ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-768

The Effects of Comprehensive Educator Evaluation and Pay Reform on Achievement

Eric Hanushek
Stanford University

Jin Luo
University of Texas at Dallas

Andrew Morgan
University College London

Minh Nguyen
Ball State University

Ben Ost
University of Illinois at Chicago

Steven Rivkin
University of Illinois at Chicago

Ayman Shakeel
University of Illinois at Chicago

A fundamental question for education policy is whether outcomes-based accountability including comprehensive educator evaluations and a closer relationship between effectiveness and compensation improves the quality of instruction and raises achievement. We use synthetic control methods to study the comprehensive teacher and principal evaluation and compensation systems introduced in the Dallas Independent School District (Dallas ISD) in 2013 for principals and 2015 for teachers. Under this far-reaching reform, educator evaluations that are used to support teacher growth and determine salary depend on a combination of supervisor evaluations, student achievement, and student or family survey responses. The reform replaced salary scales based on experience and educational attainment with those based on evaluation scores, a radical departure from decades of rigid salary schedules. The synthetic control estimates reveal positive and significant effects of the reforms on math and reading achievement that increase over time. From 2015 through 2019, the average achievement for the synthetic control district fluctuates narrowly between -0.27 s.d. and -0.3 s.d., while the Dallas ISD average increases steadily from -0.28 s.d. in 2015 to -0.08 s.d. in 2019, the final year of the sample. Though the increase for reading is roughly half as large, it is also highly significant.

VERSION: May 2023

Suggested citation: Hanushek, Eric A., Jin Luo, Andrew J. Morgan, Minh Nguyen, Ben Ost, Steven G. Rivkin, and Ayman Shakeel. (2023). The Effects of Comprehensive Educator Evaluation and Pay Reform on Achievement. (EdWorkingPaper: 23-768). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/bvtc-j223

The Effects of Comprehensive Educator Evaluation and Pay Reform on Achievement

Eric Hanushek*, Jin Luo**, Andrew Morgan***, Minh Nguyen****, Ben Ost*****, Steven Rivkin****** and Ayman Shakeel*******

March 2023

A fundamental question for education policy is whether outcomes-based accountability including comprehensive educator evaluations and a closer relationship between effectiveness and compensation improves the quality of instruction and raises achievement. We use synthetic control methods to study the comprehensive teacher and principal evaluation and compensation systems introduced in the Dallas Independent School District (Dallas ISD) in 2013 for principals and 2015 for teachers. Under this far-reaching reform, educator evaluations that are used to support teacher growth and determine salary depend on a combination of supervisor evaluations, student achievement, and student or family survey responses. The reform replaced salary scales based on experience and educational attainment with those based on evaluation scores, a radical departure from decades of rigid salary schedules. The synthetic control estimates reveal positive and significant effects of the reforms on math and reading achievement that increase over time. From 2015 through 2019, the average achievement for the synthetic control district fluctuates narrowly between -0.27 s.d. and -0.3 s.d., while the Dallas ISD average increases steadily from -0.28 s.d. in 2015 to -0.08 s.d. in 2019, the final year of the sample. Though the increase for reading is roughly half as large, it is also highly significant.

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.

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

1. Introduction

State and federal programs to strengthen school accountability highlighted by the 2003 No Child Left Behind legislation under the Bush Administration substantially expanded the use of student outcomes in the measurement of school performance. Building on these efforts, Race to the Top (RTT) legislation under the Obama administration provided large incentives for states to reform evaluation and compensation systems. This contributed to the institution of new teacher evaluation and accountability policies by the vast majority of states (National Council on Teacher Quality (2017)).

A fundamental question is whether the expansion of accountability and reforms of teacher evaluation and compensation structures raised achievement. Bleiberg et al (2021) documents "the massive effort to introduce new high-stakes teacher evaluation systems," but find that state reforms, even those classified as more rigorous, failed to significantly increase the quality of instruction and achievement. This raises doubts about the potential for changes in personnel practices to elevate the quality of instruction. However, the fact that less than one percent of teachers were rated unsatisfactory following the reforms and the typically small amount of compensation linked with higher performance, both highlighted by Bleiberg et al (2021), suggests the RTT incentives may have been too weak to induce meaningful changes. There is evidence that stronger reforms enacted in some districts may have had a more positive effect in the quality of instruction. For example, regression discontinuity design (RDD) estimates on the Washington DC IMPACT reform show that an elevated threat of dismissal following receipt of a low rating both increases the probability of exit and raises performance in the following year for those who remain in the district (Dee and Wyckoff, 2015). In addition, Adnot et al (2016) find that higher turnover of teachers who received a low IMPACT rating raised

grade-average math and reading achievement in the subsequent year. Yet a personnel system with expanded sanctions and greater salary uncertainty may adversely affect educator supply and potentially dampen the gains from collaboration.[1] Therefore, it is crucial to measure the total effect of a far-reaching reform on student outcomes, and we study the reforms introduced by Dallas ISD to do exactly that.

The comprehensive teacher and principal evaluation and compensation systems introduced in the Dallas Independent School District (Dallas ISD) attempt to address directly the primary impediments to the development and implementation of a successful human resource structure. The Principal Excellence Initiative (PEI), implemented in 2013, and Teacher Excellence Initiative (TEI), implemented in 2015, introduced multiple-measure evaluation systems that align compensation with effectiveness and establish incentives for administrators to engage in rigorous evaluation and to support teacher improvement. Dallas ISD has replaced salary scales based on experience and educational attainment with those based on evaluation scores, a radical departure from decades of rigid salary schedules. The district rates educators on the basis of their contributions to student achievement, supervisor observations and student or family feedback and uses the aggregate evaluation scores to place educators into ratings categories that are the primary determinant of salary. To protect the budget from tendencies toward evaluation inflation and deter the arbitrary treatment of teachers, the systems fix the distributions of teacher and principal ratings, and PEI includes a component that penalizes principals for a lack of alignment between their subjective teacher evaluations and effectiveness

[1] Kraft, Brunner, Dougherty, and Schwegman (2020) find evidence that state reforms adversely affected the supply of teachers, particularly in hard-to-staff schools. This is consistent with the notion that reforms tended not to account adequately for factors outside educator control and that the size of pay increases to high-performing educators did not offset the additional risk and possible other dis-amenities associated with the reforms (Rothstein 2015).

at raising achievement.[2] In addition, the inclusion of school average achievement as a determinant of teacher evaluations is included to recognize the importance of teamwork, and consideration of progress toward reducing achievement gaps in principal evaluations provides incentives to focus on the least advantaged and students of color. Finally, district assessments have been developed to assess outcomes in grades and subjects that lack a state standardized test. The system has been modified in the years since adoption, but the foundational principles remain in place.

This paper first investigates the reform effect on achievement and then considers the contribution of changes in the composition of teachers to any improvement. Although the reforms are conceptually appealing along many dimensions, attribution of any achievement gains to the reforms requires more than simple examination of trends. It is important to allow the treatment effects to evolve over multiple years. Short-term disruptions including extensive educator turnover across the experience distribution accompanied this reform, and these can mask longer-term benefits in the initial treatment years. It is also important to identify a comparison group that would produce a valid counterfactual estimate of achievement trends in the absence of the reforms. State policies or underlying demographic trends could, for example, improve outcomes for all urban or high poverty districts.

We use synthetic control methods to construct counterfactual achievement trends based on a donor pool of large Texas districts with at least 60 percent low-income students.[3] The weight assigned to each comparison school is chosen to minimize the distance between average achievement in Dallas ISD and average achievement in the synthetic control district in the pre-

[2] Morgan (2021) shows substantial evaluation inflation despite these efforts. Nevertheless, it also finds little change over time in the correlation between subjective and objective performance measures.
[3] The share of Texas students eligible for a subsidized lunch increases over time and reaches 60 percent at the end of our sample period.

treatment period prior to 2013. The main estimates further restrict the donor pool to the 20 largest low-income districts, but we examine the sensitivity of the results to the size of the donor pool by expanding it to the 50 largest low-income districts.

The synthetic control estimates reveal positive and significant effects of the reforms on math and reading achievement that increase over time. The positive effects on math achievement emerge in 2016, one year following the implementation of TEI and three years following the implementation of PEI. From 2015 through 2019 (the last year in our data), the average achievement for the synthetic control district fluctuates narrowly between -0.27 s.d. and -0.3 s.d., while the Dallas ISD average increases steadily from -0.28 s.d. in 2015 to -0.1 s.d. in 2018 and -0.08 s.d. in 2019, the final year of the sample. The expansion of the donor pool has little effect on the estimates. Though the increase for reading is roughly half as large, it is also highly significant based on permutation test p-values. The closer relationship between pay and effectiveness would be expected to increase educator effort and strengthen the relationship between educator persistence in the district and effectiveness. Consistent with this, we find that educators who exit the district have substantially lower evaluation scores on average than those who remain despite the absence of explicit removal triggers from the reforms.

The high rate and selective character of teacher turnover suggests an important role for educator composition, as the new entrants would be expected to outperform the leavers. However, we do not have direct measures of the effectiveness of new entrants prior to their arrival in Dallas ISD. We therefore deduce the contribution of fixed differences in teacher effectiveness and those related to experience by comparing overall changes over time in average achievement with estimates of average changes over time within teachers, controlling for experience. The within-teacher changes capture the influences of all factors other than

composition including stronger performance incentives, better school leadership, and enhanced professional development, and the differences between overall and within teacher changes provide estimates of the contributions of teacher composition. This analysis finds that composition accounts for approximately 15 percent of the reform effect on math achievement or roughly 0.03 standard deviations. The remaining channels including the strengthened incentives for teachers and schools account for the majority of the change, but their contributions cannot be disentangled from one another.

2. Dallas ISD Evaluation and Compensation Reforms

Dallas ISD 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 a student achievement component, a performance component based largely on supervisor observations of teaching or work product, and a survey component based on feedback from students (teachers) or families (principals). The current-year composite evaluation score determines the rating, and the two-year average score determines the salary bin with some qualifications. Finally, each delineates in great detail the requirements of the initiatives and educator responsibilities for carrying them out.

The integrated multi-measure evaluation systems and accompanying effectiveness-based compensation structure are designed to support teacher growth, strengthen incentives to improve instruction and leadership practices, and attract strong educators to Dallas ISD. These are the primary channels through which the reforms are expected to raise the quality of instruction and consequently lead to higher test scores and improvements in future educational attainment and

labor-market outcomes. We discuss TEI first and then highlight differences for PEI, drawing on district sources.[4]

2.a. Teacher Excellence Initiative (TEI)

After three years of discussion and development, the Teacher Excellence Initiative was approved by the Dallas ISD Board of Trustees in May 2014. It replaced the Dallas Professional Development and Appraisal System which used years of service and post-graduate schooling as the primary salary determinants; the system had been in place for 22 years. TEI dramatically alters the evaluation and compensation structures by requiring schools to collect far more information about teachers and to use the information for assessment, as the basis of professional development, and to set salary with some exceptions including educators in their first year in the district and some protections against salary decreases.

TEI activities can be categorized in three components - Defining Excellence, Supporting Excellence and Rewarding Excellence - each plays an important role in achieving the district goals. Defining Excellence describes the vision of effective teaching and teaching evaluation, and the principal conveys the school goals to teachers as part of goal setting. Supporting excellence refers to evidence-based professional development efforts based on the information generated by TEI. Finally, rewarding excellence refers to the connection between evaluation score and salary level.

---

[4] Sources for the discussion of TEI include TEI Presentation (2015); TEI Rulebook (2015). "Rules and Procedures for Calculating TEI Evaluation Scores and Effectiveness Lev; TEI SLO Rubric (2014); TEI Student Achievement Templates (2015); TEI Teacher Performance Rubric (2014); Weerasinghe, D. (2008). How to compute school and classroom effectiveness indices: The value-added model implemented in Dallas Independent School District (retrieved at 4/20/2015). Sources for the discussion of PEI include Final 2014-2015 DISD Principal Handbook Sept; DISD 2014-2015 Salary Handbook; Principal Professional Development-Dec 2012; Principal Evaluation Rubric-General-Dec 2012; Principal Evaluation-Concept Paper-17 Jan 2013; Professional Development Hours – 18 Mar 2013; Miles M. (2013) Superintendent's Principal Evaluation System Report to the Board and Community. http://www.dallasisd.org/site/default.aspx?PageType=3&DomainID=7954&ModuleInstanceID=24529&ViewID=047E6BE3-6D87-4130-8424-D8E4E9ED6C2A&RenderLoc=0&FlexDataID=22163&PageID=20637

2.a.1. Defining Excellence

Performance, achievement and perception comprise the three components of the evaluation system. Table 1 lists the domains and indicators within each domain that comprise the teacher performance rubric, and teacher receive scores for their performance on each. Every teacher is assigned a primary evaluator who is typically the principal or assistant principal. The evaluator monitors and collects evidence to assess performance mainly through spot, extended and informal observation. TEI specifies ten, 10- to 15-minute spot observations and one 45-minute extended observation per year. The observations focus on Domains 2 and 3, instructional practice and classroom structure. The supervisor is required to provide written feedback following all observations and conference with the teacher following the extended observation. Artifacts and informal observations also contribute to the performance score, as these constitute the evidence of performance on the first and fourth domains.

Student perception is based on a survey conducted in the second week of April. Most students in grades 3-12 complete two surveys, one online and one on paper. Results from the survey are summarized by a single score for each teacher with at least a minimum number of responses; student surveys do not contribute to the evaluation score of some teachers including those in grade 2 or below. 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.

Both school average achievement and classroom achievement contribute to the achievement component, except for teachers whose role is not associated with a TEI assessment. All school-level achievement measures are based on the state standardized test results. Teacher-level measures consists of Student Learning Objective (SLO) and Standardized Teacher-level Student Achievement Measures. SLO is a measure of student improvement during the year based

on assessments that are not standardized tests; SLO contributes to the evaluation scores of all teachers, while classroom achievement contributes to the evaluation scores of teachers whose students take a standardized test. The district computes multiple measures of school and classroom achievement, and the highest metric for a teacher is used to determine their number of achievement points. Initially the alternatives included status (percentage of tests with scores that met a specified standard); value added; and achievement score relative to the scores of a designated peer group of schools based on prior achievement. Subsequently, the district eliminated the status alternative. The district uses target distributions to assign points for the school and teacher achievement components based on the standardized tests.

The evaluation score equals a weighted sum of points earned on the three components, where the weights depend on the role and grade level. Table 2 describes the four categories of teachers and differences among the weights for the three components. Category is determined primarily by the availability of student survey responses and results of a state or district assessment.

Teachers are divided into ratings categories based on scores and whether an application for recognition as a distinguished teacher is approved, a requirement for a rating of proficient II or higher. Table 3 lists the nine evaluation categories.

2.a.2 Supporting Excellence

Evidence including Taylor and Tyler (2012) highlight the value of teacher observations and feedback for professional growth, and the reforms emphasize the importance of teacher feedback based on observations and outcomes and the principal's role as an instructional leader. Each of the three components of the evaluation system provides information used in teacher support and professional development. In addition to the written feedback and conferences

following observations, achievement data are collected and analyzed to help improve instruction. An online resource bank of videos and modules was developed to support school leaders and instructional coaches in generating a clear and common vision of the TEI program in the system and foster self-learning among teachers**.**

2.a.3 Rewarding Excellence

Except for a teacher in her first or second year in Dallas ISD, salary is based on the average of evaluation points earned in the most recent two years; for teachers in their second year, it is based on evaluation points in the previous year only. The average score divides teachers into the nine effectiveness levels listed in Table 3, conditional on certain constraints: a teacher cannot move up or down more than one effectiveness level per year; completion of three years of service as a classroom teacher is a necessary condition to be considered for the Proficient I level; the Proficient II level and above requires teachers to go through the Distinguished Teacher Review (DTR) process, and to be at Exemplary II, teachers need to have at least one year qualifying as an Exemplary teacher; And Master level has additional requirements. To maintain budget stability and deter evaluation inflation, the category boundaries are determined by a target distribution (see Figure 1).

The system also includes safeguards to protect against downside risk: 1) It takes three consecutive years in a lower ratings category for teacher salary to go down by one level; 2) a salary will not fall below the teacher's salary in 2014-15 for those employed in that year; 3) a teacher starting after 2014-15 will not receive a salary lower than their entry-level salary; and 4) the compensation scale will be adjusted at least once per three years to keep salary levels competitive with other districts.

2.b. Principal Excellence Initiative (PEI)

PEI went into effect for the 2012-2013 academic year and is quite similar to TEI. The evaluation includes performance, achievement and survey components, where the survey component contains information obtained from families rather than students. The district devotes substantial resources to build the skills and capacity of principals who went through 135 hours of professional development in the 2011-2012 school year and 175 hours in 2012-2013. As is the case for teachers, principal compensation is determined by the effectiveness level except for those in their first year.

PEI places substantial weight on effectiveness as an instructional leader. Table 4 lists the metrics used in principal evaluation. 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 subjective evaluations and to deter arbitrary judgements of teachers based on factors other than the quality of teaching. Unlike the case for TEI, attendance and enrollment also contribute to the performance score for principals.

The achievement component also differs from that used in TEI, particularly with respect to the tests included and concerns about equity. Over ten percent of the achievement score 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.

Finally, salaries differ by schooling level, conditional on evaluation rating. High school principals earn more than middle school principals who earn more than elementary school principals, on average.

3. Administrative and program data

We use both Texas state administrative data housed at the University of Texas at Dallas Education Research Center (ERC) and administrative and program data provided by Dallas ISD. The Public Education Information Management System (PEIMS), TEA's statewide educational database, reports key demographic data including race, ethnicity, and gender for students and school personnel as well as program characteristics including subsidized or free lunch eligibility. PEIMS also contains detailed annual information on teacher and administrator role, experience, salary, education, class size, grade, population served, and subject taught. Beginning in 1993, the Texas Assessment of Academic Skills (TAAS) was administered each spring to eligible students enrolled in grades three through eight.[5] In 2003 the state substituted the TAKS in place of the TAAS, and in 2012 STAAR replaced the TAKS. We focus on the years 2005 to 2018, (year refers to spring of the academic year), which covers parts of the TAKS and STAAR test regimes. We transform all test results into standardized scores with a mean of zero and variance equal to one for each subject, grade, and year, meaning that our achievement measures describe students by their relative position in the overall state performance distributions. Because TAKS and STAAR differ, it is important to account for changes associated with the test-regime change. The

[5] Many special education and limited English proficient students are exempted from the tests. In each year roughly 15 percent of students do not take the tests, either because of an exemption or because of repeated absences on testing days.

synthetic control analysis minimizes achievement differences in a pre-period that spans the two test regimes.

The longitudinal data contain unique student and educator identifiers that enable us to follow students and educators across districts and schools as long as they remain in a Texas public school. These linkages permit the estimation of value added, and they also enable the description of educator movements in and out of schools and districts including Dallas ISD. We merge educator and student data by campus, grade, and year for the entire period and additionally by teacher, grade and year beginning in 2013.

The Dallas ISD administrative data include demographic and program information contained in the state data system, achievement data, and the disaggregated TEI and PEI components used to determine evaluation and effectiveness ratings and compensation. These data also contain identifiers that enable us to link the TEI and PEI information with student and staff longitudinal data.

4. Aggregate Reform Effects

This section describes the empirical approach and presents the results of the synthetic control analysis of the Dallas ISD reform effects on achievement. The lack of a natural comparison group led us to create a synthetic control district to serve as the counterfactual for Dallas ISD. This control district is created from elementary and middle schools in large, high-poverty districts. In the main specification the donor pool includes all schools from the largest 20 high-poverty districts (other than Dallas ISD), where high-poverty districts have at least 60 percent of the students qualify for a subsidized or free lunch. Schools in the synthetic control district are selected from the donor pool and weighted to minimize the pre-period average

achievement gaps between the synthetic control district and Dallas ISD. The selection of schools rather than districts as the focal unit recognizes the substantial variation in school quality within districts and dampens the impact of the reform efforts or challenges of other districts. We subsequently investigate the robustness of the estimates by examining the sensitivity of the estimates to expanding the donor pool to include the largest 50 high-poverty districts.

4.a. Synthetic Control Model

We estimate the effect of the Dallas reforms using the synthetic control method (SCM) developed by Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010). Conceptually, rather than comparing Dallas schools to a specific set of control schools, this approach forms a synthetic control group, which is a weighted average of potential control schools throughout the state. The weights are chosen to minimize the pre-treatment difference in outcomes between Dallas and the synthetic control.

More formally, let Y_{it}^{T} be the potential outcome at school *i* when the policy is in effect and let Y_{it}^{C} be the potential outcome at school *i* when the policy is not in effect. For each year in the post-period, we know the realized outcomes at Dallas schools and need to estimate Y_{it}^{C}. The synthetic control method estimates this counterfactual by taking a weighted average of control school outcomes in each year, where these weights are constrained to be constant over time. Specifically, defining an indicator D_i that is 1 for all Dallas schools and zero otherwise, the counterfactual outcome for year t is

$$\sum_{D_i=0} w_i^* Y_{it}^{C}$$

where the weights are chosen to minimize a specific objective function. Because we match on all pre-treatment outcomes, the nested optimization component of the synthetic control approach greatly simplifies and all pre-treatment periods receive equal weight (Kaul, Klößner, Pfeifer, and

Schieler, 2022). As such, in our case the synthetic control approach simply chooses weights, w_i^*, to minimize the sum-of-squared differences between Dallas and synthetic control schools in the pre-period (defined as $t < 0$) shown in the equation below.[6]

$$\sum_{t<0} \left(Y_{it}^{D=1} - w_i^* Y_{it}^{D=0}\right)^2$$

Following the approach in Abadie, Diamond, and Hainmueller (2010), we conduct inference using a permutation test that compares the estimated effect for Dallas to a distribution of placebo estimated effects. Because there are many treated schools, the distribution of placebo estimates is based on averages where the number of placebo units used in each average is the number of Dallas schools (Cavallo, Galiani, Noy, and Pantano (2013)). With many Dallas schools and many potential controls there are a large number of possible averages, and we sample from this distribution 1,000,000 times with replacement following the approach described in Galiani and Quistorff (2017).

4.b. Results

We begin with synthetic control estimates based on a donor pool of the largest 20 high-poverty districts before illustrating the sensitivity of the estimates to expansions of the donor pool.

4.b.1 Main Estimates

Figures 2 and 3 present plots of math and reading achievement in Dallas and the synthetic control; Table 5 presents the exact estimated effects and p-values. Figure 2 shows both large treatment effects and the very close match between Dallas ISD and the synthetic control district pre-period achievement trends including the year-to-year fluctuations. In 2013, when PEI is first

[6] This is implemented using the user-written *synth_runner* routine for Stata, described in Galiani and Quistorff (2017).

instituted, we see no evidence of an instantaneous effect on achievement. This is consistent with expectations since principals have limited ability to generate instantaneous improvements in test score outcomes. Similarly, there is no evidence of an effect for 2014, the second year of PEI. In 2015, the first year of TEI, we also see no evidence of improved outcomes in Dallas relative to the synthetic control. Although it is possible for TEI to have had immediate effects, a time lag in within-teacher improvement and leftward shift in the experience distribution following the arrival of so many new teachers (described below) would have been expected to dampen temporarily the gains from the reforms. In 2016, outcomes in Dallas and the synthetic control diverge, and the positive gap between Dallas ISD and the synthetic control district grows in the following years. By 2019 (the last year in our data), the gap exceeds 0.2 standard deviations. The fact that it is an achievement increase in Dallas ISD and not an achievement decline in the synthetic control district that drives the divergence supports the belief that the reforms succeeded in elevating the quality of instruction and achievement.

Column 1 of Table 5 shows the exact estimated effects in the post-policy period, along with p-values based on the permutation methods described above. It shows that the reform effect is significant at the 5 percent level in 2016 and at the 1 percent level in the subsequent years. None of the estimates prior to 2016 are significant at any conventional level.

Figure 3 shows substantial and statistically significant reform effects on reading achievement that are roughly half as large as those for math in 2019. Column 2 of Table 5 shows the exact estimates and p-values. The pre-period achievement trend for the synthetic control tracks that for Dallas ISD, though the yearly differences tend to be larger in reading than in math. In addition, the drop in average Dallas ISD reading achievement in 2015, the first year of TEI,

mirrors that found for math, though for reading there is no corresponding drop in the synthetic control district.

4.b.2 Sensitivity analysis

Figures 4 and 5 present synthetic the control plots for an expanded donor pool that includes schools from the largest 50 low-income districts. Columns 3 and 4 of Table 5 show exact estimates and p-values. Math achievement in Dallas ISD continues to track the synthetic control district closely until diverging in 2016. By 2019 the gap approaches 0.2 standard deviations, only slightly smaller than the differential observed in the main specification. The differences for 2017 to 2019 are significant at the 1 percent level.

As is the case for math, Figure 5 shows that the expansion of the donor pool introduces only minor changes to the treatment effect estimates. Although they are slightly smaller in 2018 and 2019 with the expanded donor pool, they remain significant at the 1 percent level. All in all, the insensitivity of the estimates to a substantial expansion of the donor pool provides additional support to the finding that the reforms substantially increased achievement in Dallas ISD.

5. Contributions of educator selection

The comprehensive set of evaluation and compensation reforms involves simultaneous changes in the strength of incentives, information available for mentoring and professional development, and myriad aspects of school operations and educator composition, complicating efforts to disentangle the contributions of each. If the much closer alignment between effectiveness and salary alters the composition of entrants to and exits from Dallas ISD, educator composition could emerge as an important channel through which the reforms raise district quality. A first order issue, therefore, is understanding the impact of the reforms on educator

selection. There has been extensive educator turnover, and we can use the evaluation information and achievement data to describe teacher and principal selection out of Dallas ISD. We focus on selection out rather than selection into Dallas ISD due to the absence of comparable measures of effectiveness for most entrants into Dallas ISD. No other Texas district uses a similar evaluation system, and estimates of teacher value added are available only for the small fraction of entrants who previously taught in a tested grade in another district. Even for these teachers, estimates of value added in their previous schools would conflate teacher and district effects, just as would be the case if we were to measure effectiveness for entrants based on value added following their arrival to Dallas ISD. In addition, the fixed distributions of ratings designed to mitigate evaluation inflation and limit budget growth mean that the ratings distributions do not capture aggregate improvements or declines in educator effectiveness over time.

Following the descriptions of teacher and principal selection out of Dallas ISD we use the student-teacher matches to identify the contribution of changes in teacher composition to the overall increase in math achievement. As we discuss below, the presence of both stayers and leavers enables the separation of the contributions of teacher composition from those of all other channels through which PEI and TEI affected learning and achievement.

5.*a. Teacher selection*

The strengthened performance incentives for administrators and teachers would be expected to increase teacher turnover and result in a less experienced teaching force. Figure 6 illustrates that the share of teachers with no prior experience rose sharply following 2012, quadrupling from 3 to 13 percent by 2015. These annual increases led to a substantial rise in the share of novice teachers with 0-2 years of prior experience: it increased from 12 percent in 2012, to 16 percent in 2013, 21 percent in 2014, 30 percent in 2015 and 32 percent in 2016 prior to

declining modestly until 2019. The evidence that the largest increases in teacher effectiveness occur in the first few years indicates that the influx of teachers with little or no prior experience almost certainly offset a portion of any positive effects emanating from other channels including educator composition through which the reforms affected the quality of instruction and achievement.[7]

The implications of the sharp increase in turnover depends on whether exiting teachers are above or below average. Though we lack data from before the reform to assess how selection patterns change, Figure 7 describes teacher evaluation scores by annual transition status. There is pronounced negative selection out of the district as the average evaluation scores of teachers who remain in Dallas ISD exceed those who leave following the school year by more than 0.5 standard deviations. The lower two panels show that this strong negative selection holds for both the performance and achievement components. A low TEI rating does not trigger dismissal, but Luo (2022) shows that a lower salary increases the probability of leaving Dallas ISD, suggesting that the stronger connection between effectiveness and salary contributed to the positive selection of stayers.

Teacher value added provides an alternative measure of effectiveness that is available for teachers in tested subjects and grades starting in 2013, the first year for which the Texas data match students to a teacher. We follow Parsons, Koedel, and Tan (2019) and use a two-step estimation approach. First, we regress math achievement on a cubic polynomial in prior year achievement, student characteristics, teacher experience indicators and grade fixed effects separately for each year for all Texas teachers of math. Second, we compute year-specific estimates of value added equal to the mean residual for each teacher. We focus on mathematics

[7] See for example, Papay and Kraft (2015).

because schools often expose students to multiple teachers of reading or language arts in the same year. Importantly, the value-added sample includes only teachers of record in tested grades and subjects, and the non-persist categories includes teachers who switch to a non-value added teaching position or other role in the district.

Table 6 reports mean teacher value added by year and transition status and shows that those who return to teach in a tested grade in Dallas ISD substantially outperform those who either remain in the district in a different position or those who leave the district. This pattern predates the implementation of TEI. Importantly, the extent to which this positive selection of stayers translates into higher quality instruction depends crucially on the effectiveness of entrants. However, the absence of a measure of effectiveness for entrants prior to their entry into Dallas ISD precludes the direct estimation of the change in teacher effectiveness, leading us to undertake an indirect approach in Section 5.c below.

5.b. Principal Selection

The high rate of principal turnover following the adoption of PEI provides opportunities and risks. Figure 8 plots the survival rate of 2013 through 2019 and shows that roughly two-thirds of the principals in 2013 were no longer Dallas ISD principals in 2019. The impact of such turnover on leadership quality depends crucially on the effectiveness of new principals relative to those they replaced. Again, the fixed distribution of evaluation scores precludes the direct measurement of changes in principal effectiveness, so we turn our attention to the degree of negative selection out of leadership positions in Dallas ISD following the adoption of PEI.

Figure 9 shows the overall evaluation score and the performance and achievement components for principals in 2013, the first year of PEI (top panel), and for those hired between 2014 and 2018, by transition status in 2019: remain principal at the same schooling level;

promoted to a principal position at a higher schooling level or a central administrative position; or left the district, resumed teaching or took a lower administrative position.[8] Both panels illustrate strong negative selection of principal exits and positive selection of promotions. The average difference in evaluation scores between exits and stayers equals approximately 0.2 standard deviations for principals in 2013 and 0.4 for those hired between 2014 and 2018, and the gap between stayers and those promoted approaches 0.2 standard deviations for 2013 principals and 0.3 standard deviations for more recent principal hires. Because of the fixed ratings evaluation-score and ratings distributions and tendency for less experienced principals to have lower scores, those hired after 2013 have lower scores on average than the 2013 principals with much higher average experience. As is the case for teachers, Shakeel (2022) provides evidence that a lower salary raises the probability a principal leaves their position. This supports the belief that strengthening the connection between pay and performance can lead to beneficial improvements in educator composition in addition to any benefits due to greater effort.

5.c. Estimation of the contribution of teacher composition to the math achievement increase

The bundling of many components precludes the direct estimation of the contributions of strengthened incentives, enhanced professional development, better school leadership and other channels to the overall treatment effects. However, we can separate the contribution of teacher composition from those of the other channels by comparing estimates of the achievement changes over time from regressions of math achievement on a set of year dummies with estimates of within-teacher achievement changes over time from a regression that adds teacher fixed effects and a full set of experience dummies for years 0 to 9 and ten plus. As mentioned

[8] The evaluation scores for principals who enter their positions between 2014 and 2018 are based on the first year as principal in the school.

above, we focus on math achievement because of the more extensive contributions of educators other than the classroom teacher of record to reading and language arts instruction.[9]

Equation 1 models achievement for student i in year t with teacher j as a function of a set of year dummy variables (D), a set of experience dummies exp, a teacher fixed effect (η_j) and a random error:

$$A_{ijt} = \alpha + \sum_{t=2014}^{2019} \delta_t D_t + \sum_{x=1}^{10+} \lambda_x exp_x + \eta_j + \varepsilon_{ijt} \quad (2)$$

In the absence of teacher fixed effects and experience controls, the teacher fixed effect, η_j, and the experience effects become part of the error, and the coefficients on the year dummies, $\widehat{\delta_t^{no\ fe}}$ capture the influences of all factors including teacher composition that contribute to the difference between achievement in year t and achievement in 2013, the omitted baseline year. The inclusion of teacher fixed effects and the experience dummies shuts the teacher composition channel, and the estimate $\widehat{\delta_t^{fe}}$ captures the influences of the other factors only. Therefore, the difference between $\widehat{\delta_t^{no\ fe}}$ and $\widehat{\delta_t^{fe}}$ provides estimates for the contribution of fixed and experience related differences in teacher composition between 2013 and year t to the achievement change over that period. Importantly, student-teacher matches become available in 2013, the first year of the intervention. Therefore, the estimates capture the contribution of teacher composition to achievement growth following the first year of the program and ignore any contributions to achievement in the first year.

[9] The small number of students without teacher matches are dropped from this analysis.

The key idea underlying this exercise is that the year-dummy coefficients in the models with teacher fixed effects and experience controls provide valid estimates of the contributions of all channels other than changes in teacher composition to the achievement change. Note that the inclusion of teacher fixed effects means that only within teacher changes in average achievement across years contribute to the identification of the year-dummy coefficients.

If all of the improvement in Dallas ISD schools comes from replacing worse teachers with better teachers and changes in the experience distribution, then we would expect to find small and insignificant year-dummy coefficients for the teacher fixed effect specifications. In the diametrically opposite case, if teacher composition accounts for none of the reform effects, we would expect the year-dummy coefficients to be insensitive to the inclusion of teacher fixed effects and experience. If, however, both teacher composition and other factors contribute to the overall treatment effects, the difference between the interaction term coefficients with and without the teacher fixed effects and experience controls will provide an estimate of the contribution of teacher composition.

Table 7 reports the set of year dummy coefficients (2013 is the baseline year) for regressions of achievement on year dummies with no teacher composition controls (Column 1) and both teacher fixed effects and experience controls (Column 2). The 2019 dummy variable coefficient of 0.20 replicates the findings of the synthetic control analysis for math achievement shown in Figures 2 and 4.[10] Column 3 shows that the addition of both teacher fixed effects and experience controls reduces the 2019 dummy variable coefficient from 0.20 to 0.17 standard deviations, a 0.03 standard deviation decline. This suggests that teacher composition accounts for roughly 15 percent of the achievement gain following the implementation of the reforms. Given the

[10] Note that a small fraction of students are not matched with teachers, so the sample and changes in achievement differ slightly from the main analysis.

relatively larger shares in the lower experience levels, the results suggest there would be further achievement increases as the experience distribution continues to shift to the right following the substantial turnover at the start of the reforms. In other words, improvements in the fixed component of teacher effectiveness are likely to exceed 0.03 standard deviations, a meaningful shift given the evidence that a one standard deviation difference in the math teacher effectiveness distribution equals approximately 0.12 standard deviations in Texas (Rivkin et al , 2005).

Importantly, teacher composition accounts for only one of the channels through which the reforms could have increased the quality of instruction, but we are not able to identify the contributions of increases in effort in response to the strengthened incentives, peer teacher effects, or improvements in school leadership. Their contributions and those of other factors including improvements in academic support and school climate account for the majority of the math achievement gain but cannot be separately identified.

6. Conclusions

The comprehensive personnel reforms introduced in Dallas ISD in 2013 including the virtual elimination of the dependence of salary on experience and post-graduate degrees radically altered systems of evaluation and pay that were representative of those used throughout the United States. System details reflect careful consideration of the potential for unintended consequences including evaluation inflation, the arbitrary treatment of teachers, and strategic responses including teaching to the test. Aligning the relationship between educator effectiveness and pay dramatically strengthened performance incentives, while the development of a multiple-measure evaluation system based on student outcomes, supervisor observations and student or family feedback recognized the pitfalls of a singular reliance on achievement or subjective

evaluations by supervisors. Importantly, the focus on value added and achievement relative to comparable students rather than pass rate or achievement level in absolute terms made clear the effort to account for factors including family circumstances outside of educator control.

The synthetic control analysis shows that the reforms succeeded in raising math and to a lesser extent reading achievement following an initial period of very high teacher turnover. Effect sizes exceeding 0.2 standard deviations for math and 0.1 standard deviations for reading are large, particularly in light of the minimal costs in comparison to interventions such as a large reduction in class size.

The teacher fixed effects analysis further shows that changes in teacher composition accounted for roughly 15 percent of the math achievement increase. This means that other channels including instructional improvements driven by the strengthened incentives, enhanced support for teachers based on the much richer information produced by TEI, and more effective school leadership accounted for the majority of the improvement in Dallas ISD.

Finally, achievement is only one of the metrics over which teachers are evaluated, and the use of multiple measures is designed to encourage the kind of improvement in practice that leads to the most-positive impacts on short and longer-term outcomes. Because the TEI evaluation score and ratings distributions are structured to generate fixed distributions that do not change over time, the evaluation score components do not capture any improvements in supervisor perceptions of teacher practice and student survey responses. However, the findings in Shakeel (2022) on the relationship between the performance, achievement and student survey components on the one hand and contemporaneous and subsequent-year test scores on the other provide additional evidence that the reform generates instructional improvements that raise lasting cognitive skills. First, estimates of teacher value added to current and subsequent-year

achievement find a strong positive relationship between value added to contemporaneous achievement, the high stakes outcome, and value added to subsequent year test scores that are not high stakes to the teacher. Second, Shakeel (2022) finds that student survey responses, performance scores based on supervisor observations, and value added to contemporaneous achievement are all systematically related to subsequent year achievement. This evidence suggests that the structures of TEI and PEI contribute to gains in achievement that persist into the future.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American Statistical Association* 105, no. 490: 493-505.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The economic costs of conflict: A case study of the Basque Country." *American Economic Review* 93, no. 1: 113-132.

Bleiberg, Joshua, Eric Brunner, Erica Harbatkin, Matthew A. Kraft, and Matthew Springer. 2021. "The Effect of Teacher Evaluation on Achievement and Attainment: Evidence from Statewide Reforms." EdWorkingPapers 21-496. Brown University: Annenberg (December).

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano. 2013. "Catastrophic Natural Disasters and Economic Growth." *The Review of Economics and Statistics* 95, no. 5 (December): 1549-1561.

Dee, Thomas S., and James Wyckoff. 2015. "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." *Journal of Policy Analysis and Management* 34, no. 2: 267-297.

Dee, Thomas S., and James Wyckoff. 2017. "A Lasting Impact: High-stakes teacher evaluations drive student success in Washington, D.C." *Education Next* 17, no. 4 (Fall): 58-66.

Galiani, Sebastian, and Brian Quistorff. 2017. "The Synth_Runner Package: Utilities to Automate Synthetic Control Estimation Using Synth." *The Stata Journal* 17, no. 4: 834-849.

Hanushek, Eric A., and Steven G. Rivkin. 2012. "The distribution of teacher quality and implications for policy." *Annual Review of Economics* 4: 131-157.

Kaul, Ashok, Stefan Klößner, Gregor Pfeifer, and Manuel Schieler. 2022. "Standard Synthetic Control Methods: The Case of Using All Preintervention Outcomes Together With Covariates." *Journal of Business & Economic Statistics* 40, no. 3 (July): 1362-1376.

Kraft, Matthew A., Eric J. Brunner, Shaun M. Dougherty, and David J. Schwegman. 2020. "Teacher accountability reforms and the supply and quality of new teachers." *Journal of Public Economics* 188(August): 104212.Morgan, Andrew. 2021. "Understanding Incentives in Subjective Evaluations: Evidence from Educators." unpublished manuscript.

Luo, Jin. 2022. "Teachers' Responsiveness to Performance-based Pay: Evidence from a Large Urban School District in Texas." unpublished manuscript.

National Council on Teacher Quality. 2017. *State teacher policy yearbook, 2017*. Washington: National Council on Teacher Quality.

Papay, J.P. and Kraft, M.A., 2015. Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*, pp.105-119.

Parsons, Eric, Cory Koedel, and Li Tan. 2019. "Accounting for Student Disadvantage in Value-Added Models." *Journal of Educational and Behavioral Statistics* 44, no. 2 (April): 144-179.

Rothstein, J. (2015). Teacher quality policy when supply matters. *American Economic Review*, *105*(1), 100-130.

Shakeel, Ayman. 2022. "High Stakes Evaluation Ratings and Principal Composition Changes." unpublished manuscript.
Shakeel, Ayman. 2022. "High Stakes Objective and Subjective Teacher Teacher Evaluation Measures and Student Skill Development." unpublished manuscript.
Taylor, Eric S., and John H. Tyler. 2012. "The Effect of Evaluation on Teacher Performance." *American Economic Review* 102, no. 7 (Decenber): 3628-51.

Table 1: Teacher performance rubric.

Domain	Indicator of teacher practice	Evidence used	Max. points
Domain 1: Planning and Preparation	1.1. Demonstrate knowledge of content, concepts, and skills	Artifacts and informal observations	15
	1.2. Demonstrates knowledge of students		
	1.3. Plans or selects aligned formative and summative assessments		
	1.4. Integrates monitoring of student data into instruction		
	1.5. Develops standards-based unit and lesson plans		
Domain 2: Instructional Practice	2.1. Establishes clear, aligned standards-based lesson objective(s) (3x)	Spot, extended and informal observations	48
	2.2. Measures student mastery through a demonstration of learning (DOL) (spot) (3x)		
	2.3. Clearly presents instructional content (spot) (3x)		
	2.4. Checks for academic understanding (2x)		
	2.5. Engages students at all learning levels in rigorous work (3x)		
	2.6. Activates higher-order thinking skills (2x)		
Domain 3: Classroom culture	3.1. Maximizes instructional time (spot) (3x)	Spot, extended and informal observations	21
	3.2. Maintains high student motivation (2x)		
	3.3. Maintains a welcoming environment that promotes learning and positive interactions (2x)		
Domain 4: Professionalism and Collaboration	4.1. Models good attendance for students	Artifacts and informal observations	15
	4.2. Follows policies and procedures, and maintains accurate student records		
	4.3. Engages in professional development		

Source: compiled from TEI Teacher Performance Rubric and the TEI Presentation

Table 2: Teacher categories and evaluation templates

Teacher Category	Teacher Performance	Student Achievement	Student Perception
Category A: Most grade 3-12 teachers whose students take an ACP, STARR, or AP exam, including most K-5 special teachers	50	35	15
Category B: Most K-2 teachers whose students take an ACP or ITBS/Logramos	65	35	0
Category C: Most grade 3-12 teachers whose students do not take an ACP, STARR, or AP assessment but who are able to complete a student survey (e.g. CTE teachers)	65	20	15
Category D: Any teachers whose students do not take an ACP, STARR, or AP assessment nor are eligible to complete a student survey (e.g. pre-K teachers. Teachers not-of-record such as SPED inclusion teachers, TAG teachers)	80	20	0

Source: Compiled from TEI Teacher Guidebook p.6 and TEI Rulebook p.9

Table 3: Compensation tied with teacher effectiveness levels in the initial year of TEI

Unsatisfied	Progressing		Proficient			Exemplary		Master
	I	II	I	II	III	I	II	
$45K	$49K	$51K	$54K	$59K	$65K	$74K	$82K	$90K

Source: Teacher Guidebook p36.

Table 4. Measuring principal effectiveness – the metrics

	Area	Points
Performance (60%)	Performance rubric	30
	System review	10
	Improving teacher effectiveness	5
	Congruence between teacher performance and student achievement	5
	Student enrollment or student attendance	5
	Parent climate survey	5
Achievement (40%)	School STAAR results	10
	Feeder group STAAR results	3
	District common assessments	7
	School achievement gap	5
	College ready rate (HS); 7th grade writing (MS); 4rd grade writing (ES)	10
	Career ready rate (HS); 8th grade reading and math (MS); 5th grade reading and math (ES)	5

Source: Principal Evaluation-Concept Paper-17 Jan 201 p. 5

Table 5: Synthetic control estimates and p-values of the effects on math and reading scores

	Donor pool includes largest 20 districts		Donor pool includes largest 50 districts	
Year	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)
2013	0.017	0.029	0.001	0.002
	[0.258]	[0.058]	[0.974]	[0.998]
2014	0.010	-0.017	0.015	-0.035
	[0.626]	[0.342]	[0.458]	[0.408]
2015	0.012	-0.055	-0.002	-0.065
	[0.751]	[0.003]	[0.922]	[0.001]
2016	0.074	0.028	0.040	-0.003
	[0.030]	[0.518]	[0.126]	[0.896]
2017	0.112	0.029	0.077	0.003
	[0.000]	[0.522]	[0.001]	[0.861]
2018	0.177	0.064	0.164	0.052
	[0.000]	[0.044]	[0.000]	[0.017]
2019	0.212	0.093	0.186	0.078
	[0.000]	[0.035]	[0.000]	[0.004]

Notes: This table provides exact estimates and p-values (in brackets) corresponding figures 2-5. The estimated effects in this table are the gap between Dallas and the synthetic control and the p-values are based on the permutation test described in the text.

Table 6. Average annual teacher value-added in Dallas ISD, by transition status following the school year

	All teachers		remain in the district teaching a tested subject and grade		remain in the district but no longer teach a tested grade and subject		exit the district	
	VA	N	VA	N	VA	N	VA	N
2013	0.042	761	0.079	411	0.018	226	-0.038	124
2014	0.037	760	0.080	439	-0.024	203	-0.017	118
2015	0.031	788	0.064	477	-0.028	171	-0.011	140
2016	0.027	787	0.063	498	-0.029	171	-0.042	118
2017	0.047	806	0.071	496	0.018	187	-0.007	123
2018	0.056	802	0.071	515	0.013	169	0.054	118

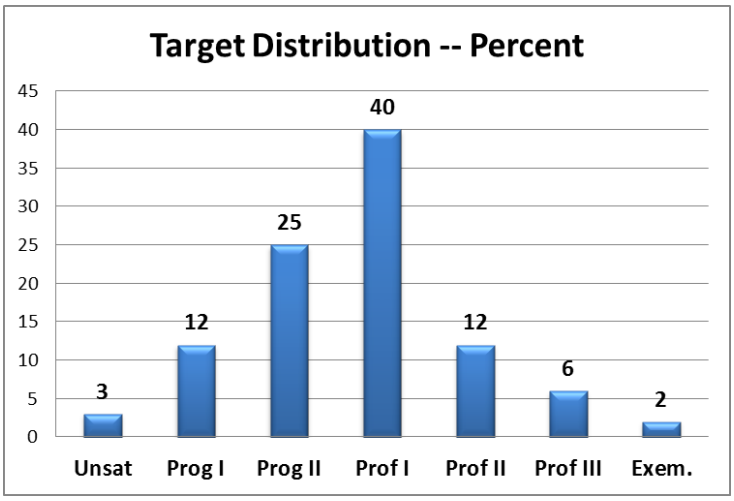
Notes: Value-added estimates come from separate specifications for each grade and year.

Table 7. Year dummy coefficients from regressions of math achievement on year indicators, by inclusion of teacher fixed effects and experience indicator variables (2013 is the omitted year; standard errors clustered by teacher in parenthesis)

teacher fixed effects	no	yes
dummy variables for single years of experience from 1 to 10 and an indicator for 11 or more years of experience	no	yes
2014	0.037	0.031
	(0.007)	(0.008)
2015	-0.037	-0.013
	(0.007)	(0.009)
2016	0.011	0.034
	(0.007)	(0.010)
2017	0.101	0.107
	(0.007)	(0.011)
2018	0.146	0.140
	(0.007)	(0.012)
2019	0.200	0.170
	(0.007)	(0.013)

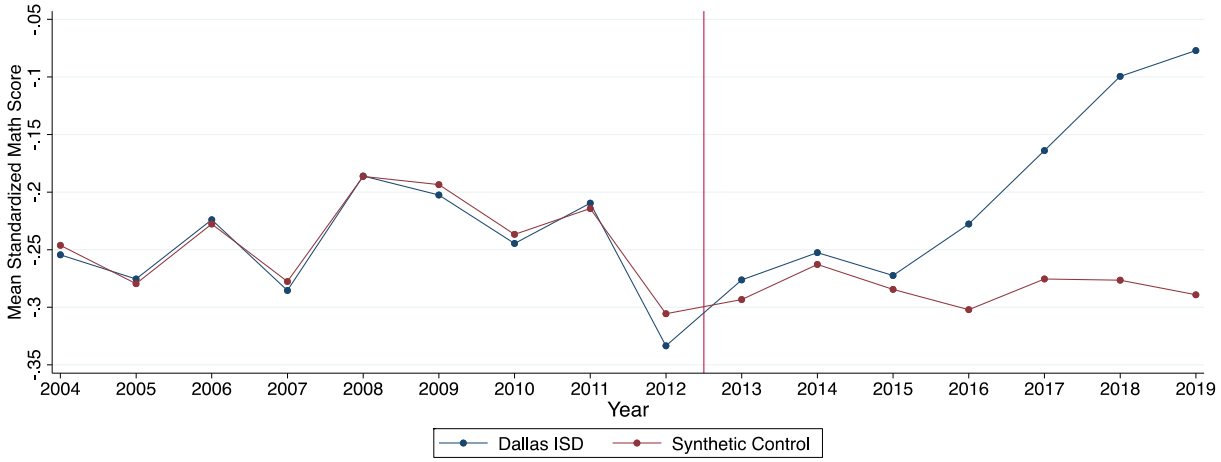
Notes: The coefficients in the left column come from a regression of math achievement on a full set of year dummies (2013 is the excluded year), and the coefficients in the right column come from a teacher fixed effect regression on a full set of year dummies and dummy variables for single years of experience from 0 to 9 and an indicator for 10 years of experience or more.

Figure 1: Target distribution of teacher effectiveness scales



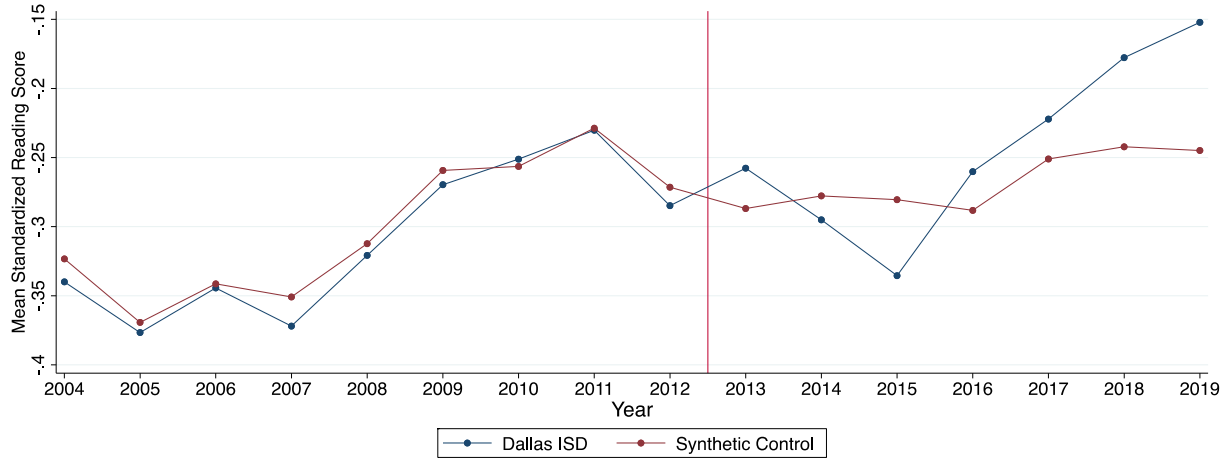
Source: TEI Rulebook v4.1 (DISD (2017)).

Figure 2. Synthetic control analysis of math achievement using the 20 largest high poverty districts



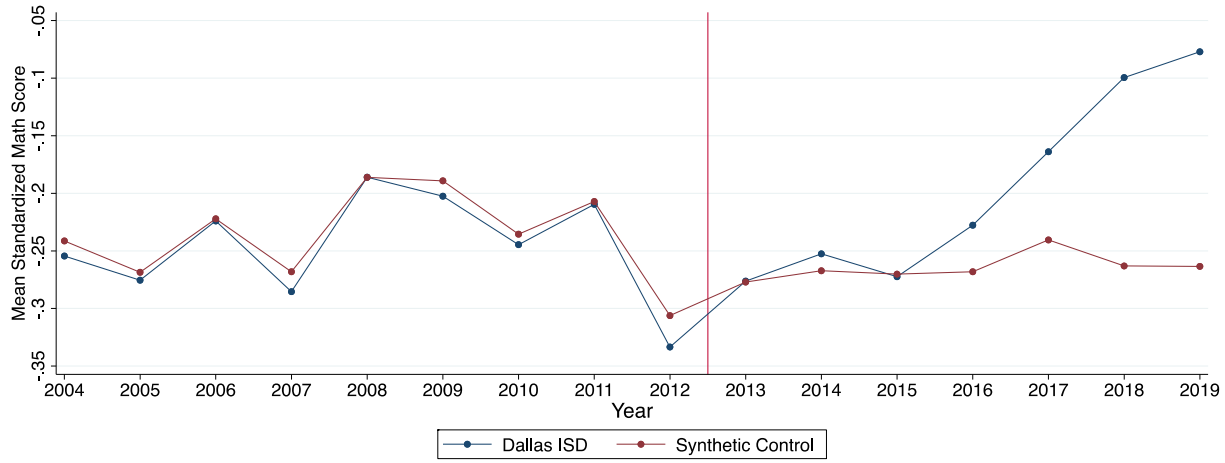
Notes: The figure plots average math achievement in Dallas ISD and the synthetic control over time. The synthetic control is constructed using schools from the 20 largest high-poverty districts as the donor pool.

Figure 3. Synthetic control analysis of reading achievement using the 20 largest high poverty districts



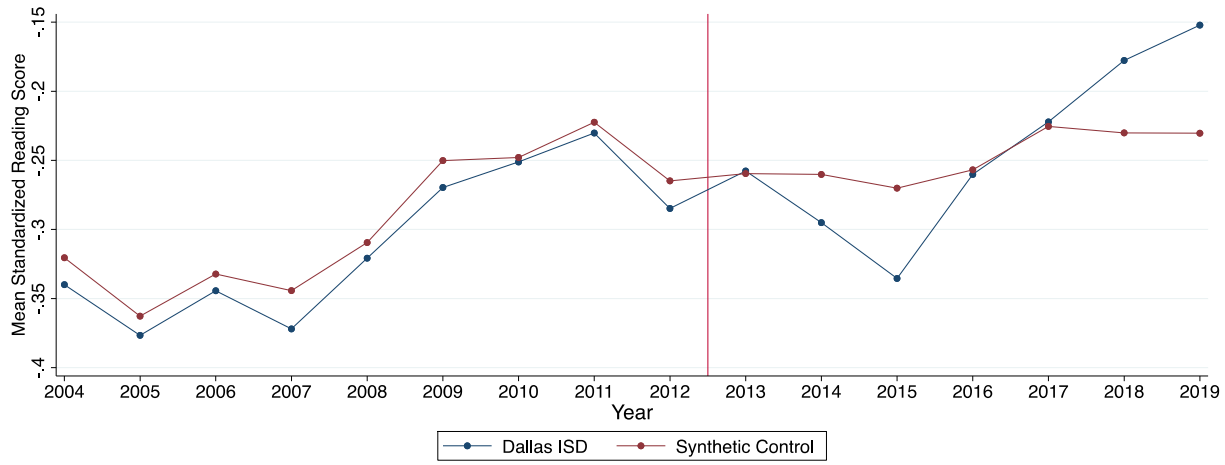
Notes: The figure plots average reading achievement in Dallas ISD and the synthetic control over time. The synthetic control is constructed using schools from the 20 largest high-poverty districts as the donor pool.

Figure 4. Synthetic control analysis of math achievement using the 50 largest high poverty districts



Notes: The figure plots average math achievement in Dallas ISD and the synthetic control over time. The synthetic control is constructed using schools from the 50 largest high-poverty districts as the donor pool.

Figure 5. Synthetic control analysis of reading achievement using the 50 largest high poverty districts



Notes: The figure plots average reading achievement in Dallas ISD and the synthetic control over time. The synthetic control is constructed using schools from the 50 largest high-poverty districts as the donor pool.

Figure 6. Dallas ISD teacher experience distribution: 2010-2019

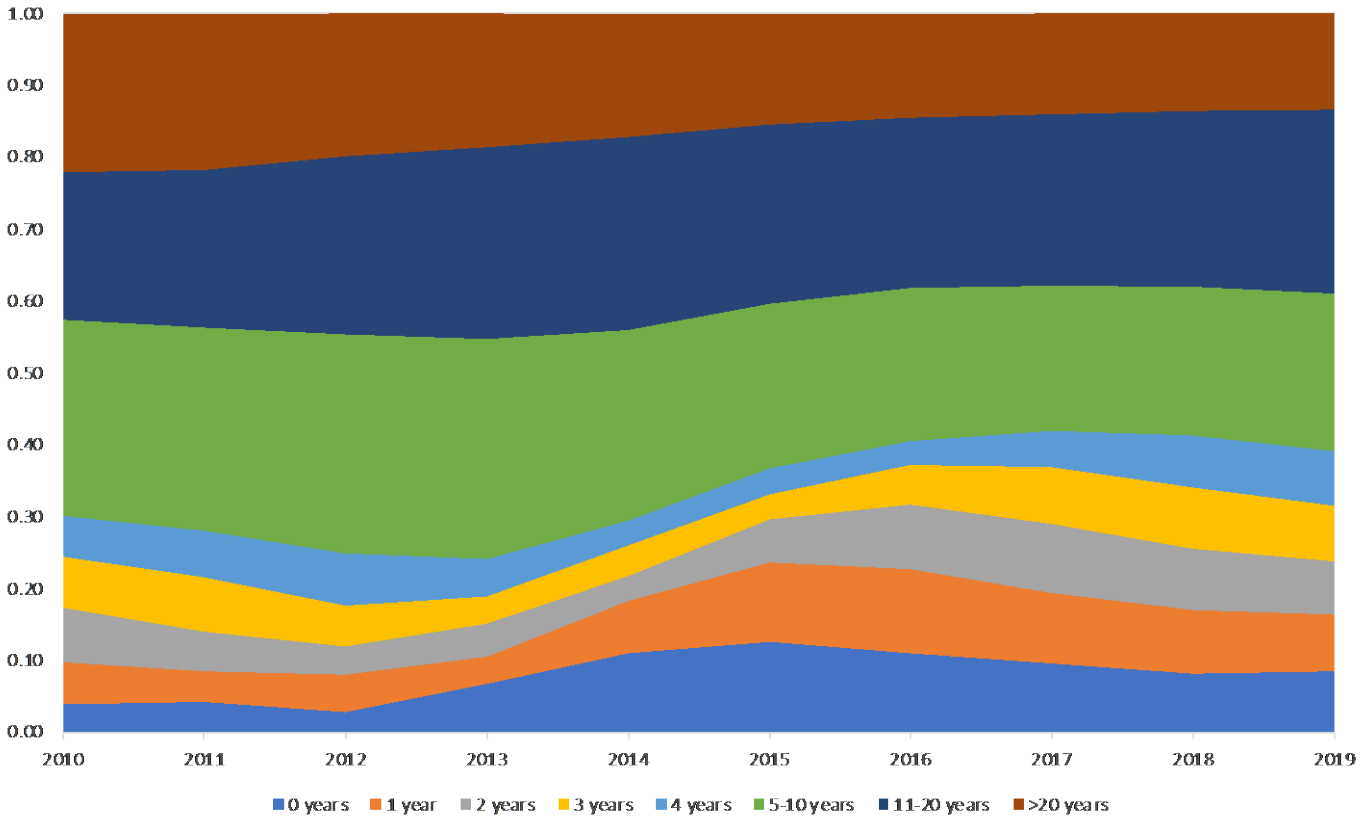
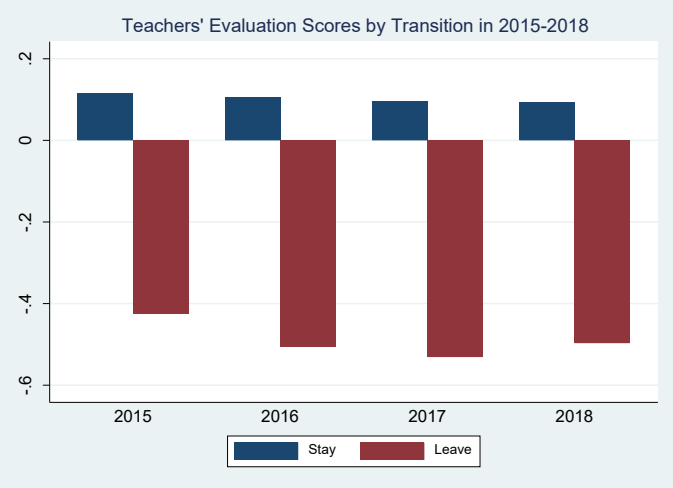
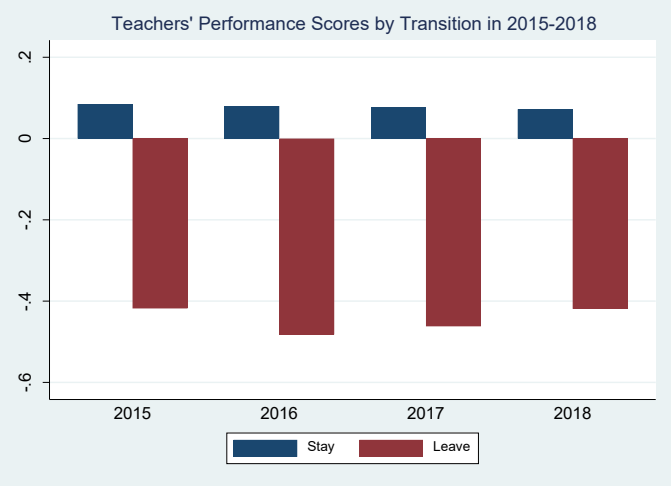


Figure 7. Mean teacher overall evaluation and component scores, by annual transition status





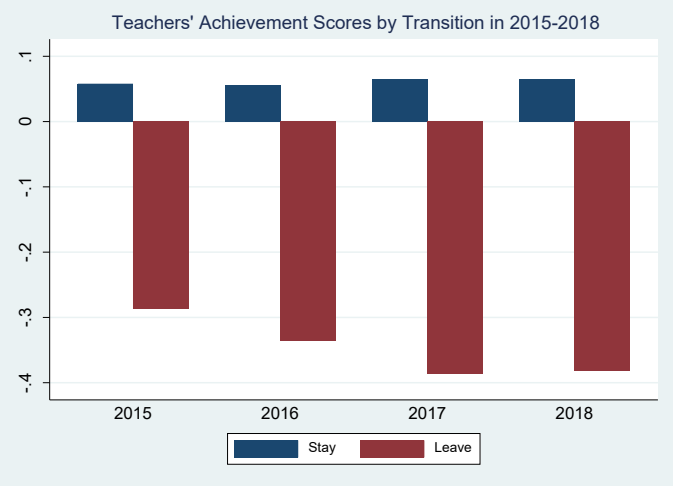


Figure 8. Survival rate of 2013 Dallas ISD elementary and middle school principals: 2014-2019

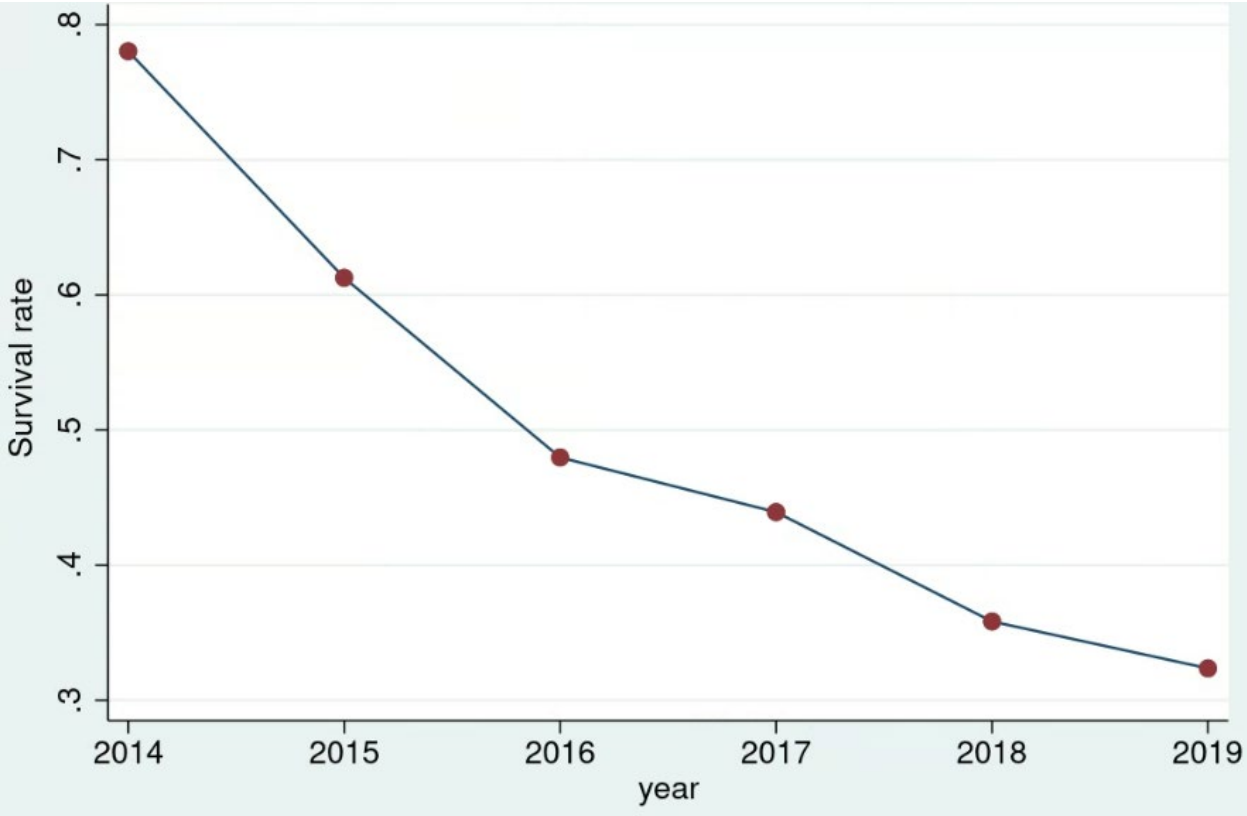
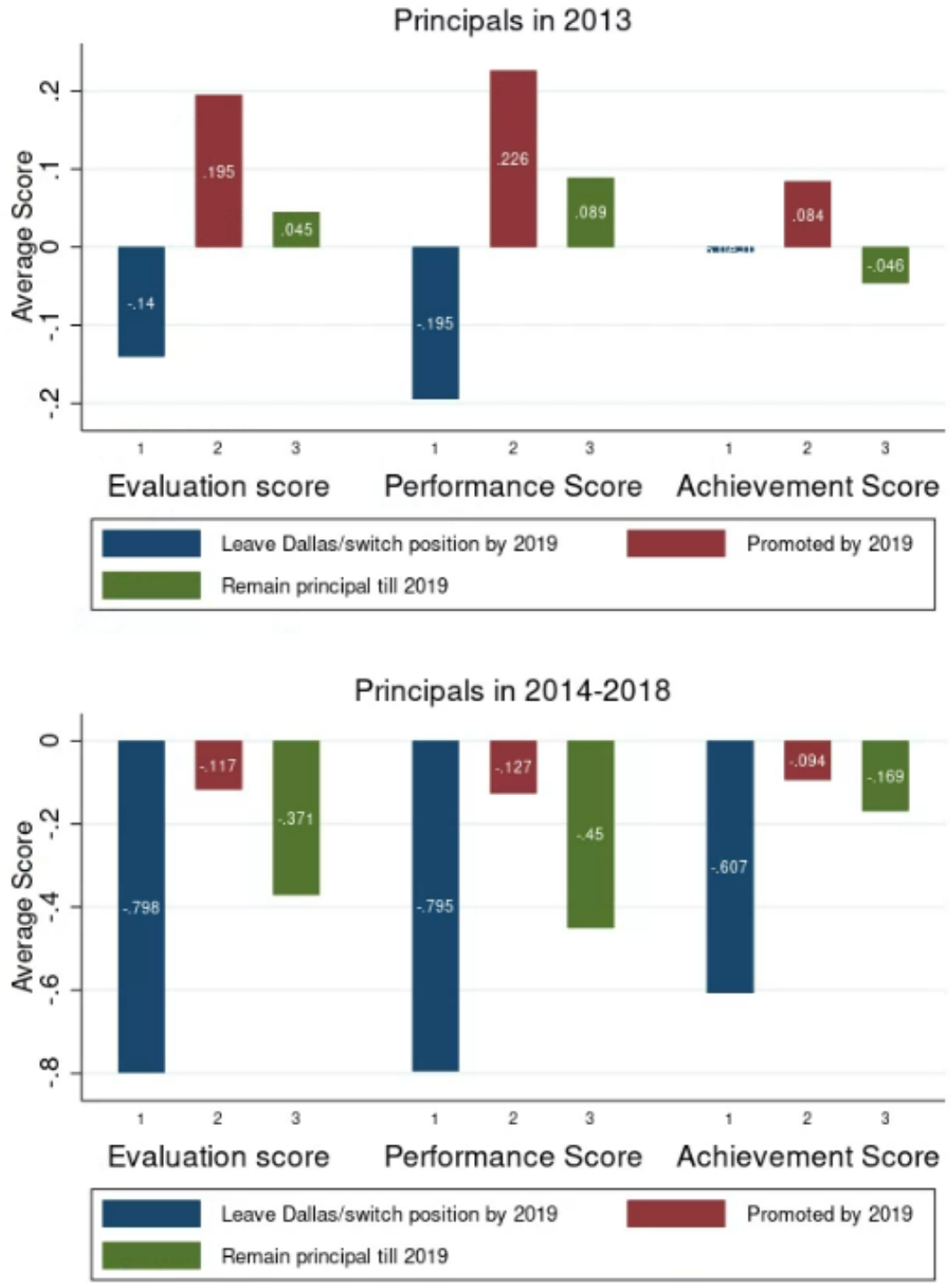


Figure 9. Mean principal overall evaluation and component scores, by 2019 transition status and year hired as principal



Notes: The top panel includes all Dallas ISD principals in 2013, and the bottom panel includes all Dallas ISD principals hired between 2014 and 2018. Effectiveness is measured in the first year a principal is observed following the implementation of PEI.