Journal of Education, Innovation, and Communication (JEICOM)

DOI: https://doi.org/10.34097/jeicom-7-2-2
Editors: Dr. Margarita Kefalaki & Dr. Fotini Diamantidaki

Exploring Differential Effects of a Teacher Incentive Program on Student Achievement

Yaxin Wang[1]

Margarita Pivovarova[2]

Audrey Amrein-Beardsley[3]

ABSTRACT

Traditional teacher compensation systems based solely on observable characteristics (such as experience and degree) have shown weak correlations with student learning outcomes. Teacher merit-pay incentives aim to strengthen the links between teacher remuneration and student learning outcomes. In this study, we examined the student achievement outcomes associated with one such incentive in the Dallas Independent School District – the Teacher Excellence Initiative (TEI) implemented in 2015. Utilizing the 1,121,557 student-level observations from 2012 to 2019 academic years, we investigated the relationships of TEI with student achievement from different demographics. We found the overall association of TEI with student achievement in mathematics and reading to be inconclusive. Conversely, we observed some benefits associated with Asian American, African American and Hispanic/Latinx students, relative to White students. We also observed the negative association between the TEI and the outcomes for students with special needs.

Keywords: teacher evaluation, teacher incentives, performance pay, student achievement, heterogeneous effects

[1] Corresponding Author - Arizona State University, USA - ywan1305@asu.edu

[2] Arizona State University, USA - margarita.pivovarova@asu.edu

[3] Arizona State University, USA - audrey.beardsley@asu.edu

1 INTRODUCTION

Teacher quality is central to many educational policies aimed at improving student learning and achievement outcomes. The current single salary schedule commonly used in public schools compensates teachers based solely on their years of experience and degrees, while decades of empirical research have shown that these characteristics are weak predictors of teacher quality (Hanushek & Rivkin, 2007; Moore et al., 2009; Murnane & Olsen, 1989; Rivkin et al., 2005). Yet, these are the factors on which many current teacher payment schedules are based. Subsequently, efforts to connect teaching effectiveness to teacher compensation have been manifested in federal policy initiatives (e.g., Race to the Top [American Recovery and Reinvestment Act, 2009], No Child Left Behind Act [NCLB, 2001]). These initiatives often framed as incentive programs yielded differential outcomes on student achievement across the nation, and overall evidence of their effectiveness has been mixed. While findings in some studies pointed that financial incentives directed at teachers have improved student achievement (Dee & Wycoff; 2015; Hanushek et al., 2023), other studies suggested the opposite (Fryer 2013; Glazerman et al., 2009; Glazerman & Saifullah 2010, 2012; Goodman & Turner 2013; Marsh et al. 2011; Springer et al., 2012a, 2012b). The most recent meta-analysis (Pham et al., 2021) of the findings from 41 studies suggests that teacher merit pay has a positive and statistically significant effect on student test scores, with the effect varying based on program design and study context.

To address the research question, we leverage student-level performance data from Dallas Independent School District (ISD) from 2012 to 2019 academic years. Dallas ISD is one of the first school districts in Texas to initiate a teacher merit pay system. In 2015, the Teacher Excellence Initiative (TEI) was established to reform the teacher compensation system. The program redesigned the conventional single salary schedule to a performance pay system where teachers were evaluated based on a comprehensive set of measures including classroom observations, student's experiences (measured via student survey), and standardized test score (measured by the yearly state assessment). This study focuses on changes in a student's test score after the program implementation, while recognizing the significance of other potential measures of student learning such as grade point average, graduation, and college attendance. Despite these limits, our analysis of student performance provides important policy perspective as teacher merit pay is at the heart of the public education reform debate. Our findings provide a novel update to previous narrative review of this literature. This study contributes to the current literature of teacher performance pay by unpacking the heterogeneous effect of the teacher incentive program, and in particular the varying effects by student characteristics which we hope will help to inform the policy conversation about the benefits and risks of teacher merit pay.

2 LITERATURE REVIEW

After years of studies of the relationship between teacher incentives and subsequent student academic outcomes, there is still no apparent consensus about the magnitude and even direction of these relationships. Various programs that have been thoroughly evaluated demonstrated a wide range of outcomes. Dee and Wycoff (2015), for example, found a positive relationship between student learning outcomes and teacher compensation from Washington DC's IMPACT teacher evaluation and feedback system. Similar to TEI in Dallas, IMPACT was used to evaluate teachers based on three major components including classroom observations, principal evaluations, and student test scores. Teachers rated as highly effective were eligible for a one-time bonus of $25,000. Analyses indicated a large effect of the monetary incentive which improved student scores taught by high-performing teachers by 10 percentage points.

Opposite results were found in New York City public schools which implemented a randomized, school-based intervention. The intervention awarded each participating school up to $3,000 for every staff member if a school met the annual performance target set by the department of education based on the school report card scores. These scores were determined by student performance and progress on state assessments, student attendance, and learning environment survey results. From the 2007-2008 through 2009-2010 school years, over 200 schools participated in the program, and more than 20,000 teachers received a total of approximately $75 million (Fryer, 2013). Evaluations of these incentives, however, did not yield any statistically significant or practically meaningful increases in student achievement.

Additionally, previous researchers have largely focused on examining performance bonuses' impact on overall student outcomes paying less attention to how these impacts could vary across student demographics including race and socio-economic status. At the same time, racial and socio-economic achievement gaps are well documented (Ashenfelter et al., 2006; Bassock et al., 2016; Ladd, 2012; Reardon et al., 2022). The racial achievement gap persists through all grades and is significant in magnitude; however, it does not seem to be related to mental functioning among children 8 to 12 months old (Fryer, 2013). This suggests that the achievement gap is not innate but might develop over time. In line with earlier patterns observed in older children, Dickens and Fryer (2001) proposed that factors other than genetic variations play a more substantial role in later learning outcomes. This highlights the potential for exploring the heterogeneous effects of interventions on students.

Most recently and related to the present study, Hanushek et al. (2023) investigated the effects of the TEI, and found an overall positive effect on student achievement with a 0.09 increase in reading in 2019 and a 0.21 standard deviation increase in math. The authors, however, did not explore the potentially heterogeneous effects of the TEI by student demographics. With the availability of student-level data, in study we intend to fill that gap and investigate the effect of merit pay on a student's achievement by subject, grade level, race/ethnicity and income status.

2.1 Program Background

The Dallas ISD serves approximately 155,000 pre-K to grade 12 students. It launched its TEI program in 2015 to improve student achievement and help the district's efforts to recruit and retain quality teachers (TEI, n.d.). The TEI requires the participating teachers to be evaluated on three components: (1) student achievement measured by the value-added score on the State of Texas Assessment of Academic Achievement (STAAR) test or the Assessment of Course Performance (ACP); (2) teacher performance, measured by class observations of instruction; and (3) student experiences, measured by a student-level survey. Teacher performance is assessed on these dimensions and classified into nine effectiveness categories, ranging from Unsatisfactory, multiple levels of Progressing, Proficient, Exemplary, to Master (TEI, n.d.). The weighted average of these component scores constitutes a teacher's overall evaluation score (see Table 1) which serves as the basis for TEI-based teacher compensation.

Table 1. Teacher Excellence Initiative Evaluation by Teacher Category

Teacher Category	Teacher Performance	Student Achievement	Student Experience
Category A			
Grade 3-12 teachers whose students take an ACP and/or STAAR	50%	35%	15%
Category B			
Grade 1-2 teachers whose students take a Growth Assessment	65%	35%	0%
Category C			
Grade 3-12 teachers whose students do not take an ACP or STAAR but who are able to complete a student survey (e.g., CTE teachers, elementary specials)	65%	20%	15%
Category D			
Any teacher whose students do not take an ACP, STAAR nor are eligible to complete a student survey (e.g., Pre-Kindergarten/Kindergarten teachers, teachers not-of-record such as Special Education/Inclusion teachers)	80%	20%	0%

TEI is not the first and not the only teacher incentive program based on performance pay. School districts in other states across the country have either experimented with similar initiatives or continue to implement them. Some of the most notable and well studied are Professional Compensation System for Teachers (ProComp) in Denver public schools (Goldhaber & Walsh, 2012; Fulbeck, 2014; Atteberry & LaCour, 2020) and IMPACT in Washington, DC schools (Koedel, 2014; Dee & Wyckoff, 2015; Dee & Wyckoff, 2017; James & Wyckoff, 2020). Less studied but also implemented at a large-scale are programs in Baltimore City public schools, Newark Public Schools, New Mexico, and Tennessee (Boudreaux & Faulkner, 2020). In Texas alone, there are other programs which tie student achievement to teacher pay and bonuses. These programs include the Governor's Educator Excellence Grant, the Texas Educator Excellence Grant, and the District Awards for Teacher Excellence. Many of the programs combine teacher evaluation systems with pay per performance and vary by the number and complexity of the indicators included in the final metric. The most successful programs share common characteristics which are also part of the TEI: the use of multiple measures to comprise the overall evaluation rating for an individual teacher; least three measures of evaluation and some measures of student learning, as well as observations and, in many cases, student surveys. These systems are designed not only to give teachers feedback and support but also to inform personnel decisions such as eligibility for leadership roles, raises, or retention in the classroom.

These programs share common objectives with TEI, such as linking teacher compensation to performance and aiming to improve student outcomes. However, their structures and outcomes vary, reflecting the complexities and challenges inherent in implementing performance-based compensation systems in education. These differences across the programs in their structure and implementation to some extent invalidate the comparison of the program outcomes. In the discussion and conclusion section, we describe some of the differences and similarities between the outcomes of the programs.

3 DATA

Consistent with the aims of the TEI, the primary research question we asked in this study is whether the introduction of Dallas ISD's pay-per-performance initiative was associated with increased student achievement and whether these achievement gains (if any) varied across students with different demographic characteristics including race and socio-economic status.

To answer this question, we acquired administrative student-level data from the Texas Education Agency (TEA). The dataset included all students enrolled in grades 3 to 8 within Texas public school districts spanning the academic years from 2011-2012 to 2018-2019. This included four years before and four years after the implementation of TEI (i.e., before and after 2015). To ensure comparability, we limited our sample to students in grades 3 to 8 who were instructed by TEI category A teachers (see Table 1) and who took the STAAR or ACP exams. We also retained variables pertaining to student demographic characteristics which included student race, gender, English language learner [ELL] status, eligibility for free or reduced lunch [FRL] which is used as an indicator for low-income status, special education [SPED] status, and grade.

We also utilized these student-level variables to identify school districts that were comparable to Dallas ISD based on district student demographics. Out of the 1022 school districts in Texas, we selected Houston ISD and Fort-Worth ISD as our comparison districts given these two districts were the most comparable in terms of their size and student demographics. Table 2 presents summary statistics on student demographics for each of these districts.

The main outcome variable in our analysis is student achievement measured by student score on state standardized tests, STAAR and ACP. We chose this outcome also because it is one of the key indicators of teacher effectiveness as per the TEI. The data requested from the TEA included raw and scaled scores on STAAR and ACP tests. It is important to note, though, that a substantial test transformation occurred in the STAAR tests in 2015, including changes to the tests' total points, test content, and measurement scales (Texas Education Agency, n.d.). Consequently, the application of either the original scale or raw scores for conducting inferential statistical analysis might be problematic due to the misalignment of raw scores that resulted given these changes.

While the STAAR test still maintained consistent difficulty levels across these years, disparities among test administrations were also likely inevitable.

STAAR assessment underwent a test change in AY 2015. As displayed in figure 1 & 2, the standardized test score jumped significantly in 2015. The department of education at Texas made changes in test content, format and scale. As a result, the raw scale scores became less comparable over time. To address these issues, we converted the scaled scores into percentiles, which serve as a common measurement that adjusts for the over-time changes in testing procedures and content.

Percentile scores rank students relative to their peers, offering a more consistent framework for assessing performance across different years. This relative measure is particularly useful when exploring the heterogeneity of student performance, as it allows for more meaningful comparisons across student subgroups. This allowed us to track the changes in performance based on relative standing of student scores in the distribution from one year to another. Additionally, this method helps minimize measurement errors that may arise from varying testing standards, ensuring that the analysis remains robust even as the testing conditions evolve.

In Figures 1 and 2 we plotted math and reading performance by race and ethnicity in Dallas ISD. White students did not display significant improvements following TEI implementation. At the same time, we observed a narrowing of the gap between White students and students of colour over that period. The upward trend after 2015 for Hispanic and Black students suggests a potential positive impact of TEI on achievement among these students. Visually, the graph shows that the introduction of TEI could have been associated with changes in student's learning outcome and that the potentially beneficial effects were concentrated among students of colour.

Table 2. Student Demographics in Dallas ISD, Houston ISD, and Fort Worth ISD, AY 2012 -2019

	Dallas ISD	Houston ISD	Fort Worth ISD
Hispanic/Latinx	0.71 (0.45)	0.63 (0.48)	0.64 (0.48)
Black	0.23 (0.43)	0.25 (0.43)	0.25 (0.43)
White	0.04 (0.20)	0.08 (0.27)	0.10 (0.30)
Asian	0.01 (0.10)	0.03 (0.18)	0.01 (0.12)
Low-Income	0.89 (0.32)	0.81 (0.39)	0.84 (0.37)
ELL	0.18 (0.38)	0.12 (0.33)	0.15 (0.36)
SPED	0.07 (0.26)	0.07 (0.26)	0.07 (0.26)

Note: The numbers in the cells represent the percentage of the respective subgroups. The population of American Indian, Hawaiian Pacific Islander, and two/races are not listed above since the share of students was less than 1 percent.

Figure 1. STAAR Reading Percentile by Student's Race/ Ethnicity, AY2012-19

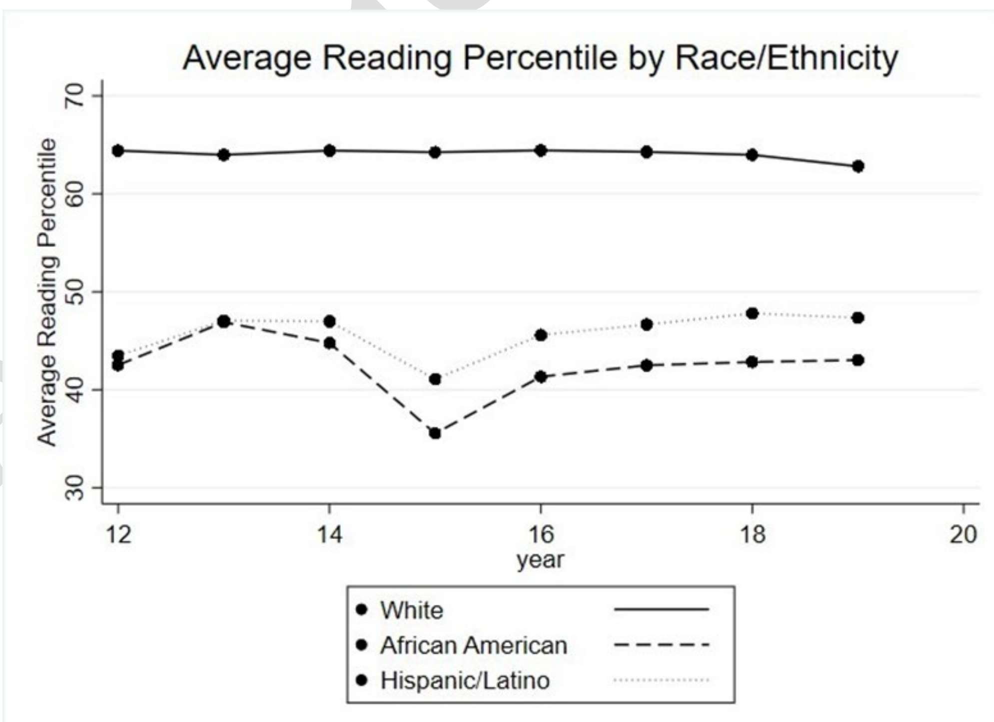
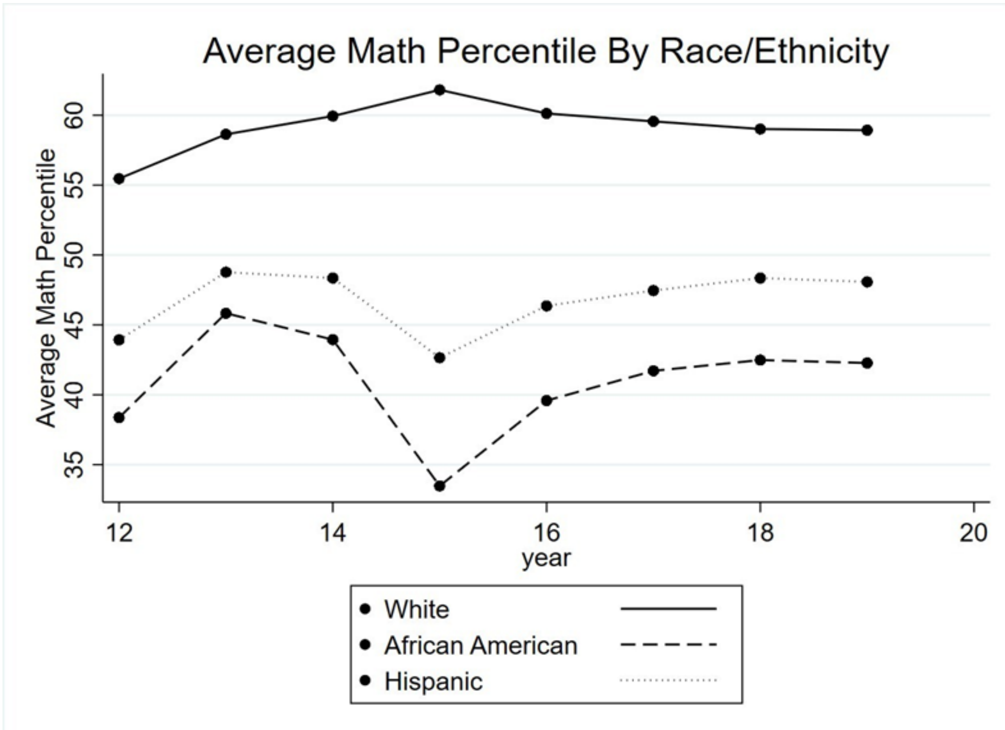


Figure 2. STAAR Math Percentile by Student's Race/ Ethnicity, AY2012-19



4 METHOD

To understand whether the introduction of the TEI was associated with changes in student achievement, we statistically compared the changes of Dallas ISD's student-level achievement over time against our two other school districts—Houston ISD and Fort Worth ISD. Our general approach is similar to the traditional difference-in-differences (DID) method used to estimate the causal effects of policy changes in the absence of random assignment. The DID method defines a group which experienced a change in policy which could have potentially affected the outcome of interest as a treatment group and a group with similar pre-change trends in the outcome but which did not see a change in policy as a control group. In our case, we defined the implementation of the TEI in Dallas ISD as treatment and Dallas ISD as a treatment group. Two other districts – Houston ISD and Fort Worth ISD – were designated as a control group. Our strategy is based on comparing changes in average scores before and after TEI implementation in Dallas ISD (first difference) to changes in other two districts (second difference) over the same period. We enhanced the conventional DID method with student fixed effects since we could identify students consistently enrolled throughout the data period. Our DID model was defined as follows:

$$Y_{igtd} = \beta_0 + \beta_1 Dallas_{igtd} + \beta_2 After_{igtd} + \beta_3 After_{itgd} * Dallas_{igtd} + \beta_5 X_{igtd} + \beta_6 X_{igtd} * Dallas_{igtd}$$

In this equation, the dependent variable Y_{igtd} is a percentile score on STAAR test of student i in grade g in district d in the calendar year t. The indicator of treatment status of a student i at district d in calendar year t was represented using $Dallas_{igtd}$ which equalled 1 for Dallas ISD and 0 for our other districts. Its coefficient captured the average effect of being a student at Dallas ISD on the outcome variables. $After_{igtd}$ is an indicator of years after implementation of the TEI in Dallas and was equal to 1 for the period after 2015 including year of 2015, and 0 otherwise. The coefficient on the indicator would measure the difference in average performance on standardized test across all districts after 2015. Our main variable of interest is an interaction term between treatment and time $After_{igtd} * Dallas_{igtd}$ which is a binary variable which takes value of 1 for a student i in Dallas ISD after the implementation of the TEI, i.e., after 2015 inclusive of 2015. The coefficient on this interaction would capture the average change in student achievement in the treated district versus two control districts after the implementation of the TEI program. The key identifying assumption for this interpretation was the differences in student achievement between treated districts and controlled districts was consistent before the intervention. X_{igtd} stands for the student demographics for student i in grade g in district d in the calendar year t including the gender, race/ethnicity, low-income household status, English language learner status and special education status. The interaction term between the student characteristics of X_{igtd} and $After_{igtd} * Dallas_{igtd}$ captures of the heterogeneous effect of the performance-pay program by student subgroups.

To assess the validity of this assumption, we compared whether the trajectory of a students' test score in the control and treatment groups followed a parallel trend during the pre-intervention phase. When these trends exhibited similarity before the intervention, it suggested that they would likely continue to align in the post-intervention period if the treated district had not implemented the TEI compensation reform. In addition, we conducted a parallel trend test using a pre-installed statistical package in Stata (2022) to estimate the null hypothesis (i.e., the linear trend was parallel prior to the treatment). If the p value was greater than 0.05, we would not have sufficient evidence to reject the null suggesting the linear trend was parallel (Stata, 2022). Our test results are listed in Tables 3 and 4. In math the trend before the program implementation satisfied the parallel trend assumption but not in reading; therefore, we treated model estimations in reading with more caution.

5 RESULTS

After TEI implementation, overall reading score for White students declined by 4.44 percentiles. In contrast, we observed an increase in the average reading score after TEI among students of colour. Asian American students' scores increased by 3.65 percentile points, average score among Black students rose by 3.42 percentiles, and Hispanic/Latinx students demonstrated a 3.85 percentile increase. However, it is important to note that students with special needs had a substantial decline in performance post-program, scoring 10.01 percentiles lower compared to their peers. For mathematics, no statistically significant shifts were observed for White students, while students of colour, again, experienced an improvement in average score after the program implementation. Asian American students improved their score rankings by 4.01 percentiles, African American students improved by 1.18 percentiles, and Hispanic/Latinx students improved by 2.51 percentiles. Results are presented in Tables 3 and 4 along with the trend plot of the predicted reading and mathematics plotted in Figures 3 and 4. These findings are consistent with graphical representation of trends in test scores over time broken down by student race.

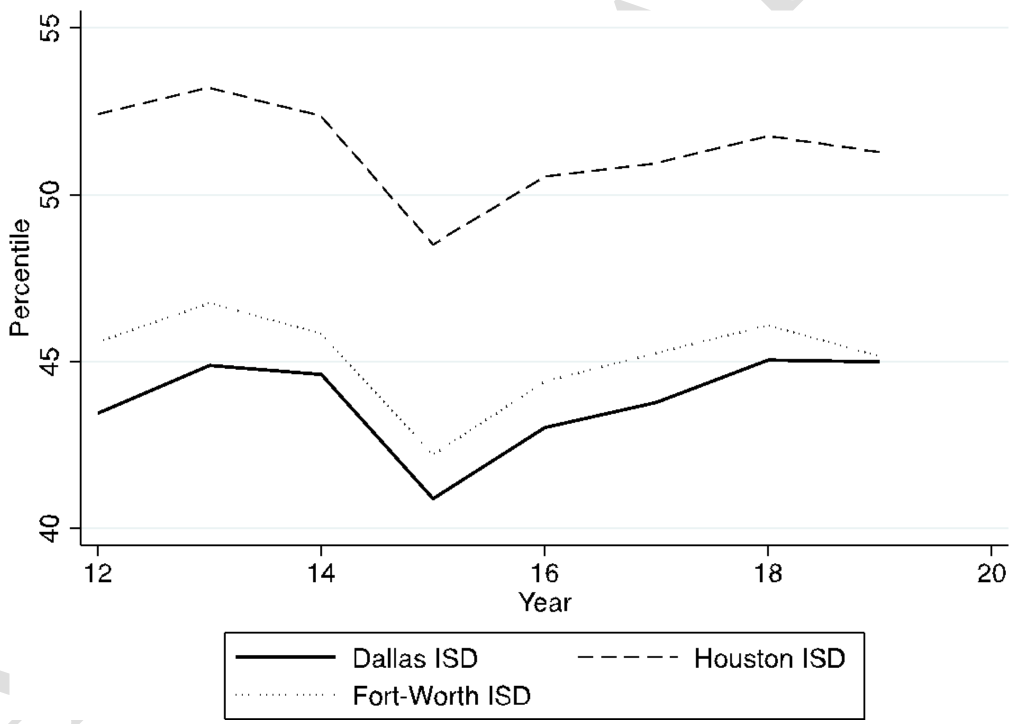
Table 3 Difference-in-Differences Model Estimations for Reading

Variables	DID Coefficients	DID Standard Error (SE)	*p*-value
Asian	6.04	0.25	0
Black	-13.71	0.17	0
Hispanic/Latinx	-8.71	0.15	0
Others	-1.41	0.43	0.001
Gender	2.78	0.10	0
Low-Income	-11.72	0.13	0
ELL Student	-17.01	0.19	0
SPED	0.60	0.21	0.005
Asian*After	1.98	0.34	0
Black*After	0.53	0.24	0.025
Hispanic*After	0.80	0.21	0
Others*After	-1.38	0.57	0.015
Gender*After	0.41	0.14	0.003
ELL*After	2.03	0.25	0
Low-Income*After	0.00	0.17	0.99
SPED*After	-19.79	0.29	0
Asian* Dallas	-16.61	0.69	0
Black* Dallas	-7.42	0.37	0
Hispanic* Dallas	-5.89	0.35	0
Others*Dallas	-13.58	1.06	0
Gender*Dallas	0.67	0.15	0
ELL* Dallas	4.72	0.24	0
Low-Income*Dallas	5.37	0.23	0
SPED*Dallas	9.46	0.30	0
Asian*After*Dallas	3.65	0.93	0
Black*After*Dallas	3.42	0.49	0
Hispanic*After*Dallas	3.85	0.46	0
Others*After*Dallas	5.58	1.35	0
Gender*After*Dallas	0.13	0.21	0.546
ELL*After*Dallas	1.65	0.32	0
Low-Income*After*Dallas	1.44	0.31	0

We also observed differences in TEI effects across grade levels. Among elementary schools, we observed marginal score improvements. In reading, 4th grade students have shown a 2.52 percentile increase in performance, but students in other grade levels did not display statistically significant changes. Conversely, for math, we observed an average increase of 4.64 percentiles. Within middle schools, students, on average, have demonstrated a lower performance on annual assessments in comparison to preceding cohorts. In reading, students, again on average, scored 10.55 percentiles lower except for 7th graders. In math, we observed improvements among 6th and 8th graders, but their overall performance average decreased by 4.72 percentiles after TEI implementation.

Students who were eligible for free and reduced-price lunch, on average underperformed across all districts. After the implementation of the TEI we observed a positive change in both math and reading scores for low-income students in Dallas ISD with a larger magnitude in reading. The magnitude of the change is comparable to the original gap in performance between FRL eligible and other students which implies that TEI has a potential to narrow and even close the gap.

Figure 3 Predicted STAAR Reading Percentile, Dallas ISD, Houston ISD and Fort Worth ISD, AY2012-19



Due to the varied evidence that we observed across analyses, we cannot draw a firm conclusion as to whether this specific performance-pay program improved all student's learning outcomes. Our results suggest a mixed set of evidence pertinent to this teacher incentive program and provide positive association between the teacher performance pay for some student demographics in Dallas ISD, but not overall.

One limitation of our study is that we do not have teacher-level data and unable to link students to teachers. As a result, we could not investigate whether and how the effects of TEI varied by teacher – for instance, if more effective teachers as measured by TEI indicators who received higher incentive payment were able to raise the achievement of their students to a higher degree compared to less effective teachers defined by the TEI metrics.

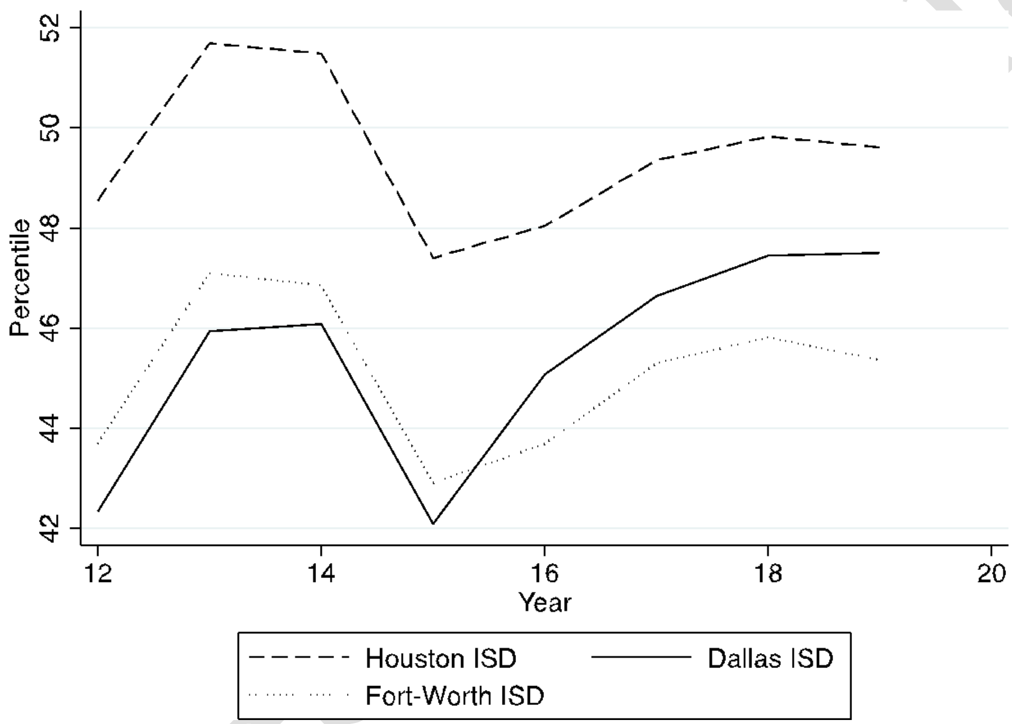
Another limitation of the study brought by the specificity of the data is that we were not able to account for teacher characteristics in our models – such as experience, education, certification, and teacher's demographics. Given the recent findings from the literature on student-teacher matching, especially for students of colour, we anticipate that our results would underestimate the effects of TEI for students of colour who were taught by teachers of the same race.

Table 4 Difference-in-Differences Model Estimations for Math

Variables	DID Coefficients	DID Standard Error (SE)	*p*-value
Asian	8.98	0.25	0
Black	-12.92	0.18	0
Hispanic/Latinx	-5.73	0.16	0
Others	-2.03	0.44	0
Gender	-0.50	0.10	0
Low-Income	-7.85	0.13	0
ELL Student	-12.68	0.19	0
SPED	1.86	0.22	0
Asian*After	-0.29	0.34	0.404
Black*After	0.85	0.24	0
Hispanic*After	-0.55	0.21	0.009
Others*After	-2.04	0.58	0
Gender*After	-0.63	0.14	0
ELL*After	3.03	0.25	0
Low-Income*After	-0.89	0.17	0
SPED *After	-15.81	0.29	0
Asian*Dallas	-14.13	0.70	0
Black*Dallas	-3.68	0.37	0
Hispanic*Dallas	-2.17	0.35	0
Others*Dallas	-10.53	1.07	0
Gender*Dallas	0.68	0.16	0
ELL*Dallas	5.99	0.25	0
Low-Income*Dallas	5.07	0.23	0
SPED *Dallas	9.00	0.31	0
Asian*After*Dallas	4.01	0.94	0
Black*After*Dallas	1.18	0.50	0.017
Hispanic*After*Dallas	2.51	0.47	0
Others*After*Dallas	4.11	1.38	0.003
Gender*After*Dallas	0.55	0.21	0.01
ELL*After*Dallas	-2.45	0.33	0

Low-Income*After*Dallas	1.64	0.31	0
SPED *After*Dallas	-10.45	0.43	0
After*Dallas	-0.22	0.44	0.618
After	2.12	0.18	0
Dallas	-5.37	0.33	0
Intercept	61.48	0.13	0
Parallel Trend Test (p-value)		0.368	

Figure 4 Predicted STAAR Math Percentile, Dallas ISD, Houston ISD and Fort Worth ISD, AY2012-19



In addition, the heterogeneous effect of the program has been observed across grades. For reading, the program works better for elementary schools than middle schools. Grade 3 to Grade 5 have shown improvement in both reading and math test scores after program implementation; third graders score 5.4 percentile higher for math ($t= 9.01$, $P<0.00$), on average primary school students score 2.7 percentile higher for reading and 5.5 percentile higher for math. The program, however, did not seem to produce an effective outcome for higher grade students. Middle school students perform worse on reading. Particularly eighth graders score 5.65 percentile lower after program implementation. Overall, all grades have seen improvement on math except for grade 7.

Table 5 The Estimated Effects of TEI on Student's Academic Achievement

Grade Level	Reading	Math
Grade 3	1.91	5.4
	(0.04)	(0.00)
Grade 4	2.77	7.32
	(0.00)	(0.00)
Grade 5	3.30	4.36
	(0.00)	(0.05)
Grade 6	-0.69	1.88
	(0.06)	(0.54)
Grade 7	0.08	-9.90
	(0.84)	(0.10)
Grade 8	-5.65	3.24
	(0.13)	(0.07)
Primary School	2.70	5.50
Middle School	-2.14	-0.40

Note: Student's standardized test scores are calculated as percentiles to compare changes across years.

6 ROBUSTNESS CHECK

To assess our model estimates' reliability, we conducted a robustness check via multilevel model (MLM) with repeated measures clustered at the student level. The model was clustered at student level with individual characteristics including student race, English language learner (ELL) status, special education (SPED) status, free and reduced lunch eligibility (low-income household), gender, year, and TEI implementation. The time variable in the model is centred around 2015 when Dallas ISD implemented the program. We used the interaction term between time variable and indicator for Dallas ISD to measure TEI effectiveness which is $Dallas * Year$. We also introduced three-way interactions (i.e., a student characteristic, period after TEI, and a Dallas ISD dummy variable) to estimate differential effects by student demographics. We estimated the MLM model using Maximum-Likelihood estimation Version 4.0 (Stata, n.d.) which predicts students' performance based on their demographics and random effects clustered at the student level. The MLM model we defined as follows:

Level 1 Model:

$$Student\ s\ Achievement = \beta_{0i} + \beta_{1i}\ (Year) + \beta_{2i}\ (Dallas) + \beta_{3i}\ (Dallas * Year) + \varepsilon_{ti}$$

Level 2 Model:

$$\beta_{0i} = \gamma_{00} + \gamma_{01} AA + \gamma_{02} Latino + \gamma_{03} Asian + \gamma_{04}\ ELL + \gamma_{05} Low\ Income + \gamma_{06} SPED + \gamma_{07}\ Gender$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11} AA + \gamma_{12} Latino + \gamma_{13} Asian + \gamma_{14}\ ELL + \gamma_{15} Low\ Income + \gamma_{16} SPED + \gamma_{17}\ Gender$$
$$\beta_{2i} = \gamma_{20} + \gamma_{21} AA + \gamma_{22} Latino + \gamma_{23} Asian + \gamma_{24}\ ELL + \gamma_{25} Low\ Income + \gamma_{26} SPED + \gamma_{27}\ Gender$$
$$\beta_{3i} = \gamma_{30} + \gamma_{31} AA + \gamma_{32} Latino + \gamma_{33} Asian + \gamma_{34}\ ELL + \gamma_{35} Low\ Income + \gamma_{36} SPED + \gamma_{37}\ Gender$$

In the Level 1 model, β_{0i} was the model intercept that represented the predicted posttreatment math percentile for a white, not low income, not English language learner (ELL), and not special education student in a control district. β_{1i} was the coefficient of time centered at the year of 2015 when the TEI program was implemented. The coefficient for $Dallas^{*}Year$ represented the predicted post-treatment difference between white students in Dallas ISD and white students in control districts after controlling for gender, low-income status, ELL status, and special education status. ε_{ti} are the residuals. Our Level 2 model contained student-level characteristics, which included indicators for gender, race (Black and Hispanic/Latinx), ELL status, whether a student was from a low-income family, and whether a student was classified as a special education student.

Table 6 Multilevel Model Estimations of Reading Score

Variables	MLM Coefficients	MLM SE	p-value
Asian	5.54	0.23	0
Black	-15.34	0.16	0
Hispanic/Latinx	-10.72	0.14	0
Others	-3.34	0.38	0
Gender	3.13	0.10	0
Low-Income	-8.68	0.10	0
ELL Student	-13.00	0.14	0
SPED	-4.02	0.18	0
Asian*Year	0.64	0.08	0
Black* Year	0.45	0.06	0
Hispanic* Year	0.37	0.05	0
Others* Year	-0.03	0.13	0.790
Gender* Year	0.05	0.03	0.102
ELL* Year	1.59	0.06	0
Low-Income* Year	0.12	0.04	0.002
SPED* Year	-4.41	0.07	0
Asian* Dallas	-15.07	0.62	0
Black* Dallas	-4.64	0.32	0
Hispanic* Dallas	-2.95	0.30	0
Others*Dallas	-10.74	0.87	0

Gender*Dallas	0.65	0.15	0
ELL* Dallas	2.65	0.18	0
Low-Income*Dallas	4.87	0.18	0
SPED*Dallas	6.65	0.26	0
Asian* Year *Dallas	1.07	0.22	0
Black* Year *Dallas	0.93	0.12	0
Hispanic* Year *Dallas	1.16	0.11	0
Others* Year *Dallas	1.20	0.30	0
Gender* Year r*Dallas	-0.02	0.05	0.615
ELL* Year *Dallas	-0.10	0.075	0.186
Low-Income* Year *Dallas	0.70	0.07	0
SPED* Year *Dallas	-2.11	0.10	0
Year *Dallas	-1.46	0.10	0
Year	0.12	0.04	0.003
Dallas	-2.73	0.29	0
Intercept	64.38	0.12	0
Parallel Trend Test			

Table 7 Multilevel Model Estimations of Math Score

Variables	MLM Coefficients	MLM SE	*p*-value
Asian	8.76	0.22	0
Black	-13.21	0.15	0
Hispanic/Latinx	-6.69	0.13	0
Others	-3.48	0.36	0
Gender	-0.70	0.09	0
Low-Income	-6.83	0.10	0
ELL Student	-10.38	0.14	0
SPED	-2.51	0.18	0
Asian*After	-0.39	0.08	0
Black*After	0.57	0.06	0
Hispanic*After	0.01	0.05	0
Others*After	-0.28	0.13	0.790
Gender*After	-0.06	0.03	0.102
ELL*After	1.55	0.06	0
Low-Income*After	0.06	0.04	0.002
SPED *After	-3.53	0.07	0
Asian*Dallas	-13.19	0.60	0
Black*Dallas	-3.14	0.32	0
Hispanic*Dallas	-1.08	0.30	0
Others*Dallas	-8.68	0.88	0
Gender*Dallas	0.74	0.14	0.615
ELL*Dallas	4.08	0.19	0.186
Low-Income*Dallas	5.18	0.18	0
SPED *Dallas	4.70	0.26	0

Asian*After*Dallas	1.63	0.24	0
Black*After*Dallas	0.31	0.12	0
Hispanic*After*Dallas	0.76	0.12	0
Others*After*Dallas	0.99	0.32	0
Gender*After*Dallas	0.18	0.05	0
ELL*After*Dallas	-1.41	0.08	0
Low-Income*After*Dallas	0.36	0.07	0
SPED *After*Dallas	-2.46	0.10	0
After*Dallas	-0.16	0.11	0
After	-0.02	0.04	0.003
Dallas	-4.91	0.28	0
Intercept	61.16	0.11	0
Parallel Trend Test			

Consistent with our DID estimations, using the MLM approach we observed a small negative overall effect of TEI on student achievement ($\gamma =- 1.46$, $p < 0.001$) in reading and no significant changes in math. In reading, we found a small but statistically significant improvement of test scores for students of colour after TEI implementation whereby Asian American students improved their score rankings by 1.07 percentile, Black students improved their score rankings by

0.93 percentile, and Hispanic/Latinx students improved their score rankings by 1.16 percentile. However, we observed SPED students, again, to be disadvantaged after TEI implementation ($\gamma =- 2.11$, $p < 0.001$). In math we observed no substantial statistically significant differences among white students in Dallas ISD and students in Houston ISD and Fort-Worth ISD after implementation of TEI. Students of other races, again, seemed to have improved their percentile rankings, whereby Asian American students improved their score rankings by 1.63 percentile, Black students improved their score rankings by 0.31 percentile, and Hispanic/Latinx students improved their score rankings by 0.76 percentile. Lastly, we continued to observe statistically significant decreases in performance for SPED students ($\gamma =- 2.46$, $p < 0.001$). Across all our models, students who were eligible for free and reduced-price lunch (FRL), underperformed on both math and reading tests. The magnitude of the difference in scores between FRL-eligible and other students was comparable in magnitude to the gap between White and Hispanic students. This implies that students who come from low-income Hispanic households faced double disadvantage.

Due to the varied evidence that we observed across analyses, we cannot draw a firm conclusion as to whether this specific performance-pay program improved student's learning outcomes across the board. Our results suggest a mixed set of effects pertinent to this teacher incentive program and provide evidence of beneficial effects of the program for some student demographics in Dallas ISD, but not overall. Indirectly, our findings also imply that the TEI program may have improved the effectiveness of teachers of these students. This finding is in line with the recent assessment of TEI outcomes by Hanushek et al. (2023).

7 DISCUSSION & CONCLUSION

Our main goal in this study was to extend the research on pay-per-performance initiatives and explore potential heterogeneity of such programs. With over 70% of participating students belonging to diverse racial ethnicity backgrounds, we quantified some of the differential evidence of TEI on student learning outcomes by race as well as other demographics such as low income (i.e., FRL-eligible), SPED, and ELL. Taken together, our findings suggested a mixed set of effects of TEI incentives on student performance. In general, students did not appear to benefit significantly from the program; however, students of colour appear to make progress after the program implementation, albeit the magnitude of these improvements for the most part was not large. More specifically, we found that there were no significant changes in White students' achievement. Conversely, Asian American, Black, and Hispanic/Latinx yielded notable effects in both reading and math. One of the most alarming findings is that students with special needs encountered significant disadvantages, yielding approximately 10 percentile lower scores in both reading and math. We also found mixed evidence of the program's effects on student performances by grade level and subject area. The program showed a larger effect for elementary school students compared with middle school students which is consistent with the meta-analysis findings from previous literature (Pham et al., 2020). In terms of subject differences, the reading score for White students decreased by 4.44 percentiles and no statistically significant changes were found in mathematics. In general, these findings added ambiguity in answering whether students benefited from this TEI program.

Our findings aligned with the evaluation of the Texas's Governor Educator Excellence Grant Program (GEEG) launched in 2005 which targeted high-needs schools across Texas via similarly defined incentive systems - GEEG had a weakly positive, negligible, or negative effect on students' learning outcomes (Springer et al., 2009). In Texas, then, there still seems to be no strong, or not enough strong evidence yielding statistically or practically significant associations among student achievement gains and such incentive plans, like the TEI. At the same time, the evaluation of the TEI (as cited in Putnam et al., 2018) indicated an increase in student achievement by 7 percentage points in 2015-2017 closing the proficiency gap between Dallas and the state of Texas by 3 points. According to the authors of the evaluation report, these findings do not represent the causal effect of the program.

Our findings are not in line with the evidence from other similar programs mentioned prior. For instance, Denver Public Schools' students have consistently outpaced their classmates statewide in academic growth in English language arts and math after the implementation of the ProComp program in Denver Public Schools (Putnam et al., 2018). Similar findings were recorded for Tennessee Educator Acceleration Model (TEAM) and New Mexico. While the reports on the outcomes of these programs focus on teacher-level outcomes such as teacher quality and teacher retention, evidence on improvements in student achievement are limited (Kraft & Gilmour, 2017).

As we mentioned repeatedly, the observed outcomes from teacher performance pay programs vary. Most of this evidence comes from the observational studies when identifying the direct effect of the program is not possible due to confounding factors. At the same time, a small group of studies were able to evaluate the effect of teacher performance pay in an experimental setting. One such experimental program, POINT in Metro Nashville public schools district, provided causal estimates of teacher incentives on student performance (Springer et al., 2011) and found no evidence on average of the positive effect of performance pay on student scores.

Yet many factors may have contributed to the still-mixed evidence observed. For example, it is unclear if such programs' incentive structures provide teachers with just enough motivation to improve their teaching. Or whether the outcomes on which they are most motivated can be captured using students' standardized test scores (e.g., test representation and sensitivity to teachers' instructional effects). After TEI's first year of implementation, for many newly hired teachers, the average bonus was around $2,000 (Texas Education Agency, n.d.). This bonus amount may not have been a critical or prominent factor contributing to their professional or instructional decisions[4]. By comparison, the TEI program can compensate Dallas ISD's more experienced teachers (i.e., teachers who have at least two years of teaching experience) with a starting salary up to $58, 000. The unbalanced structure of TEI's (and perhaps other similar incentive programs') salary schedule may also contribute to the differential effects we observed. In addition, teachers are often unaware of the key aspects of these types of incentive pay-per-performance programs, with previous researchers noting how a lack of clear communications about these programs and their incentive structures (Glazerman & Saifullah 2012; Springer et al., 2012b) prevent teachers from effectively changing their teaching practices for the better.

While accountability programs were designed to improve educational quality and equity, they sometimes produced unintended consequences that undermined their goals. While the TEI aims to improve student achievement by linking teacher compensation and evaluation to performance metrics, it has also led to several unintended consequences which were also noted in other similar programs (Pham et al., 2021) and suggested by previous literature (Hanushek & Raymond, 2005; Figlio & Ladd, 2014; Hanushek et al., 2023). Under the No Child Left Behind (NCLB) administration, local education agencies had to achieve a certain learning outcome threshold to be granted funding. As theorized and empirically demonstrated in several studies, this could lead to attempts of gaming the system, narrowing curriculum, teaching to the test, lowering of teacher and student morale, teacher burnout and stress. For instance, to meet the accountability objective, some schools may reclassify students to exempt them from taking tests. As one of the examples, Florida Opportunity Scholarship Program (Chakrabarti and Schwartz, 2013) threatened underperforming schools with student transfers and funding losses, leading administrators to reclassify low-performing students into exempt categories to artificially boost accountability scores. One exempted category was limited-English- proficient (LEP) students whose native language was not English and who enrolled in English-for-speakers- of-other-languages (ESOL) program for less

[4] As reported in an evaluation of another teacher incentive program, teachers believed that the bonus was "a reward for their usual efforts, not as an incentive for changing their behaviour." (Marsh et al., 2011)

than two years. Several other types of students were also excluded from the traditional test, including special education students. The study found that although the average test scores increased after the incentive was implemented, there was also a significant increase in the number of students in the excluded category.

Another key concern is the potential narrowing of the curriculum, as educators may feel pressured to focus primarily on tested subjects—such as math and reading—at the expense of non-tested areas like social studies, science, and the arts. This emphasis on standardized assessments may encourage "teaching to the test," potentially limiting students' broader educational development, where schools allocate disproportionate instructional time to tested subjects at the expense of arts, sciences, and critical thinking skills. Chakrabarti and Schwartz's research noted that schools under accountability pressures often abandon project-based learning and interdisciplinary units to focus on only memorization and test-taking strategies. For instance, teachers may drill students on past exam questions or formal classroom assessments to mimic standardized tests. While this approach may raise scores in the short term, it sacrifices creativity and fails to teach the analytical skills needed for higher education and workforce readiness. Such strategic reclassification did not reflect genuine improvements in teaching quality or student comprehension but instead manipulated statistical outcomes to meet thresholds.

This phenomenon is not isolated in Florida. Similarly, a study by Jacob (2005) in Chicago Public Schools revealed that high-stakes testing under NCLB led to increased rates of student retention in grade levels, particularly among minority and low-income students, as schools sought to avoid including low-performing students in accountability metrics. Particularly further analysis suggests that the observed achievement gains were driven by increases in test-specific strategy. Scholars also found that teachers responded strategically to the incentives along a variety of dimensions—by increasing special education placements, intentionally retaining students and substituting away from low-stakes subjects like science and social studies.

Another unintended consequence of TEI is its potential impact on teacher retention and collaboration. Performance-based compensation structures, while designed to reward high-performing educators, can create high-pressure environments that contribute to burnout and increased turnover, particularly in high-need schools where student growth is more challenging to achieve. Furthermore, such systems may inadvertently foster competition rather than collaboration among teachers, as financial incentives are often tied to individual rather than collective student success. This shift can undermine professional learning communities that thrive on shared knowledge and mentorship. These challenges highlight the complexities of implementing performance-based evaluation systems and suggest that while TEI may yield measurable improvements in student achievement, it also introduces systemic pressures that could counteract its intended benefits.

Additionally, the use of value-added models (VAM) to assess teacher effectiveness has been widely debated, as these models do not always account for external factors such as socioeconomic disparities, student health, and home environment, leading to inconsistent and sometimes inequitable evaluations.

Strategic reclassification of students, narrowing of the curriculum, and even outright cheating were among the behaviours incentivized by high stakes testing regimes. These practices are not the initial intention of designing the program. Our analysis suggested this incentive did not benefit all students, but there were sizable improvements in academic outcomes for some student sub-populations. Moving forward, policymakers must consider these unintended consequences and design accountability systems that prioritize support and capacity-building. Beyond instructional shifts, performance-based evaluation systems like TEI have also been associated with increased teacher turnover and decreased collaboration. Research on similar initiatives, such as Denver's ProComp, suggests that the pressure to achieve high ratings can contribute to stress and job dissatisfaction, leading some educators to leave the profession or transfer to schools where achieving performance benchmarks is perceived as more attainable. Moreover, competition for high ratings and financial incentives may weaken professional collaboration, as teachers become more focused on individual performance rather than collective improvement. These unintended consequences emphasize the need for careful policy design to ensure that evaluation systems support sustained instructional quality without undermining teacher morale or equity in educational opportunities.

Finally, academic performance is only one of the measures on which teachers are evaluated, and the application of multiple measures, whereby teachers might be evaluated using more and better (although also imperfect) educational measures (e.g., like the student survey instruments being used as part of the TEI, although we do not claim anything about the data derived via these instruments in terms of their reliability and validity), is intended to make improvements in teaching effectiveness both observable and (ideally) sustainable. Perhaps most evident here is that future research is needed in this area to explore the multiple dimensions of a program better and more fully such as TEI and its differential effects.

8 REFERENCES

American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 § 115 (2009). https://www.congress.gov/bill/111th-congress/house-bill/1

American Statistical Association (ASA). (2014). ASA statement on using value-added models for educational assessment. http://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf

Ashenfelter, O., Collins, W. J., & Yoon, A. (2006). Evaluating the role of Brown v. Board of Education in school equalization, desegregation, and the income of African Americans. American Law and Economics Review, 8(2), 213-248.

Atteberry, A., & LaCour, S. E. (2020). Putting performance pay to the test: Effects of Denver ProComp on the teacher workforce and student outcomes. Teachers College Record, 122(12), 1-48.

Bassok, D., Finch, J. E., Lee, R., Reardon, S. F., & Waldfogel, J. (2016). Socioeconomic gaps in early childhood experiences: 1998 to 2010. Aera Open, 2(3), 2332858416653924.

Boudreaux, M., & Faulkner, J. (2020). Examining student achievement, teacher effectiveness, and merit pay in a rural Tennessee school district. Voices of Reform, 3(2), 11-39.

Chakrabarti, R., & Schwartz, N. (2013). Unintended consequences of school accountability policies: Evidence from Florida and implications for New York. Economic Policy Review, 19(1). https://doi.org/10.2139/ssrn.2324231

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. Journal of Policy Analysis and Management, 34(2), 267–297. https://doi.org/10.1002/pam.21818

Dee, T., & Wyckoff, J. (2017). A lasting impact: High-stakes teacher evaluations drive student success in Washington, DC. Education Next, 17(4).

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. Psychological Review, 108(2), 346–369. https://doi.org/10.1037/0033-295X.108.2.346

Figlio, D. N., & Ladd, H. F. (2014). School accountability and student achievement. In Handbook of research in education finance and policy (pp. 194-210). Routledge.

Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from New York City Public Schools.

Journal of Labor Economics, 31(2), 373–407. https://doi.org/10.3386/w16850

Fulbeck, E. S. (2014). Teacher mobility and financial incentives: A descriptive analysis of Denver's ProComp.

Educational Evaluation and Policy Analysis, 36(1), 67-82.

Glazerman, S., McKie, A., & Carey, N. (2009). An evaluation of the Teacher Advancement Program (TAP) in Chicago: Year one impact report. Mathematica Policy Research. https://www.mathematica.org/publications/an-evaluation-of-the-teacher-advancement-program-tap-in-chicago-year-one-impact-report

Glazerman, S., & Saifullah, A. (2010). An evaluation of the Teacher Advancement Program (TAP) in Chicago: Year two impact report. Mathematica Policy Research. https://files.eric.ed.gov/fulltext/ED510712.pdf

Goldhaber, D., & Walch, J. (2012). Strategic pay reform: A student outcomes-based evaluation of Denver's ProComp teacher pay initiative. Economics of Education Review, 31(6), 1067-1083.

Goodman, S. F., & Turner, L. J. (2013). The design of teacher incentive pay and educational outcomes: Evidence from the New York City bonus program. Journal of Labor Economics, 31(2), 409–20. https://doi.org/10.1086/668676

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance?. Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management, 24(2), 297-327.

Hanushek, E.A., & Rivkin, S. (2007). Pay, working conditions, and teacher quality. The Future of Children. 17(1), 69-86. https://doi.org/10.1353/foc.2007.0002

Hanushek, E. A., Luo, J., Morgan, A. J., Nguyen, M., Ost, B., Rivkin, S. G., & Shakeel, A. (2023). The effects of comprehensive educator evaluation and pay reform on achievement (Working Paper No. 31073). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w31073/w31073.pdf

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. Journal of public Economics, 89(5-6), 761-796. https://doi.org/10.1016/j.jpubeco.2004.08.004

James, J., & Wyckoff, J. H. (2020). Teacher evaluation and teacher turnover in equilibrium: Evidence from DC public schools. AERA Open, 6(2), 2332858420932235.

Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. Educational researcher, 46(5), 234-249.

Koedel, C. (2014). Measuring Student Achievement in the Washington DC IMPACT Evaluation System: A Review.

Ladd, H. F. (2012). Education and poverty: Confronting the evidence. Journal of Policy Analysis and Management, 31(2), 203-227.

Maximum likelihood estimation | Stata. (n.d.). Retrieved August 20, 2023.

https://www.stata.com/features/overview/maximum-likelihood-estimation/

Marsh, J. A., Springer, M. F., McCaffrey, D. F., Yuan, K., Epstein, S., Koppich, J., Kalra, N., DiMartino, C., & Peng, A.

X. (2011). A big apple for educators: New York City's experiment with schoolwide performance bonuses. Final evaluation report. RAND Corporation. https://www.rand.org/pubs/monographs/MG1114.html

Moore Johnson, S. M., & Papay, J. P. (2009). Redesigning teacher pay: A system for the next generation of educators. Economic Policy Institute. https://www.epi.org/publication/book-redesigning_teacher_pay/

Murnane, R. J., & Olsen, R. J. (1989). The effects of salaries and opportunity costs on

duration in teaching: Evidence from Michigan. Review of Economics and Statistics, 71(2), 347–352. https://doi.org/10.2307/1926983

No Child Left Behind Act, Pub. L. No. 107-110, § 115 Stat. 1425. (2002). https://www.congress.gov/bill/107th-congress/house-bill/1

Pham, L., Nguyen, T., & Springer, M. (2020). Teacher Merit Pay: A Meta-Analysis. American Educational Research Journal, 58, 000283122090558. https://doi.org/10.3102/0002831220905580

Putnam, H., Ross, E., & Walsh, K. (2018). Making a Difference: Six Places Where Teacher Evaluation Systems Are Getting Results. National Council on Teacher Quality.

Reardon, S.F., Weathers, E.S., Fahle, E.M., Jang, H., & Kalogrides, D. (2022). Is Separate Still Unequal? New Evidence on School Segregation and Racial Academic Achievement Gaps.CEPA Working Paper No. 19-06.

Rivkin, S. G., Hanushek, E. A., & Kain J. F. (2005). Teachers, schools, and academic

achievement. Econometrica, 73(2), 417. https://doi.org/10.1111/j.1468-0262.2005.00584.x Springer, M. G., Lewis, J. L., Ehlert, M. W., Taylor, L. L., Lopez, O. S., Podgursky, M., & Peng, (2009). Governor's educator Excellence Grant (GEEG) Program: Year three evaluation report.

https://tea.texas.gov/system/files/GEEG_Y3_ExecSum.pdf

Springer, Matthew, Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J. R. Lockwood, Daniel McCaffrey, Matthew Pepper, and Brian Stecher. 2012a. Final report: Experimental evidence from the project on incentives in teaching. Nashville, TN: National Canter on Performance Incentives at Vanderbilt University.

Springer, Matthew, John Pane, Vi-Nhuan Le, Daniel McCaffrey, Susan Burns, Laura Hamilton, and Brian Stecher. 2012b. Team pay for performance: Experimental evidence from the Round Rock pilot project on team incentives. Educational Evaluation and Policy Analysis 34, no. 4:367–90.

Stata. (2022). Difference-in-differences (DID) and DDD models.

https://www.stata.com/features/overview/difference-in-differences-DID-DDD/ Teacher Excellence Initiative (n.d.). Resources. https://tei.dallasisd.org/home/resources/

Texas Education Agency. (n.d.). STAAR. Retrieved March 29, 2022. https://www.texasassessment.gov/staar.