneous test scores. In this case, we are primarily interested in the coefficient δ_{n-1}, which corresponds to the student having exactly n years of elementary-school exposure following program start. For example, when studying the effect of 3 years of exposure, the δ_2 coefficient captures the effect of attending a treated school for grades 3-5 on 6th grade test scores, relative to the last fully-untreated cohort. It is the 6th grade students in 2019 that were treated for three years, and their achievement is compared with achievement for the 6th grade students in 2016, the last untreated cohort. As with the contemporaneous analysis, we can assess pretrends in 6th grade test scores by plotting estimates of δ_k for $k < 1$.[18]

[16] As an example, consider δ_{-2} for the 3 years of exposure analysis. This parameter captures the "effect" of 3 years of exposure between $k = -4$ and $k = -2$ versus having 3 years of exposure between $k = -3$ and $k = -1$.
[17] Longer exposure corresponds to exposure in earlier grades, raising the possibility that heterogeneous treatment effects could contribute to differences by exposure length. Preliminary estimates (not reported) reveal little variation by grade, leading us to combine students across grades.
[18] Some students who attend an ACE elementary school also attend an ACE middle school, but this cannot drive estimates as ACE elementary attendance does not increase the odds of ACE middle school attendance and the rates

Restriction of a sample to students with n successive years at a treatment or control school potentially introduces selection bias, and we therefore also report intent-to-treat estimates based on samples that include all students with n years of potential exposure. In other words, estimates of the effects of three years of treatment include all students in ACE and control school who would be in one of the tested grades two years later if they were to remain in the same school.[19]

To further assess the role of student composition, we modify the event-study model outcome to be predicted, rather than actual achievement. We generate predicted achievement from a two-step process. First, we regress math (reading) achievement on student characteristics separately by grade in the year prior to the beginning of treatment for each ACE wave. Then, we use the coefficients to predict achievement for ACE and control students in grades 3-5 with valid test scores.

Table 2 reports event-study estimates of the "effects" of ACE on predicted math and reading achievement and reveals little evidence of the type of endogenous selection that would introduce substantial bias. None of the coefficients in either the treatment or pre-treatment period are significant at conventional levels, and none are larger in magnitude than 0.04. Moreover, all of the coefficients in the first year of treatment are negative and very small, indicating the absence of any enrollment response in anticipation of the program and therefore any selection bias in the intent-to-treat estimates.

Another potential threat to identification is that the ACE treatment could affect control schools. For example, ACE may adversely affect the quality of educators in control 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 teachers employed in an ACE elementary school in 2016 represent less than 2 percent of Dallas teachers, implying that spillover effects in the teacher labor market are likely to be small. Among all schools in the district, the modal number of teachers lost to ACE schools is zero. For control schools specifically, less than 3% of control-school teachers in 2015 transferred to an ACE school in

---

of overlap in ACE attendance across schooling levels is low. For example, only 2% of those with 3 years of potential ACE elementary attendance attend an ACE middle school.

[19] An alternative to the intent-to-treat approach is instrumenting actual years of exposure with potential years of exposure, but this relies on a linearity assumption (e.g. three years of exposure is exactly three times as beneficial as one year of exposure) that is likely to be violated.

2016. Of course, we cannot rule out the possibility that ACE schools hired some high-quality teachers who otherwise would have ended up at control schools, but the negligible direct movement suggests that the implementation of ACE is unlikely to have a significant effect on the average effectiveness of educators in the control schools.

c. Event-study estimates of contemporaneous effects

Plots of the math and reading event-study coefficients for both waves of ACE shown in Figure 5 reveal little or no evidence of differential trends prior to ACE implementation, and there are sharp increases in test scores in the year of implementation. Table 3 shows the point estimates, standard errors and permutation test p-values corresponding to Figure 5. For math, the estimated effect is approximately 0.5 in both waves, whereas reading estimates are closer to 0.3. Importantly, the much smaller t+3 coefficient for ACE wave 1 corresponds to 2019, the year in which most components of the ACE treatment were removed in these schools. Because of entry to and exit from ACE schools during the treatment periods, ACE 1 and ACE 2 schools include students treated for one, two or three years following the initial program year. Consequently, we turn now to specifications that estimate the effects of specific treatment dosages.

d. Treatment dosage effects

In Table 4, we show the estimated effect of 1, 2 or 3 years of exposure to the ACE program on math achievement (top panel) and reading achievement (bottom panel), separately by ACE wave. Columns 1,2 and 4 report the effects of actual exposure, and Columns 3 and 5 report ITT estimates of potential exposure effects. Since ACE 2 begins in 2018, we observe at most 2 years of exposure, meaning that Columns 4 and 5 are blank for this wave. The exposure effects are estimated as described earlier and are interpreted as the change in outcomes in ACE vs control schools for students who attend for *n* years.

All estimates of ACE effects on math achievement are large and highly significant based on permutation test p-values for both ACE waves. The 1-year-of-exposure coefficients exceed 0.4, and the coefficients on two years of exposure are larger in magnitude than the estimated one-year effect, but the differences are modest and not significant for both waves. The estimated effect of three years of exposure is slightly smaller than that for two years for ACE 1, but the difference is not statistically significant. As expected, the ITT estimates in Columns 3 and 5 are smaller than the corresponding actual exposure coefficients. Nevertheless, they all exceed 0.35 with p values below 0.05. The magnitudes of the ITT estimates are not easily interpretable in absolute terms or relative to one another, but their substantial sizes strongly support the belief that ACE has a large effect on math achievement.

The estimates for reading in the lower panel also reveal little variation by ACE wave, though they are smaller than those for math. Nevertheless, the ITT estimates are all significant at the 0.05 level based on the permutation tests and at least 0.2 standard deviations in magnitude, strong evidence of sizeable ACE treatment effects on reading achievement.

e. Effects on 6th grade test scores

Exposure to ACE for two or more years has a lasting impact that is seen after students transition to middle school. Columns 1, 2 and 4 of Table 5 report coefficients from OLS specifications that measure actual years of exposure, and Columns 3 and 5 report ITT estimates from specifications that measure two or three years of potential exposure. In contrast to the previously seen large effect of 1 year of exposure on contemporaneous math test scores, we see little evidence of improved 6th grade scores for students with 1 year of ACE exposure (top panel). However, students with 2 or 3 years of exposure to an ACE elementary school, show lasting improvements of approximately 0.3-0.4 standard deviations in math achievement that increase with dosage years. Note that these estimates are less precisely estimated than the contemporaneous effects, with permutation test p-values equal to 0.07 and 0.11 for the estimates of effects of two and three years of exposure, respectively. The ITT estimates are again smaller than the estimates for actual exposure and less precisely estimated. Although the magnitudes are large and consistent with positive dosage effects, the lower precision amplifies the importance of the sensitivity analysis for these long-term effects.

In addition to investigation of the robustness of the results to changes in the model and sample in the next section, we now illustrate the effects of three years of exposure on the entire distribution to ensure that the positive effect does not emanate from outliers or large gains for only a small fraction of students. Panel A of Figure 6 illustrates kernel density estimates of the distributions of math achievement for ACE and control students in 2016, the last pre-treatment cohort, (left diagram), and 2019, the fully treated cohort (right diagram). The striking contrast illustrates the rightward shift in the math achievement distribution for ACE students in comparison to students in the control schools between 2016 and 2019: the ACE distribution lies slightly to the left of the control distribution in 2016 but shifts to lie substantially to the right of the control distribution in 2019.

For reading, Column 4 of the lower panel of Table 5 shows that the estimated effects of three years of elementary school ACE exposure are substantial though less precisely estimated than those for math. Consistent with this estimate, Panel B of Figure 6 shows a rightward shift in the distribution of ACE schools relative to controls that is less pronounced than that for math. In addition, the estimated effects of two years of exposure are much smaller and less precisely estimate for reading, as are the ITT estimates for both two and three years of exposure.

To assess whether the common trends assumption holds for the analysis of long-term effects, we present the event-study estimates for three years of ACE exposure in Figure 7; the pattern is very similar if we use potential rather than actual exposure. The plotted coefficients from -4 to -2 do not show significant differences in pretrends for either math or reading. All coefficients for the pre-treatment cohorts are small in magnitude and insignificant relative to the scores in t-1. As discussed earlier, the coefficients for k=1 and k=2 are not informative of the 3 years of exposure effect because these coefficients capture the effect of attending an ACE school during a three-year period that includes both pre and post ACE years. Though the k=1 and k=2 coefficients are not informative of the effect of 3 years of exposure, they do provide estimates of the effect of 1 or 2 years of exposure, albeit for a restricted sample (one that has 3 years of ACE attendance). The coefficient for k=2 shown in Figure 7 reproduces the estimate shown earlier in Table 5.

The divergence between the patterns of contemporaneous and future effects of attendance at an ACE elementary school highlights the importance of a careful consideration of treatment dynamics. On the one hand, the evidence strongly supports the existence of large, immediate ACE effects on contemporaneous elementary school achievement in both waves. On the other hand, the evidence also suggests that only multiple years of exposure to an ACE elementary school raise middle school achievement. Therefore, evidence based on a single year of treatment may not be adequate to understand the potential for an education intervention to affect the acquisition of valued skills that persist into the future.[20]

[20] This magnitude of the one-year ACE effect on math in combination with the limited evidence of further increases of anywhere near the same size in subsequent years raises the possibility that a portion of the effect comes from something other than teacher impacts – improved school organization, testing conditions, or even effort on the tests. Persistent improvements in these dimensions would be expected to increase standardized scores in the first year but not further increase future years if the testing conditions remained at the new higher level.

f. Robustness to alternative control groups

The previous analyses excluded some schools from the treatment waves, used control groups consisting of the 15 percent of remaining district schools with the lowest average scores two years prior to the implementation of the respective waves of ACE, and did not include controls for student characteristics in the regressions. The results below illustrate the robustness of the results to variation in the share of district schools included in the control group, inclusion of student demographic controls in the regressions and inclusion of Umphrey Lee, Onesimo Hernandez and J. W. Ray Learning Center in the treatment groups.

Tables 6 (math) and 7 (reading) report coefficients from regressions that differ along the aforementioned dimensions. For ease of comparison, the top panel in each reports the baseline math and reading estimates from Table 4. Estimates in Panel 2 show that controlling for student covariates has little effect on the estimates for either wave, providing further evidence that compositional changes are not driving the improvement. Similarly, Panels 3 and 4 show that the estimates are not sensitive to the composition of the control group in terms of the cutoff for low achievement (lowest 10% or lowest 20% in place of the lowest 15% of schools). Panel 5 shows that the estimated effects for wave 1 are not sensitive to the inclusion of the second wave of ACE schools in the donor pool for the control group. Our preferred model allows ACE 2 schools to be in the control group for ACE 1 because their exclusion would amount to selection on achievement during the treatment period. (ACE 2 schools were selected based on having very low achievement in 2016). That said, allowing ACE 2 schools to be in the ACE 1 control group has the downside that the control group becomes partially treated in 2018. In practice, the choice about exclusion of ACE 2 schools from the control group is empirically unimportant because they are a small fraction of the control schools. Panel 6 shows that the estimates also change little following the inclusion of Umphrey Lee (the school with the 2013 cheating scandal) in the ACE 1 treatment group. Similarly, Panel 7 shows that the estimates for ACE 2 also remain largely unchanged following the inclusion of the two ACE 2 schools that leave the program in 2019.

Table 7 exhibits a very similar pattern for reading achievement. Again, differences in control groups, the composition of ACE schools or the inclusion of covariates has little effect on any of the estimates for either ACE 1 or ACE 2 regardless of the treatment dosage or focus on actual or potential years of exposure.

Tables 8 and 9 show the robustness of the 6th grade math and reading grade test scores analyses to a variety of empirical choices. Given that the 6th grade estimates were only marginally significant, it would be concerning if they were also sensitive to various choices made in the empirical model. Beginning with the math estimates reported in Table 8, all coefficients fluctuate within very narrow ranges of less than 0.05 standard deviations with similar levels of significance. By comparison, Table 9 shows slightly larger fluctuations in the estimated effects on 6th grade reading achievement, including the estimates of three-year exposure effects.

5. *How important are stipends to the retention of ACE teachers?*

A critical question concerns the necessity of performance-based stipends in the ACE schools over the longer-term. Is just the initial introduction of stipends and public commitment to hiring high-quality teachers sufficient to move the ACE schools from a bad to a much better equilibrium in terms of educator effectiveness? Or is the continuation of stipends necessary to retain high quality teachers at these previously low-performing schools? If effective teachers value collaboration with high-quality peers, it is possible that they may not leave if the stipends are eliminated, even if they would not have moved to ACE schools initially in the absence of the program. Alternatively, if time-invariant factors such as location or a high poverty rate make hinder efforts to attract and retain educators, it is likely that elimination of the stipends would lead to an increase in turnover that comes disproportionately from highly effective teachers due to the larger decline in compensation and their better alternative opportunities.

The previously discussed Figure 5 shows a sharp achievement decline in 2019 following the elimination of the stipends and other ACE 1 program components at 3 of the 4 ACE 1 schools and a scaling back of the program at the fourth school. This decline suggests that the elimination of significant parts of the program including stipends for most teachers reversed much of the benefit of the ACE treatment.

Figure 8 illustrates the changes in turnover by effectiveness level between 2017 and 2018 for all ACE 1 schools (top panel), for all but Pease Elementary, the school that continued to participate in a modified version of the ACE program in 2019 (middle panel), and for the control schools (bottom panel). Among the small number of ACE 1 teachers with an effectiveness level of Progressing II or lower or who did not have an effectiveness level by virtue of not having taught in Dallas ISD in the previous year, turnover following 2018 declined relative to the previous year. In contrast, among teachers with an effectiveness level of Proficient I or higher, turnover almost doubled in the four ACE 1 schools (top panel) and almost tripled in the three schools other than Pease Elementary (middle panel). Over 40 percent of these teachers left ACE 1 schools after 2018 following the elimination of stipends, while less than one quarter left after 2017.

At control schools, the turnover rate was also higher following 2018 compared to 2017, but there is much less evidence of changes in the selectivivity of exits. Turnover increases by approximately 10 percentage points for highly-rated teachers at control schools and it increases by approximately 5 percentage points for lower-rated teachers at control schools. At the ACE schools that lost most elements of the program (middle panel), turnover increased by approximately 30 percentage points for highly rated teachers and turnover *decreased* by approximately 20 percentage points for lower-rated teachers.

6. Summary and Conclusions

Improving achievement in low-performing urban schools has been identified as a high priority for education policy. Yet, a range of tactics have been employed over the past half century with limited overall success. Dallas ISD addressed this challenge by using the information produced by its evaluation and compensation reforms as the basis for effectiveness-adjusted payments that provided the compensating differentials to attract and retain effective teachers in the lowest achievement schools. By adopting a rigorous screening process included information produced by the Dallas ISD multi-measure evaluation system and interviews, the district leveraged the supply increase induced by the targeted salary increases.

Roughly 80 percent of the teachers and all principals were replaced at the schools in each ACE wave, and the treated schools saw dramatic achievement increases. This improvement was largest in math, but the reading improvements were also substantial. Importantly, attendance at an ACE elementary school for two or more years led to large increases in achievement following matriculation to middle school (6^{th} grade). The existence of large long-term effects suggest that the program produces lasting improvements in cognitive skills. Moreover, the second wave of ACE showed similar effects to the first wave, suggesting that the program can be scaled.

The program as implemented brought the average achievement in the previously lowest performing schools close to the district average. But doing so required extra funds to pay the compensating differentials that appear to be key to attract and retain highly effective teachers at the ACE schools. When the stipends paid to ACE 1 educators were largely removed following the achievement increases, turnover jumped among the most effective teachers and test scores fell substantially.

References

Bietenbeck, Jan, Matthew Collins (2020). New Evidence on the Importance of Instruction Time for Student Achievement on International Assessments Working Paper No. 2020:18. Department of Economics: Lund University (September).

Boyd, Don, Pam Grossman, Hamilton Lankford, Susanna Loeb, James Wyckoff (2006). How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement. *Education Finance and Policy* 1 (2): 176-216.

Clotfelter, Charles T., Elizabeth Glennie, Helen F. Ladd, Jacob L. Vigdor (2008). Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina. *Journal of Public Economics* 92 (5-6): 1352-1370.

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

Cowan, James, Dan Goldhaber (2018). Do Bonuses Affect Teacher Staffing and Student Achievement in High Poverty Schools? Evidence from an Incentive for National Board Certified Teachers in Washington State. *Economics of Education Review* 65: 138-152.

Dee, Thomas (2012). School Turnarounds: Evidence from the 2009 Stimulus. NBER Working Paper No. 17990. Cambrdige, MA: National Bureau of Economic Research (April).

Dee, Thomas S., James Wyckoff (2015). Incentives, Selection, and Teacher Performance: Evidence from Impact. *Journal of Policy Analysis and Management* 34 (2): 267-297.

Dougherty, Shaun M., Jennie M. Weiner (2019). The Rhode to Turnaround: The Impact of Waivers to No Child Left Behind on School Performance. *Educational Policy* 33 (4): 555-586.

Figlio, David, Susanna Loeb (2011). School Accountability. In *Handbook of the Economics of Education, Vol. 3*, edited by Eric A. Hanushek, Stephen Machin, Ludger Woessmann. Amsterdam: North Holland: 383-421.

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

Glazerman, Steven, Daniel Mayer, Paul Decker (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.

Glazerman, Steven, Ali Protik, Bing-ru Teh, Julie Bruch, Jeffrey Max (2013). *Transfer Incentives for Highperforming Teachers: Final Results from a Multisite Randomized Experiment*, NCEE 2014-4003. Washington, DC: U.S. Department of Education (November).

Goodman-Bacon, Andrew (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics* 225 (2): 254-277.

Hanushek, Eric A, Jin Luo, Andrew Morgan, Minh Nguyen, Ben Ost, Steven G. Rivkin, Ayman Shakeel (2023). The Effects of Comprehensive Educator Evaluation and Pay Reform on Achievement. Unpublished manuscript (March).

Hanushek, Eric A., John F. Kain, Steve G. Rivkin (2004). Why Public Schools Lose Teachers. *Journal of Human Resources* 39 (2): 326-354.

Hanushek, Eric A., Steven G. Rivkin (2004). How to Improve the Supply of High Quality Teachers. In *Brookings Papers on Education Policy 2004*, edited by Diane Ravitch. Washington, DC: Brookings Institution Press: 7-25.

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

Hanushek, Eric A., Steven G. Rivkin (2012). The Distribution of Teacher Quality and Implications for Policy. *Annual Review of Economics* 4: 131-157.

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

Hayes, Michael S., Seth Gershenson (2016). What Differences a Day Can Make: Quantile Regression Estimates of the Distribution of Daily Learning Gains. *Economics Letters* 141: 48-51.

Heissel, Jennifer A., Helen F. Ladd (2018). School Turnaround in North Carolina: A Regression Discontinuity Analysis. *Economics of Education Review* 62: 302-320.

Kane, Thomas J., Jonah E. Rockoff, Douglas O. Staiger (2008). What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City. *Economics of Education Review* 27 (6): 615-631.

Kershaw, Joseph A., Roland N. McKean (1962). *Teacher Shortages and Salary Schedules*. NY: McGraw-Hill.

Lavy, Victor (2015). Long Run Effects of Free School Choice: College Attainment, Employment, Earnings, and Social Outcomes at Adulthood. NBER Working Paper 20843. Cambridge, MA: National Bureau of Economic Research

Papay, John P. (2015 of Conference). The Effects of School Turnaround Strategies in Massachusetts. Paper presented at *APPAM Fall Research Conference*, November, at Miami.

Rivkin, Steven G., Eric A. Hanushek, John F. Kain (2005). Teachers, Schools, and Academic Achievement. *Econometrica* 73 (2): 417-458.

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

Schueler, Beth E., Joshua S. Goodman, 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, Brian M. Stecher (2010). *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives, Vanderbilt University.

Springer, Matthew G., Walker A. Swain, Luis A. Rodriguez (2016). Effective Teacher Retention Bonuses: Evidence from Tennessee. *Educational Evaluation and Policy Analysis* 38 (2): 199-221.

Steele, Jennifer L., Richard J. Murnane, John B. Willett (2010). Do Financial Incentives Help Low-Performing Schools Attract and Keep Academically Talented Teachers? Evidence from California. *Journal of Policy Analysis and Management* 29 (3): 451-478.

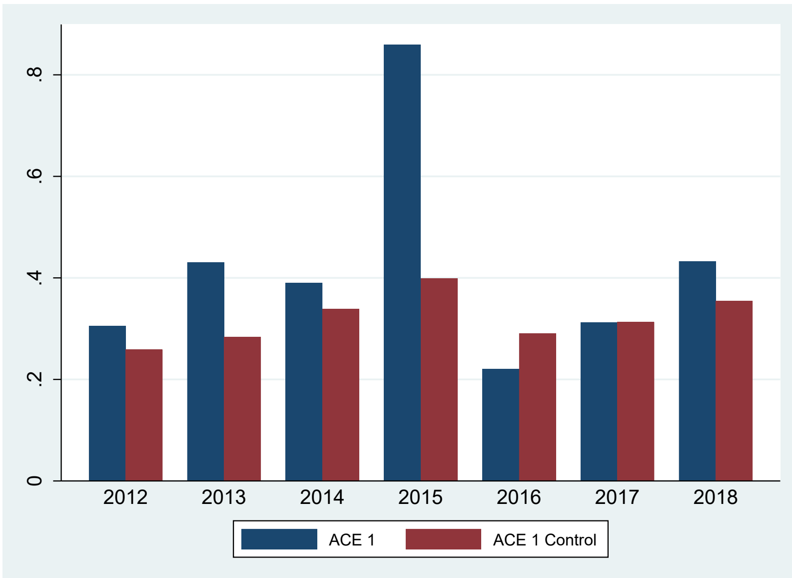
Strunk, Katharine O., Julie A. Marsh, Ayesha K. Hashim, Susan Bush-Mecenas, Tracey Weinstein (2016). The Impact of Turnaround Reform on Student Outcomes: Evidence and Insights from the Los Angeles Unified School District. *Education Finance and Policy* 11 (3): 251-282.

Yeşil Dağlı, Ümmühan (2019). Effect of Increased Instructional Time on Student Achievement. *Educational Review* 71 (4): 501-517.

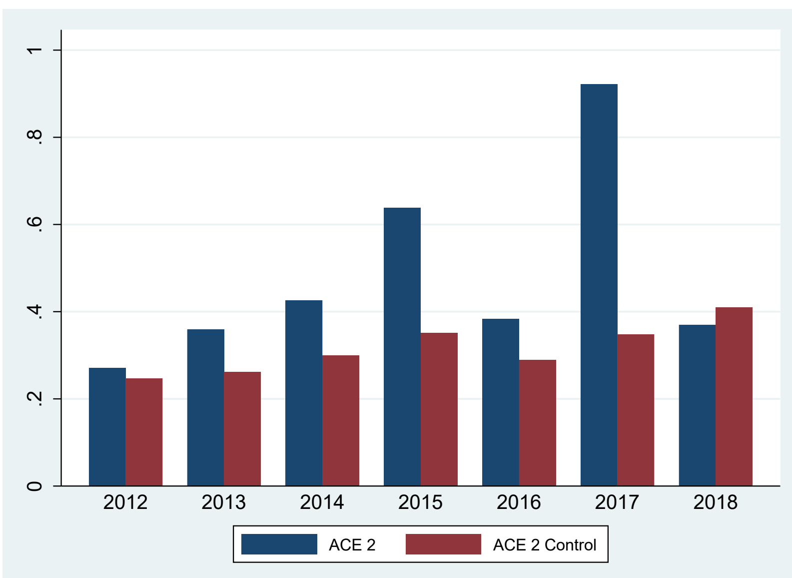
Zimmer, Ron, Gary T. Henry, Adam Kho (2017). The Effects of School Turnaround in Tennessee's Achievement School District and Innovation Zones. *Educational Evaluation and Policy Analysis* 39 (4): 670-696.

Figure 1: Share of teachers who exit ACE and control schools following the school year, by ACE wave: 2012 to 2018

A. ACE 1



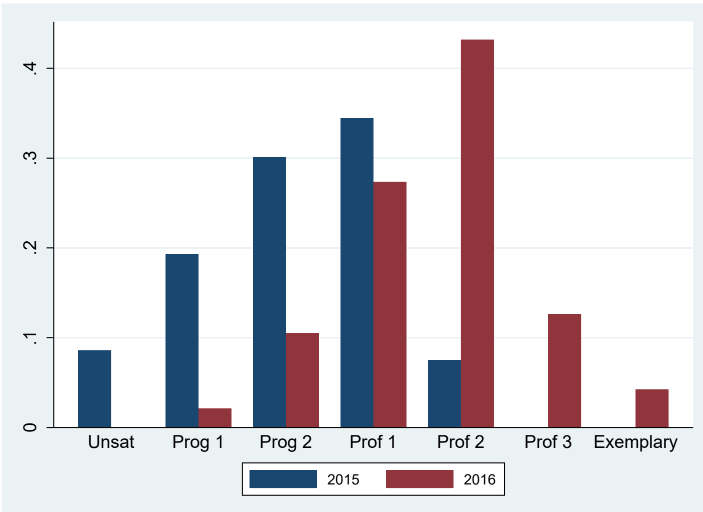
B. ACE 2



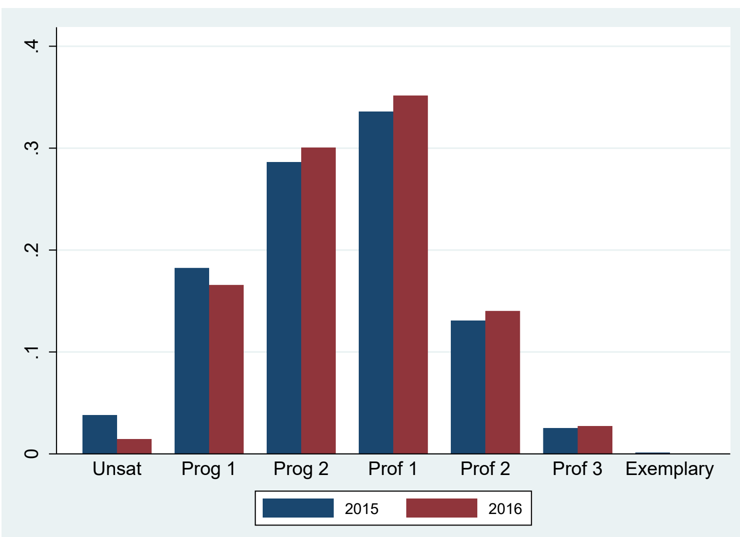
Notes: ACE 1 excludes Umphrey Lee Elementary School, and ACE 1 controls are the lowest 15 percent in terms of 2014 achievement excluding schools eventually included in ACE 2. ACE 2 excludes Onesimo Hernandez Elementary School and J.W. Ray Learning Center that leave the program in 2019, and ACE 2 controls are the lowest 15 percent in terms of 2016 achievement.

Figure 2. 2015 Evaluation ratings distribution for teachers in ACE 1 and control schools in 2015, the year prior to treatment, and in 2016, the first year of treatment

A. ACE 1



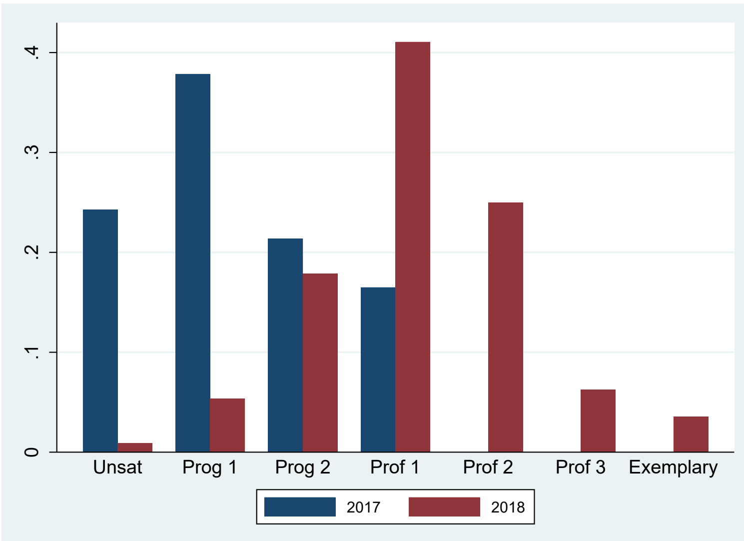
B. Controls



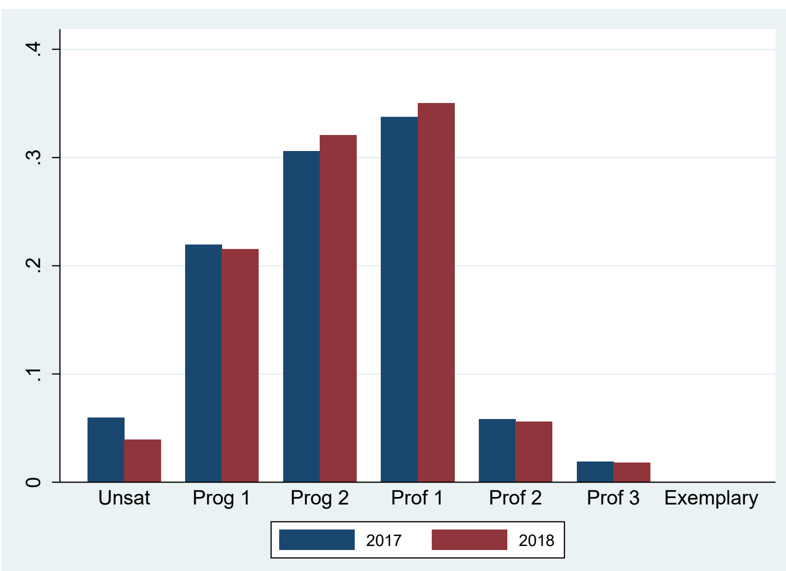
Notes: ACE 1 includes Umphrey Lee Elementary School, and ACE 1 control schools are the lowest 15 percent non-ACE 1schools in terms of 2014 achievement.

Figure 3. 2017 Evaluation ratings distribution for teachers in ACE 2 and control schools in 2017, the year prior to treatment, and in 2018, the first year of treatment

A. ACE 2

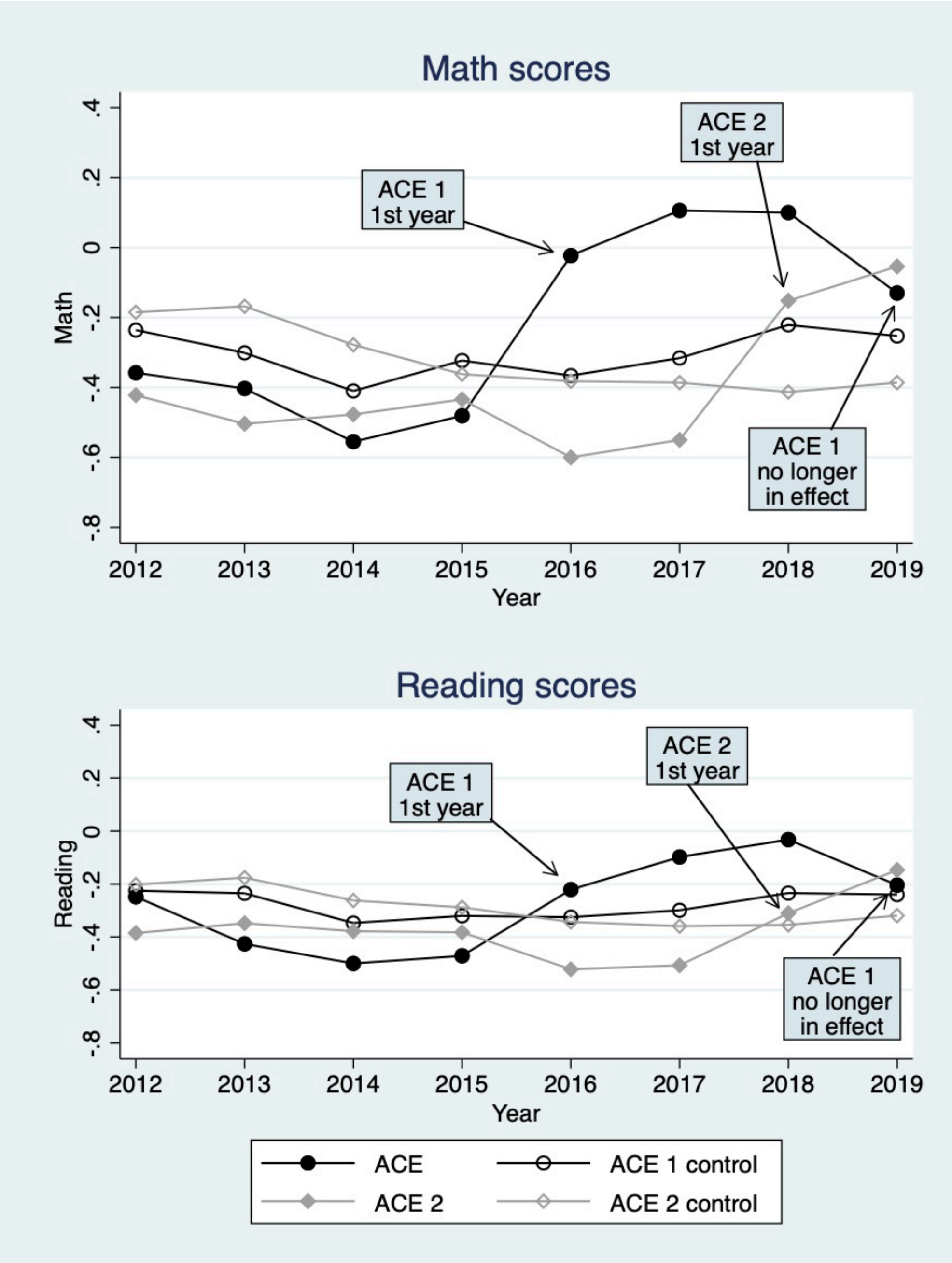


B. Controls



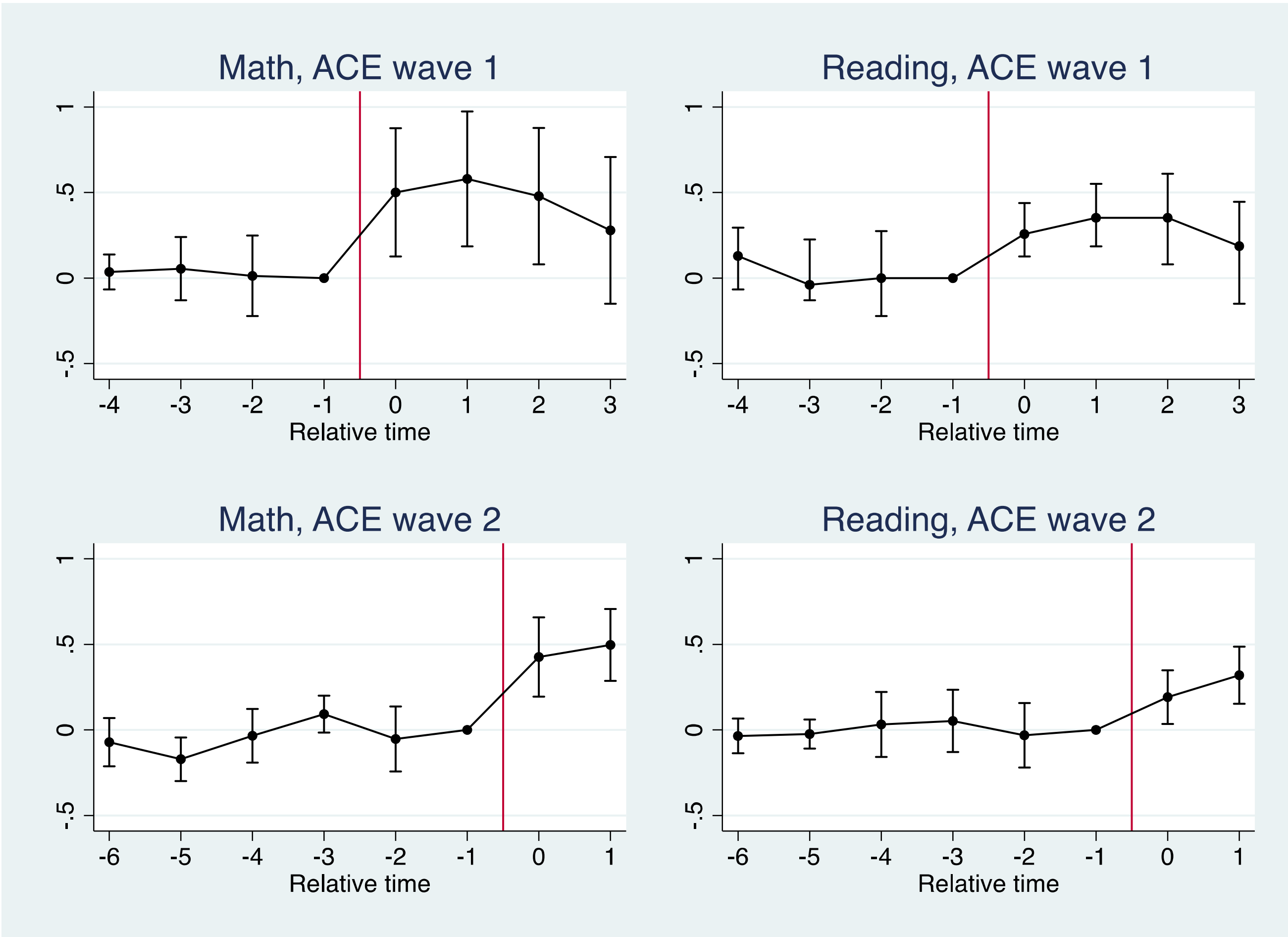
Notes: ACE 2 excludes Onesimo Hernandez Elementary School and J.W. Ray Learning Center that leave the program in 2019. ACE 2 control schools are the lowest 15 percent non-ACE schools in terms of 2016 achievement.

Figure 4. Trends in standardized math and reading scores in ACE 1, ACE 2, and their respective control groups: 2012-2019



Notes: The figure plots standardized math and reading scores for the schools that first received ACE in 2016 (ACE 1) and for schools that first received ACE in 2018 (ACE 2). ACE 1 receives the program for 2016-2018 and it is mostly eliminated for these schools in 2019. The control groups consist of the bottom 15 percent of nontreated schools in terms of achievement in 2014 for ACE 1 and 2016 for ACE 2. ACE 1 excludes Umphrey Lee and ACE 2 excludes Onesimo Hernandez Elementary School and J.W. Ray Learning Center.

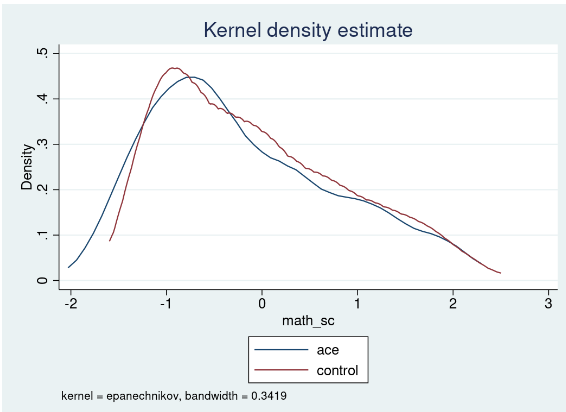
Figure 5. Event-study estimates of the effects of ACE on math and reading achievement, by wave



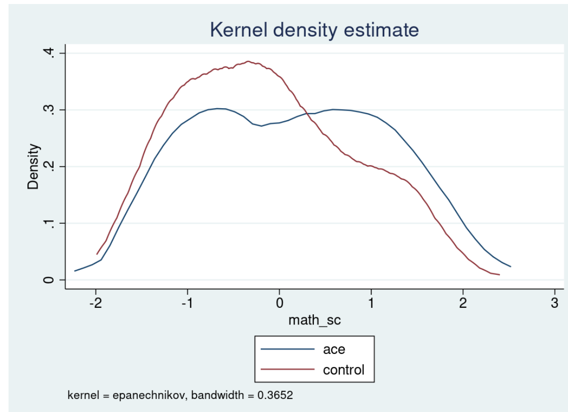
Notes: The figures plot coefficients and 95% confidence intervals for the baseline event-study specifications without controls in which the control groups consist of the bottom 15 percent of nontreated schools in terms of achievement in t-2. ACE 1 excludes Umphrey Lee Elementary School, and ACE 2 excludes Onesimo Hernandez Elementary School and J.W. Ray Learning Center.

Figure 6. Kernel Density Plots of 6th grade math and reading achievement for ACE 1 and control students who attended an ACE or Control school for three successive years for the final pre-treatment cohort and the fully treated cohort

A. Math

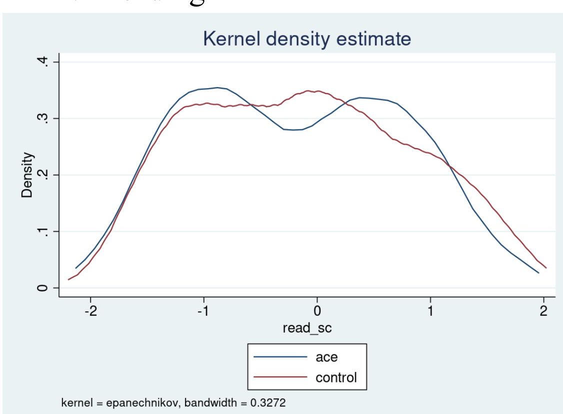


2016

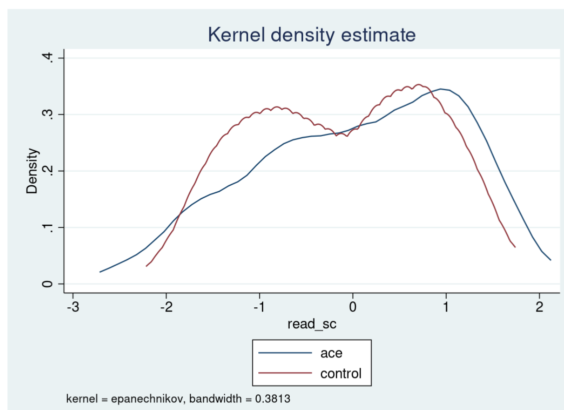


2019

B. Reading



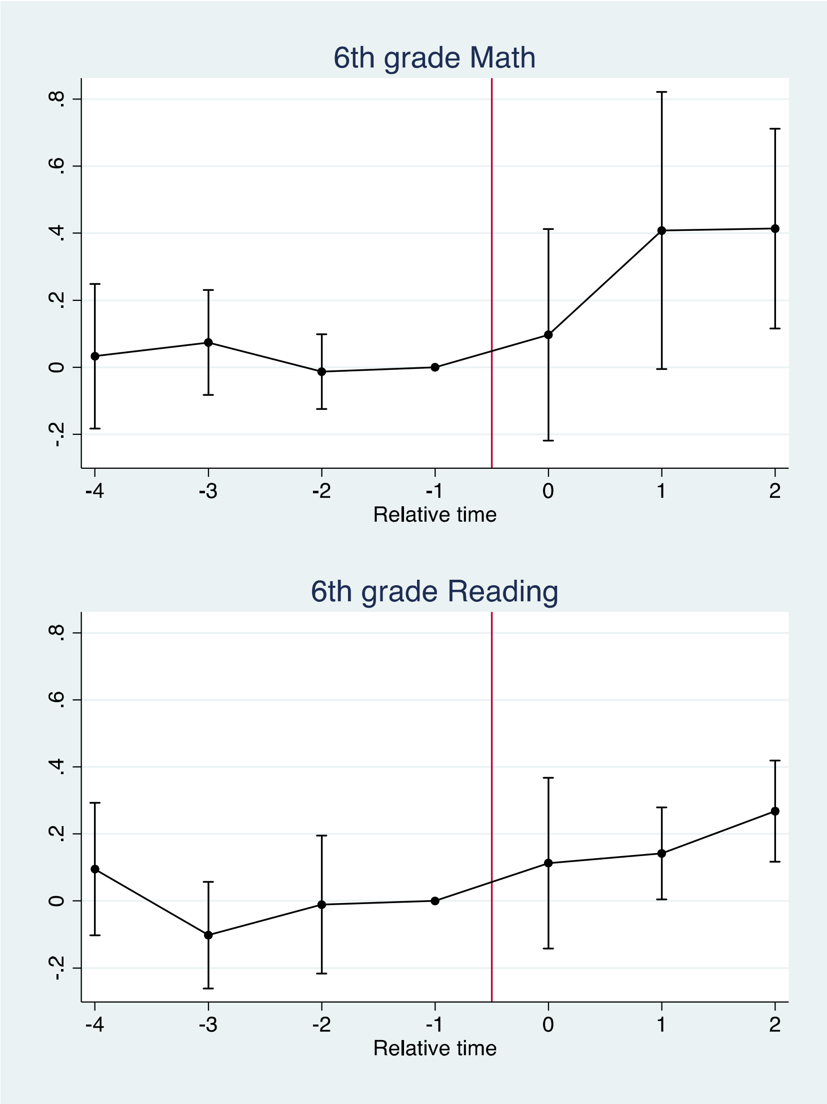
2016



2019

Notes: ACE 1 includes Umphrey Lee Elementary School, and ACE 1 control schools are the lowest 15 percent non-ACE 1schools in terms of 2014 achievement.

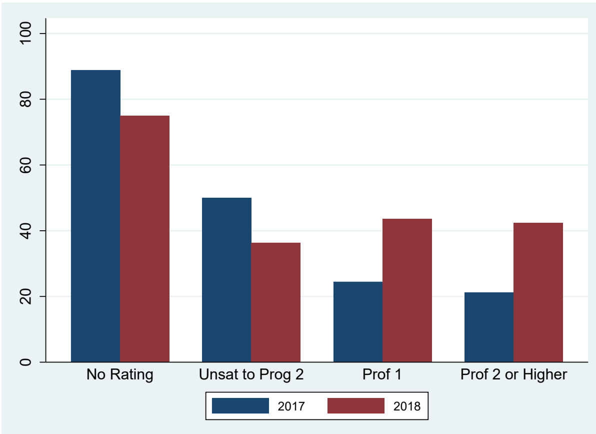
Figure 7: Event-study estimates of the effect of ACE wave 1 on 6th grade math and reading scores



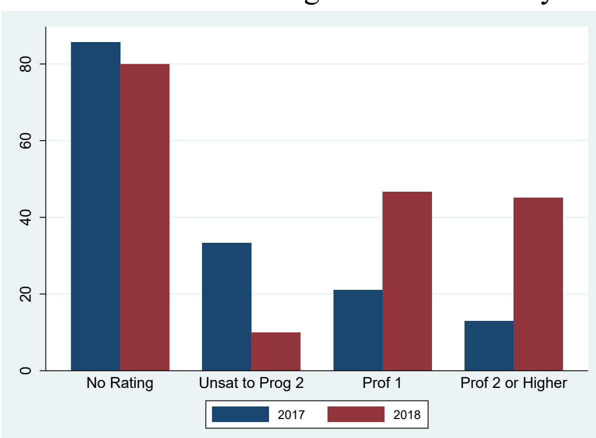
Notes: The figures plot coefficients and 95% confidence intervals for the event-study specifications described in the text. The model is restricted to students with 3 years of stable school attendance from 3rd to 5th grade at either ACE 1 or control schools. The reference year is students who were in 5th grade the year before treatment begins such that they never attend a treated elementary school. T+0 corresponds to students who were in 5th grade when ACE began and therefor have 1 year of exposure. T+2 corresponds to students who were in 3rd grade when ACE began and therefore have 3 years of exposure. ACE 1 excludes Umphrey Lee Elementary School.

Figure 8. Probability an ACE 1 or control teacher exits their campus following the 2017 or 2018 school year, by effectiveness level and inclusion of ACE 1 school that remains treated in 2019

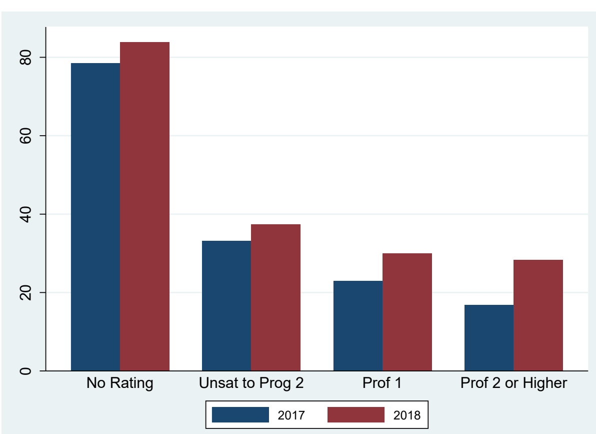
A. ACE 1



B. ACE 1 excluding Pease Elementary School



C. Controls



Notes: ACE 1 schools exclude Umphrey Lee Elementary School.

Table 1. 2015 and 2017 pre-treatment student variable means for ACE 1 and ACE 2 elementary schools and their controls

	2015		2017	
	ACE 1	controls	ACE 2	controls
Student variables				
standardized math score	-0.481	-0.323	-0.555	-0.386
standardized reading score	-0.471	-0.320	-0.507	-0.359
Shares:				
eligible for subsidized lunch	0.952	0.934	0.918	0.945
White	0.025	0.032	0.010	0.012
Black	0.489	0.439	0.618	0.467
Hispanic	0.484	0.511	0.368	0.494
Native American	0.001	0.003	0.003	0.016
Asian	0.001	0.015	0.001	0.011
special education	0.061	0.076	0.088	0.078
LEP	0.377	0.447	0.286	0.369

Notes: ACE 1 schools exclude Umphrey Lee Elementary School and ACE 2 schools exclude Onesimo Hernandez Elementary School and J. W. Ray Learning Center. The control schools for ACE 1 include the lowest 15 percent of elementary schools in terms of average achievement in 2014, and the control schools for ACE 2 include the lowest 15 percent of elementary schools in terms of average achievement in 2016.

Table 2. Event study estimated "effects" of attendance at an ACE elementary school on predicted math and reading achievement, by year and ACE wave (standard errors clustered by school are in parentheses; p-values from permutation tests that cluster by school are in brackets)

	ACE 1		ACE 2	
	math	reading	math	reading
2012	0.0072	0.0124	0.018	0.009
	(0.0149)	(0.0156)	(0.017)	(0.018)
	[0.788]	[0.544]	[0.348]	[0.656]
2013	-0.0178	-0.009	0.010	0.009
	0.0150	(0.009)	(0.013)	(0.015)
	[0.310]	[.588]	[0.532]	[0.586]
2014	-0.0192	-0.011	0.019	0.013
	(0.015)	(0.010)	(0.013)	(0.012)
	[0.144]	[0.366]	[0.196]	[0.368]
2015	0	0	0.010	0.005
			(0.008)	(0.008)
			[0.242]	[0.638]
2016	-0.013	-0.0073	0.002	0.003
	(0.015)	(0.0133)	(0.005)	(0.003)
	[0.324]	[0.564]	[0.760]	[0.658]
2017	0.0187	0.0173	0	0
	(0.015)	(0.008)		
	[0.234]	[0.246]		
2018	0.034	0.038	-0.003	-0.002
	(0.015)	(0.012)	(0.011)	(0.009)
	[0.238]	[0.128]	[0.808	[0.838]
2019	-0.0015	0.0024	0.009	0.004
	(0.016)	(0.016)	(0.005)	(0.006)
	[0.956]	[0.916]	[0.450]	[0.752]
observations	46,933	46,933	50,996	50,996

Notes: ACE 1 schools exclude Umphrey Lee Elementary School and ACE 2 schools exclude Onesimo Hernandez Elementary School and J. W. Ray Learning Center. The control schools for ACE 1 include the lowest 15 percent of elementary schools in terms of average achievement in 2014, and the control schools for ACE 2 include the lowest 15 percent of elementary schools in terms of average achievement in 2016.

Table 3. Event study estimated effects of attendance at an ACE elementary school on math and reading achievement, by year and ACE wave (standard errors clustered by school are in parentheses; p-values from permutation tests that cluster by school are in brackets)

	ACE 1		ACE 2	
	math	reading	math	reading
2012	0.0356	0.128	-0.071	-0.035
	(0.0527)	(0.0854)	(0.072)	(0.052)
	[0.758]	[0.256]	[0.692]	[0.774]
2013	0.0551	-0.0403	-0.171	-0.024
	(0.0948)	(0.135)	(0.065)	(0.043)
	[0.696]	[.782]	[0.382]	[0.844]
2014	0.0122	-0.002	-0.034	0.032
	(0.120)	(0.140)	0.080	(0.097)
	[0.892]	[0.990]	[0.890]	[0.794]
2015	0	0	0.093	0.053
			0.055	(0.093)
			[0.532]	[0.674]
2016	0.500	0.2549	-0.053	-0.031
	(0.190)	(0.0921)	(0.097)	(0.096)
	[0.000]	[0.016]	[0.618]	[0.708]
2017	0.5796	0.3515	0	0
	(0.201)	(0.101)		
	[0.006]	[0.016]		
2018	0.4782	0.3529	0.426	0.192
	(0.203)	(0.131)	(0.118)	(0.080)
	[0.036]	[0.068]	[0.010]	[0.028]
2019	0.2801	0.1863	0.497	0.320
	(0.218)	(0.131)	(0.107)	(0.085)
	[0.356]	[0.476]	[0.010]	[0.028]
observations	45,019	45,014	48,957	48,950

Notes: ACE 1 schools exclude Umphrey Lee Elementary School and ACE 2 schools exclude Onesimo Hernandez Elementary School and J. W. Ray Learning Center. The control schools for ACE 1 include the lowest 15 percent of elementary schools in terms of average achievement in 2014, and the control schools for ACE 2 include the lowest 15 percent of elementary schools in terms of average achievement in 2016.

Table 4. Event study estimated effects of actual and potential years of attendance at an ACE elementary school on math and reading achievement, by ACE wave (standard errors clustered by school are in parentheses; p-values from permutation tests that cluster by school are in brackets)

	years attending an ACE school	1	2	2	3	3
	specification	actual	actual	potential	actual	potential
Math						
	ACE wave 1	0.500	0.626	0.487	0.521	0.385
		(0.190)	(0.244)	(0.227)	(0.244)	(0.190)
		[0.000]	[0.012]	[0.016]	[0.036]	[0.030]
	observations	45,019	34,932	41,397	28,062	38,665
	ACE wave 2	0.426	0.599	0.455	N. A.	N. A.
		(0.118)	(0.088)	(0.098)		
		[0.010]	[0.002]	[0.000]		
	observations	48,957	37,550	44,384		
Reading						
	ACE wave 1	0.257	0.374	0.294	0.426	0.361
		(0.093)	(0.132)	(0.150)	(0.142)	(0.145)
		[.050]	[0.012]	[0.024]	[0.054]	[0.042]
	observations	45,014	34,930	41,387	28,063	38,660
	ACE wave 2	0.192	0.367	0.243	N. A.	N. A.
		(0.080)	(0.097)	(0.092)		
		[0.028]	[0.006]	[0.004]		
		48,950	37,551	44,383		

Notes: The table reports only the treatment effect. ACE 1 schools exclude Umphrey Lee Elementary School and ACE 2 schools exclude Onesimo Hernandez Elementary School and J. W. Ray Learning Center. The control schools for ACE 1 include the lowest 15 percent of elementary schools in terms of average achievement in 2014, and the control schools for ACE 2 include the lowest 15 percent of elementary schools in terms of average achievement in 2016.

Table 5. Event Study Estimated Effects of potential and actual years of attendance at an ACE 1 elementary school on 6th grade math and reading achievement, by years of attendance (standard errors clustered by elementary school in parentheses; permutation test p values in brackets)

years attending an ACE school	1	2	2	3	3
specification	actual	actual	potential	actual	potential
math	0.021	0.297	0.250	0.414	0.299
	(0.110)	(0.180)	(0.134)	(0.152)	(0.085)
	[.980]	[0.070]	[0.092]	[0.108]	[0.144]
observations	10,088	7,865	9,401	6,237	8,830
reading	0.056	0.083	0.034	0.268	0.139
	(0.135)	(0.099)	(0.088)	(0.077)	(0.060)
	[0.834]	[0.478]	[0.756]	[0.172]	[0.372]
observations	10,104	7,882	9,421	6,250	8,844

Notes: The table reports only the treatment effect. ACE 1 schools exclude Umphrey Lee Elementary School. The control schools for ACE 1 include the lowest 15 percent of elementary schools in terms of average achievement in 2014.

Table 6. Sensitivity Analysis of estimated effects of actual and potential years of attendance at an ACE elementary School on math achievement, by Ace wave, control sample composition, ACE composition and inclusion of student characteristics (standard errors clustered by school in parentheses; permutation test p-values in brackets)

	ACE 1					ACE 2		
years attending an ACE school	1	2	2	3	3	1	2	2
specification	actual	actual	potential	actual	potential	actual	actual	potential
1. control group lowest 15%	0.500	0.626	0.487	0.521	0.385	0.426	0.600	0.455
(from Table 4)	(0.190)	(0.244)	(0.227)	(0.244)	(0.190)	(0.118)	(0.099)	(0.099)
	[0.000]	[0.012]	[0.016]	[0.036]	[0.030]	[0.010]	[0.002]	[0.000]
2. control group lowest 15%	0.508	0.61	0.485	0.49	0.37	0.444	0.599	0.480
with student characteristics	(0.184)	(0.242)	(0.217)	(0.246)	(0.183)	(0.117)	(0.099)	(0.092)
	[0.000]	[0.012]	[0.016]	[0.060]	[0.044]	[0.006]	[0.002]	[0.000]
3. control group lowest 10%	0.499	0.617	0.489	0.473	0.358	0.442	0.585	0.441
	(0.194)	(0.250)	(0.231)	(0.257)	(0.198)	(0.121)	(0.100)	(0.106)
	[0.000]	[0.008]	[0.022]	[0.122]	[.138]	[0.014]	[0.002]	[0.004]
4. control group lowest 20%	0.452	0.586	0.450	0.547	0.398	0.402	0.577	0.422
	(0.189)	(0.242)	(0.225)	(0.238)	(0.186)	(0.116)	(0.082)	(0.094)
	[0.004]	[0.004]	[0.022]	[0.014]	[0.028]	[0.012]	[0.004]	[0.000]
5. ACE 2 schools excluded before	0.396	0.552	0.426	0.557	0.404	N. A.	N. A.	N. A.
selection of lowest 15% of	(0.174)	(0.244)	(0.227)	(0.241)	(0.189)			
ACE 1 controls	[0.036]	[0.010]	[0.031]	[0.014]	[0.038]			
6. ACE 1 schools include	0.518	0.625	0.507	0.594	0.436	N. A.	N. A.	N. A.
Umphrey Lee School	(0.140)	(0.186)	(0.171)	(0.189)	(0.146)			
controls group lowest 15%	[0.000]	[0.006]	[0.004]	[0.002]	[0.006]			
7. ACE 2 schools include Onesimo	N. A.	N. A.	N. A.	N. A.	N. A.	0.452	0.608	0.450
Hernandez Elementary School						(0.104)	(0.103)	(0.089)
and J. W. Ray Learning Center						[0.000]	[0.000]	[0.000]

Notes: The control schools for ACE 1 are selected based on average achievement in 2014, and the control schools for ACE 2 are selected based on average achievement in 2016.

Table 7. Sensitivity Analysis of estimated effects of actual and potential years of attendance at an ACE elementary School on reading achievement, by Ace Wave, control sample composition, ACE composition and inclusion of student characteristics (standard errors clustered by school in parentheses; permutation test p-values in brackets)

	ACE 1					ACE 2		
years attending an ACE school	1	2	2	3	3	1	2	2
specification	actual	actual	potential	actual	potential	actual	actual	potential
1. control group lowest 15%	0.257	0.374	0.294	0.426	0.361	0.192	0.367	0.243
(from Table 4)	(0.093)	(0.132)	(0.150)	(0.142)	(0.145)	(0.080)	(0.097)	(0.093)
	[0.050]	[0.012]	[0.024]	[0.054]	[0.042]	[0.028]	[0.006]	[0.004]
2. control group lowest 15%	0.2616	0.373	0.299	0.396	0.341	0.204	0.376	0.275
with student characteristics	(0.084)	(0.134)	(0.141)	(0.144)	(0.136)	(0.077)	(0.103)	(0.090)
	[0.012]	[0.014]	[0.026]	[0.072]	[0.054]	[0.022]	[0.004]	[0.000]
3. control group lowest 10%	0.247	0.366	0.297	0.407	0.358	0.190	0.336	0.210
	(0.098)	(0.139)	(0.153)	(0.156)	(0.153)	(0.082)	(0.102)	(0.096)
	[0.067]	[0.052]	[0.032]	[0.130]	[0.086]	[0.072]	[0.010]	[0.030]
4. control group lowest 20%	0.245	0.391	0.309	0.466	0.392	0.177	0.361	0.226
	(0.089)	(0.126)	(0.147)	(0.132)	(0.160)	(0.079)	(0.092)	(0.089)
	[0.010]	[0.004]	[0.014]	[0.006]	[0.024]	[0.052]	[0.004]	[0.000]
5. ACE 2 schools excluded before	0.254	0.359	0.288	0.435	0.366	N. A.	N. A.	N. A.
selection of lowest 15% of	(0.083)	(0.130)	(0.149)	(0.140)	(0.143)			
ACE 1 controls	[0.046]	[0.012]	[0.046]	[0.036]	[0.044]			
6. ACE 1 schools include	0.288	0.405	0.315	0.503	0.394	N. A.	N. A.	N. A.
Umphrey Lee School	(0.076)	(0.101)	(0.112)	(0.125)	(0.113)			
controls group lowest 15%	[0.002]	[0.002]	[0.002]	[0.008]	[0.012]			
7. ACE 2 schools include Onesimo	N. A.	N. A.	N. A.	N. A.	N. A.	0.209	0.357	0.225
Hernandez Elementary School						(0.064)	(0.103)	(0.080)
and J. W. Ray Learning Center						[0.000]	[0.000]	[0.000]

Notes: The control schools for ACE 1 are selected based on average achievement in 2014, and the control schools for ACE 2 are selected based on average achievement in 2016.

Table 8. Sensitivity Analysis of estimated effects of actual and potential years of attendance at an ACE elementary school on 6th grade math achievement, by control sample composition, ACE composition and inclusion of student characteristics (standard errors clustered by school in parentheses; permutation test p-values in brackets)

years attending an ACE School	1	2	2	3	3
specification	actual	actual	potential	actual	potential
1. control group lowest 15%	0.021	0.297	0.250	0.414	0.299
(from Table 5)	(0.110)	(0.180)	(0.134)	(0.152)	(0.085)
	[.980]	[0.070]	[0.092]	[0.108]	[0.144]
2. control group lowest 15%	0.028	0.268	0.235	0.371	0.284
with student characteristics	(0.103)	(0.179)	(0.127)	(0.151)	(0.083)
	[0.856]	[0.090]	[0.084]	[0.140]	[0.136]
3. control group lowest 10%	-0.017	0.213	0.211	0.392	0.285
	(0.097)	(0.189)	(0.125)	(0.131)	(0.079)
	[0.720]	[0.128]	[0.116]	[0.092]	[0.106]
4. control group lowest 20%	-0.011	0.296	0.263	0.434	0.319
	(0.111)	(0.184)	(0.140)	(0.162)	(0.102)
	[0.784]	[0.152]	[0.122]	[0.120]	[0.106]
5. ACE 2 schools excluded before	-0.008	0.276	0.232	0.396	0.309
selection of lowest 15% of	(0.100)	(0.173)	(0.126)	(0.137)	(0.067)
ACE 1 controls	[0.826]	[0.086]	[0.082]	[0.076]	[0.108]
6. ACE 1 schools include	-0.036	0.268	0.229	0.388	0.285
Umphrey Lee School	(0.112)	(0.184)	(0.133)	(0.154)	(0.085)
controls group lowest 15%	[0.710]	[0.110]	[0.102]	[0.124]	[0.126]

Notes: The control schools for ACE 1 are selected based on average achievement in 2014.

Table 9. Sensitivity Analysis of estimated effects of actual and potential years of attendance at an ACE elementary school on 6th grade reading achievement, by control sample composition, ACE composition and inclusion of student characteristics (standard errors clustered by school in parentheses; permutation test p-values in brackets)

years attending an ACE School	1	2	2	3	3
specification	actual	actual	potential	actual	potential
1. control group lowest 15%	0.056	0.083	0.034	0.268	0.139
(from Table 5)	(0.135)	(0.099)	(0.088)	(0.077)	(0.060)
	[0.834]	[0.478]	[0.756]	[0.172]	[0.372]
2. control group lowest 15%	0.056	0.083	0.034	0.268	0.139
with student characteristics	(0.135)	(0.099)	(0.088)	(0.077)	(0.060)
	[0.834]	[0.478]	[0.756]	[0.172]	[0.372]
3. control group lowest 10%	-0.013	0.066	0.027	0.217	0.115
	(0.131)	(0.089)	(0.076)	(0.089)	(0.061)
	[0.918]	[0.516]	[0.762]	[0.198]	[0.350]
4. control group lowest 20%	0.034	0.089	0.054	0.313	0.161
	(0.134)	(0.105)	(0.093)	(0.088)	(0.071)
	[0.982]	[0.494]	[0.652]	[0.108]	[0.250]
5. ACE 2 schools excluded before	0.020	0.088	0.047	0.276	0.171
selection of lowest 15% of	(0.131)	(0.094)	(0.082)	(0.064)	(0.049)
ACE 1 controls	[0.958]	[0.446]	[0.690]	[0.104]	[0.276]
6. ACE 1 schools include	-0.013	0.102	0.073	0.236	0.128
Umphrey Lee School	(0.145)	(0.092)	(0.076)	(0.068)	(0.062)
controls group lowest 15%	[0.844]	[0.382]	[0.464]	[0.198]	[0.414]

Notes: The control schools for ACE 1 are selected based on average achievement in 2014.

Appendix A: PEI and TEI institutional background

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. PEI and TEI delineate in great detail the requirements of the initiatives, points awarded for each criterium, and educator responsibilities for carrying them out. 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.

We now highlight some main features of TEI and PEI along with 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 capped, and the higher point values had to be based on the value-added or peer group measures. The number of achievement points also 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 whether a principal is an effective 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.[21] 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. Student surveys, student achievement and supervisor evaluations combine to determine the evaluation score and rating, though each of the latter two components may not count for teachers who either do not teach in tested subjects or grades or who work with students below the grade level at which surveys are administered. 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

[21] Morgan (2020) investigates evaluation inflation including the impact of the congruence component.

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. 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.[22] 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.

Though each rating is assigned a salary, there are other considerations that can override this process. First, experience and education determine the salary for teachers new to Dallas ISD; second, teachers with fewer than three years of experience are limited in the maximum rating and compensation they can receive; and third, teachers who taught in Dallas ISD prior to the TEI reform cannot have their nominal pay lowered below its pre-reform level.

[22] Since we focus on mathematics and reading/language arts teachers in tested grades, their evaluations include an achievement component based on classroom achievement.