Paying for Lack of Performance? Effects of Principal Incentive Pay on Students and Teachers*

Ashley Schwanebeck
University of Pennsylvania

December 15, 2025

Click for latest version

Abstract

School principals play a central role in managing teachers and shaping student learning, yet their compensation is typically based on experience rather than performance. This paper studies the effects of principal incentive pay within schools, and its implications for principal effort, teacher effectiveness, and student test scores. To do so, I leverage statewide implementation of incentive pay for principals in North Carolina (NC). Using administrative panel data in NC and Georgia and several difference-in-differences techniques, I estimate the causal effects of principal incentive pay on student test scores. I find that this particular incentive pay design induced significant declines in student outcomes, reducing reading and math test scores by 0.12-0.15 and 0.13-0.16 standard deviations (SD), respectively. To understand these counterintuitive results, I first estimate latent principal effort along four dimensions using teacher survey responses to analyze the effect of incentive pay on principals. Second, I estimate teacher effectiveness measured as test score value-added pre- and post-salary reform, and then I relate teacher effectiveness to principal effort and characteristics. I find that principal effort toward administrative tasks and teacher professional development declines after the incentive pay scheme was introduced by 0.05-0.10 SD, with no effect on instructional support. Teacher effectiveness declines by 0.05 SD in math and 0.01 SD in reading, with reduced principal effort toward administration driving this decline. My results suggest that incentive pay induced principals to reduce effort on margins to which teachers respond. Finally, I discuss how the design of the incentive scheme may have reduced its effectiveness.

JEL Codes: I20, I21, I28, I38, J31, J33

Keywords: Pay incentives, pay for performance, principal performance, leadership, student achievement

*I am deeply grateful to Petra Todd, Margaux Luflade, and Francesco Agostinelli for their invaluable support, guidance, and feedback. I thank Holger Sieg, Juan Pablo Atal, Juan Camilo Castillo, Shresth Garg, Jin Liu, Jere Behrman, Ioana Marinescu, and other seminar participants at the University of Pennsylvania for feedback and suggestions. This work is supported by the Population Studies Center at the University of Pennsylvania.
Contact information: Department of Economics, 133 S. 36th Street, Phila., PA 19104. Email: aschwane@sas.upenn.edu.

1 Introduction

Public school principals are pivotal in managing staff and improving student learning. Because principals supervise many classrooms, classic span-of-control theory implies their effort and quality scale across subordinates (Rosen, 1982). With national test scores at historic lows, incentives that motivate effective school leadership are increasingly important (National Center for Education Statistics, 2025). Principal salary in United States public schools typically follows a set schedule based on experience and education, rather than student outcomes. Unlike many private-sector managers, principals rarely receive performance-based pay, leaving little direct financial incentive to improve school performance. In an effort to raise student achievement, a limited number of school districts, cities, and states in the U.S. have incorporated performance pay for school principals, tying a fraction of salary to student test scores, such as New York City, Chicago, Pittsburgh, Houston, and all of North Carolina. Since performance-based pay is not widespread and is typically implemented along with teacher incentive pay, there is limited understanding of how it impacts school performance. Standard managerial theory suggests that linking pay to performance increases employee effort (Lazear, 2000), but this mechanism may not hold in a school setting, where principals face stronger management frictions than private-sector managers and student learning is harder to influence than firm output. Additionally, performance pay may even reduce effort depending on the structure of the incentive program (Ariely, Gneezy, Loewenstein, & Mazar, 2009).

Evidence from teacher performance pay is also mixed. Cohodes, Eren, and Ozturk (2023) find that teacher performance pay improves educational attainment and reduces criminal activity, while Brehm, Imberman, and Lovenheim (2017) finds that only teachers near a merit-based cutoff adjust effort, with no resulting gains in student outcomes. If principal pay incentives function similarly, they may encourage strategic behavior rather than improve student outcomes. For example, principals closer to salary thresholds will have a higher return to effort than principals far from cutoffs. Principals far from cutoffs may then experience a reduction in intrinsic motivation if they perceive the incentive as infeasible to reach. Moreover, the effectiveness of performance pay for principals depends on how teachers respond to principals. Teachers interact with students daily, and may respond heterogeneously to principal effort. While principals may successfully encourage teachers to increase their effort and improve student achievement, micromanagement or excessive administrative burden may hinder teacher effectiveness.

In this paper I study three questions: (1) What are the effects of introducing performance-based pay for public school principals on student outcomes, teacher effectiveness, and principal effort? (2) Through which channels do principals influence teacher behaviors? (3) Are effects heterogeneous across different types of schools and principal characteristics? The setting of my analysis is North Carolina, which implemented statewide principal performance pay in July 2017. Prior to the reform, principals' salary was determined by years of experience, degree type, and the number of teachers managed; after the reform, salary is determined by schedules based solely on student growth performance and student membership. I answer these questions in two steps. First, I estimate the impact of principal incentive pay on student test scores using a difference-in-differences design that compares North Carolina to a nearby comparison state, Georgia, which has similar testing regimes and no other major K–12 policy changes during the sample period. To address identification concerns about using Georgia as a control, I also estimate a matching difference-in-differences model and a synthetic control difference-in-differences (DID) model, both of which further balance pretrends in test scores. Across matching and synthetic DID specifications, I find that the principal incentive pay scheme reduced math scores by 0.12-0.15 standard deviations (SD) and reading scores by about 0.13–0.16 SD.

Second, I turn to within-state analysis in North Carolina to understand why the policy reduced student achievement. I use detailed administrative data and teacher responses to the Teacher Working Conditions Survey to study how the reform affected principals and teachers. NCERDC contains detailed administrative data on student test scores, employee salary, degree, and experience, and demographic information on students, teachers, and principals, and the TWC contains anonymized teacher survey responses on school leadership and climate. I first use teacher survey responses to estimate latent principal effort along four dimensions: administrative tasks, school climate, teacher professional development, and the instructional program. Next, I estimate teacher effectiveness, measured by value-added in test scores, before and after the salary reform and relate teacher effectiveness to both principal effort and principal characteristics. I then construct a principal exposure index that captures, for each principal, the salary gain implied by the switch from the experience-based schedule to the performance-based schedule, and use this index, together with principal and school characteristics, to study heterogeneity in both principal effort and teacher value-added.

Three main findings emerge. First, across standard, matching, and synthetic difference-in-differences designs, the principal incentive pay reform consistently reduces student achievement, with sizable negative effects on both math and reading scores, as discussed above. Moreover, North Carolina schools with positive estimated treatment effects relative to Georgia are led by principals who exert roughly twice as much latent effort as principals in schools with negative treatment effects. Second, within North Carolina, principal effort and teacher effectiveness both decline on average after the reform. Principal effort toward administrative tasks, school climate, and teacher professional development falls by roughly 0.10, 0.08, and 0.05 SD, respectively, with little change in effort toward the instructional program. Teacher value-added falls by roughly 0.05 SD in math and 0.01 SD in reading. In regressions of teacher value-added on latent principal effort, I find teachers are most responsive to principal effort that reduces administrative burdens and supports professional development, with smaller or negligible responses to instructional program effort and school climate. Third, heterogeneity analyses of teacher value-added reveal that these declines are similar in magnitude for low- and high-exposure principals, but are mitigated in schools with higher shares of economically disadvantaged students and high-performing students. Teachers with high-achieving students faced lower reductions in effectiveness, and teachers with high shares of students classified as economically disadvantaged experienced improved effectiveness. Finally, in my analysis relating principal effort to teacher effectiveness, I find that teachers are most receptive to principal effort exerted on providing professional development, and administrative tasks of teachers, which is where effort reduced the most. My results suggest that incentive pay induced principals to exert effort toward margins to which teachers are less responsive.

The finding that incentive pay reduces principal effort and student achievement is consistent with how this particular scheme reshaped the returns to effort. The reform simultaneously delivered large, mostly unconditional salary increases to many principals and introduced a new bonus structure that is both high stakes and difficult to influence in the short run. For principals who already faced heavy workloads and limited control over measured growth, the new system can operate more like a pay raise plus insurance against bad test-score years than a clear, attainable reward for extra effort. The reform introduced an average $10,000 salary increase across all principals, regardless of student growth. Additionally, principals who would be paid less under the new policy were given a hold harmless salary, which prevented them against salary loss. Standard labor-leisure intuition then implies that the income and insurance components of the reform can reduce optimal effort when the marginal payoff to additional effort is small or uncertain. In my data, this shows up as declines in principal effort along the dimensions that matter most for teachers' work– reducing administrative burdens, supporting school climate, and providing professional development.

This research contributes to several strands of literature. First, this study relates to literature analyzing the effect of principals on student outcomes, namely graduation rates and test scores. Branch, Hanushek, and Rivkin (2012), Dhuey and Smith (2014, 2018), and Chiang, Lipscomb, and Gill (2016) estimate principal value-added (VA) from fixed-effects OLS specifications and consistently estimate principal VA to be 0.05-0.2 standard deviations. Coelli and Green (2012), Bluestein and Goldschmidt (2021), and Miller (2013) estimate the effects of principal turnover and find that student performance declines immediately following turnover, but reverts to its prior level a few years later. Several papers study the effects of principal management practices on student outcomes, and find that principal training raises student achievement, especially in high socioeconomic status schools (Bloom, Lemos, Sadun, & Reenen, 2015; Di Liberto, Giua, Schivardi, Sideri, & Sulis, 2023; Roland G. Fryer, 2017). I contribute to this literature by exploring a novel channel for how principals affect students; rather than directly affecting student outcomes, principal effort affects teacher effectiveness, which then translates to changes in student test scores.

Next, I contribute to literature on manager effects on firm efficiency and production. Most seminal in this strand are Bloom, Eifert, Mahajan, McKenzie, and Roberts (2012); Lucas (1978); Rosen (1982), who develop hierarchy span-of-control models and find that better managers run larger firms, and managerial quality scales aggregate output. Additionally, incentive pay in the private sector elicits higher managerial effort and attracts higher talent (Bandiera, Prat, Guiso, & Sadun, 2011; Wu, 2017). I contribute to this literature by studying the efficacy of principal managerial effects on the school system in the presence of incentive pay.

Finally, this study contributes to literature studying incentive pay in public schools. There is a robust literature studying effects of teacher incentive pay on student achievement, which finds mixed effects. While in some settings, performance-based incentives for teachers improve student achievement (Atkinson et al., 2009; Dee & Wyckoff, 2015; Figlio & Kenny, 2007; Glewwe, Ilias, & Kremer, 2010; Lavy, 2009; Muralidharan & Sundararaman, 2011), other work has found teacher pay incentives induced strategic effort decisions without raising test scores Brehm et al. (2017); Fryer (2013); Imberman and Lovenheim (2015); Jones and Hartney (2017). To my knowledge, the only other work studying the effects of North Carolina's principal incentive pay reform is Westall (2022), where he uses a difference-in-differences approach to study the labor decisions of principals before and after the implementation of incentive pay in North Carolina. He finds that in the year following the policy, effective principals moved to different schools, especially low-performing ones, and ineffective principals were more likely to leave the profession. I contribute to this literature as the first work to estimate the effects of a statewide reform introducing incentive pay for principals on teacher effectiveness and student outcomes.

The rest of my paper is structured as follows. Section 2 describes the data used in my analysis and the institutional details of North Carolina's reform to principal salary. Sections 3 and 4 describe the empirical framework. Section 5 discusses results, and Section 6 concludes.

2 Data and Policy Description

2.1 Data

I use rich administrative data from the North Carolina Education Research Data Center (NCERDC), which records student test scores, class assignments, and demographic information as well as teacher and principal salary, experience, degree type, and demographics. For across-state analysis, I supplement this data with test scores and student demographic information from Georgia, a nearby control state. Georgia outcomes come from the Georgia Milestones End-of-Grade (EOG) assessments.[1] I additionally supplement data in all three states with school characteristics from the National Center for Education Statistics' Common Core of Data (CCD). I focus my analysis on the 2014-15 through 2017-18 academic years, a period with consistent test versions and scales in both states. Table A1 describes characteristics of students and schools in each state during the sample period. NC schools have higher shares of White students and Title I schools and lower shares of Black students, a lower share of urban schools, and smaller tested cohorts. Specifically, NC has a 9 pp higher hare of White students (49% vs. 40%) and an 11 pp lower share of Black students (26% vs. 37%). Hispanic shares are similar (18% vs. 16%), and English learners are slightly more prevalent in GA (8% vs. 6%). Economic disadvantage status (EDS) shares are similar (68% NC vs. 66% GA), as is disability prevalence (12% in both). GA has a higher share of urban schools (59% vs. 49%), whereas Title I schools are more common in NC (80% vs. 73%). The average tested cohort size per school-grade is larger in GA (210) than in NC (170).

For within North Carolina analysis, I conduct analysis at the student, teacher, and principal level using data from NCERDC. I additionally utilize teacher responses to the North Carolina

[1] I perform a robustness check using school-grade-subject level test score data in South Carolina from SC READY, the South Carolina College- and Career-Ready Assessments in ELA and mathematics.

Teacher Working Conditions Survey (TWC), a biennial survey given to teachers on perceptions of time use, resources, leadership, professional development, and school climate. Teacher responses to the TWC are anonymous and will be utilized to construct latent measures of principal effort that are free of reporting bias.[2] Specific questions used to construct principal effort will be selected from those in Table B3.

I focus on principals and teachers in regular elementary schools, and students in grades 3-5 during the 2012-13 through 2018-2018 academic years.[3] I restrict my sample to schools with principals working at least 9 months of the academic year if they are the only principal listed and 8 months of the academic year if there are multiple principals listed within a school year.[4]

I use data on students and teachers to estimate teacher effectiveness (value-added). I restrict my sample of students to those in elementary school that have a valid math and/or reading score. I link students to classrooms using unique classroom identifiers, and use regular expressions to identify whether a course is a math or reading course a student is administered an end-of-grade exam in.[5] Following Chetty, Friedman, and Rockoff (2014a, 2014b) I consider classrooms with at least 10 students that have nonmissing test scores. Finally, I consider teachers, principals, and students who are observed in at least 2016-17 and 2017-18, one year prior to and after policy implementation.

2.2 Description of Policy

School educator salary in North Carolina follows robust pay schedules set by the state. The state specifies a separate pay schedule for each employee type, e.g.- teachers, principals, assistant principals, instructional support, school psychologists, and school administrators. Local salary adjustments are given according to the discretion of the school district an employee is located in. On July 1st, 2017, North Carolina signed

Throughout the 2016, a special legislative committee on school-based administrative pay con-

[2]Each survey year, the North Carolina Department of Public Instruction publishes aggregate survey results at each school separately for teachers and principals. Since a majority of schools have one principal, there is incentive for principals to rate themselves more highly.

[3]I follow the classification of regular schools defined by the National Center for Education Statistics in the Common Core of Data, which defines a regular school as "a public elementary or secondary school that does not have a primary focus on special education, vocational/technical education, or other alternative programs" (National Center for Education Statistics, 2025).

[4]Occasionally, in school transition years, two individuals will be listed as principals, the exiting and entering principals. When this occurs, I keep the incoming principal if the old one was salaried for 1 month. Principals are paid for 12 months during a school year and are student- and teacher-facing 10 months of the year.

[5]For example, courses titled "Math 3","Math gr 3", and "grade three math" are included, but courses titled "math Olympics" or "math enrichment" are dropped.

Table 1: Principal Characteristics

	Pre-Policy			Post-Policy		
	All	Bottom 25%	Top 25%	All	Bottom 25%	Top 25%
Total Experience	21.03	23.99	16.54	20.88	25.08	18.97
	(6.85)	(5.66)	(3.90)	(5.94)	(6.24)	(4.07)
Experience as Principal	6.18	6.81	4.60	6.24	7.53	6.84
	(4.43)	(4.84)	(2.95	(4.36)	(5.14)	(3.05)
Annual Nominal Salary	61,707	64,113	58,526	72,183	69,960	77,938
	(6,830)	(6,518)	(3,732)	(6,972)	(7,446)	(5,952)
Annual Real Salary	62,761	64,964	59,597	69,990	67,934	75,663
	(6,985)	(6,598)	(3,764)	(6,546)	(6,986)	(5,469)
Hold Harmless Suppl.	0.13	0.17	0.13	0.23	0.46	0.02
	(0.33)	(0.37)	(0.34)	(0.42)	(0.50)	(0.38)
Growth Not Met	0.24	0.33	0.10	0.25	0.23	0.19
	(0.43)	(0.47)	(0.31)	(0.43)	(0.42)	(0.39)
Growth Met	0.49	0.47	0.44	0.48	0.50	0.45
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Growth Exceeded	0.28	0.20	0.45	0.27	0.27	0.36
	(0.45)	(0.40)	(0.48)	(0.44)	(0.45)	(0.45)
N	8,480	919	1,060	3,357	538	512

Observations consider principal-year units in K-12 during the 2013-14 through 2017-18 academic years. Column (1) presents averages during the entire period, while column (2) and column (3) report averages pre- and post-policy. Standard errors of the mean are reported in parentheses.

vened to discuss restructuring principal compensation to depend on performance. In December 2016, this group submitted their findings and suggestions to the Regular Session of the 2017 General Assembly (North Carolina General Assembly, 2016). The North Carolina General Assembly officially passed performance pay into law on June 28, 2017, which took effect July 1st, 2017, right before the start of the 2018 academic year (North Carolina General Assembly, 2016). Prior to the implementation of principal performance pay, principal salary was set according to experience, number of teachers managed, and degree type. Figure A1 provides an example of a salary schedule for a principal who manages 22-32 teachers in the 2017 academic year. An additional year of experience or jump to the next bin of teachers managed corresponds to a 1.3-2% salary increase. Having an advanced degree raises salary by roughly 3% for the lowest paid principal and 1.5% for the highest paid. Having a doctorate degree doubles this raise, for a salary increase of 3-5.6%.

After the implementation of performance pay, principal salaries are determined by school growth status and average daily membership, with the former having a larger impact on salary.[6]

[6]Average daily membership (ADM) measures average enrollment of students not in violation. It is calculated as the sum of non-violating membership days across enrolled students divided by days in a school month. Principal salary

Each increase in school growth status raises a principal's salary 10% from the base salary, and each increase in student membership levels increases salary 5%. Principals are also given small bonuses for being in the top performing schools. For example, principals in the top 50% performing schools are given a $1,000 annual bonus (1.4% of average base salary), and principals in the top 5% performing schools are given a $5,000 bonus, which is a 7% increase from average salary.

School growth status is determined by the value of a school's growth indexes, which are calculated using the Education Visualization and Analytics Solution (EVAAS), a value-added measure estimated through a fixed-effects model based on student standardized test scores, ACT/SAT scores, and AP scores. North Carolina does not publicly provide the full model they estimate, but document details on standard error calculations and which students to include in the estimation (SAS Institute Inc., 2023). If a school's growth index is more than 2 standard errors below expected growth (0), it does not meet expected growth and falls into the base salary category for principals. Growth within two standard errors of expected growth is categorized as "Growth Met," while growth exceeding two standard errors is classified as "Growth Exceeded." The two highest growth indexes of the three prior academic years determine principal salary according to Figure A3. For example, a principal with growth indexes of 0.1 in 2011, 2.5 in 2012, and -1 in 2013 will be categorized as having growth met and will receive a 10% bonus above base salary. Salaries are increased by 10% for principals in schools receiving Met versus Base and 20% in schools receiving Exceeds versus Base. Every step increase in ADM further raises salary an additional 5% from base.

Table 1 describes principal experience, salary, and growth expectations from 2014-2019, before and after the implementation of performance pay. Columns 1 and 3 show aggregate averages for 2014-2017 and 2018-2019, respectively. Columns 2 and 4 show averages for principals in the bottom 25th percentile of real year-over-year salary change based in 2018, while columns 3 and 5 report averages for principals in the 75th-100th percentile of year-over-year salary change in 2018.

Table 2 shows the average growth status transition from 2014 through 2017 to 2018 through 2019. Roughly 50% of principals meet growth expectations before and after the implementation of performance pay, with 23% of principals not meeting expectations and 27-28% exceeding expectations. After introducing performance pay, there are compositional shifts in the transition of

is determined from the highest ADM measure of the school years 1st and 2nd month, both of which must be 20 days long. Students are considered in violation after accumulating 10 consecutive unlawful absences. Part-time students contribute fractionally to ADM.

Table 2: Principal Characteristics

	Pre-policy				Post-policy			
	t				t			
$t-1$	Not Met	Met	Exceeded	Total	Not Met	Met	Exceeded	Total
Not Met	37.25	45.81	16.94	100	37.23	45.02	17.75	100
Met	22.63	53.73	23.64	100	21.59	54.90	23.51	100
Exceeded	11.57	47.48	40.94	100	12.36	42.52	45.12	100
Total	22.89	50.12	26.99	100	22.71	49.30	27.99	100.00

Observations consider principal-year units in K-12 during the 2013-14 through 2017-18 academic years. Column (1) presents averages during the entire period, while column (2) and column (3) report averages pre- and post-policy. Standard errors of the mean are reported in parentheses.

pay for principal previously exceeding expectations (row 3 of Table 2). For principals previously exceeding expectations in the year prior, post policy, principals exceeding expectations increases 4 percentage points and principals meeting expectations declines 5 percentage points. There is a small, 1 percentage point increase in principals not meeting expectations of those previously exceeding expectations. This suggests that the policy induced minimal changes in growth expectation status for all principals except those that were previously exceeding expectations, who now continue to exceed expectations at a higher frequency.

3 Empirical Model

3.1 Difference in Differences Model

The main challenge for identifying the causal effect of principal performance pay on student achievement is that the policy was implemented statewide in North Carolina beginning in the 2017–18 school year, rendering no untreated comparison group within the state. I therefore construct a counterfactual for North Carolina using Georgia as a control state and employ a difference-in-differences (DiD) strategy.

Georgia is a natural comparison for several reasons. First, both states administered stable statewide standardized exams in grades 3–5 in mathematics and reading from 2014–15 through 2018–19. Second, during this period Georgia did not undertake major reforms to principal pay or other statewide policies that directly link school-level pay to test performance. Third, the two systems are similar in terms of grade structure, test subjects, and the organization of traditional public schools, which supports the use of Georgia as a counterfactual for aggregate trends in student achievement. I discuss remaining concerns about this choice and how I address them in Section 3.1.1.

Georgia makes a natural control group for a variety of reasons. Georgia administered consistent statewide standardized testing during the 2014-15 through 2018-19 academic years, and did not have any major school-policy changes during this period.[7] I address potential concerns of using Georgia as a control state below.

I estimate the two-way fixed-effects difference-in-differences model in Equation 1 separately for math and reading, comparing student test scores within school, grade, and subject across North Carolina and Georgia.

$$y_{gst} = \alpha_0 + \beta\, NC_s \times Post_t + x'_{gst}\alpha_1 + \gamma_t + \delta_{gs} + \varepsilon_{gst} \tag{1}$$

y_{gst} is the average math or reading standardized score of students in grade g in school s in year t. To ensure comparability across states, subjects, and grades, I standardize test scores at the state-grade-subject level using 2017 means and standard deviations. Anchoring the moments in 2017 avoids scaling out the policy's impact and preserves the policy-induced level shifts. NC_s is an indicator variable taking a value of 1 if school s is observed in North Carolina and $Post_t$ is an indicator taking a value of 1 in policy years. x'_{st} is a vector of school characteristics including Title I status, an indicator for being located in an urban area, and shares of students who are economically disadvantaged, disabled, white, black, Hispanic, female, and have limited English proficiency. γ_t and δ_{sg} are time and unit (school-grade) fixed effects, respectively. The coefficient β is the average treatment on the treated (ATT) estimate of principal incentive pay on school-grade test score averages.

3.1.1 Identification

Causal estimation of β requires that test score trends in each state are constant before the policy began in 2018. I test this assumption in two ways: first, a placebo test treating 2015 and 2016 as the policy implementation years, and second, estimating and plotting an event study model.

[7]Georgia redesigned its College and Career Ready Performance Index (CCRPI) accountability system in 2018, changing how school ratings are calculated without altering the underlying Georgia Milestones tests. Because CCRPI does not directly determine state funding or statewide employee salaries (and testing content/administration in grades 3-5 remained stable), I do not expect large, mechanical shifts in test scores attributable to the redesign.

I estimate the following event study specification

$$y_{gst} \;=\; \delta_{gs} \;+\; \tau_t \;+\; \sum_{r \neq 2016} \beta_r \left(\text{NC}_s \cdot \mathbf{1}\{r\}\right) \;+\; x'_{gst}\gamma \;+\; \epsilon_{gst} \tag{2}$$

I condition on the same covariates as Equation 1. To account for a possible anticipation effect, I omit 2016 as the reference year to account for the possibility of an anticipation effect. Based on the event study plot in Figure 2, there is a dip in test scores in North Carolina compared to Georgia in 2017, one year before policy implementation. This is indicative of an anticipation effect in North Carolina or a sudden rise in test scores in Georgia. Coupled with the decline in relative scores in 2017, North Carolina and Georgia have baseline differences in student demographic compositions, school sizes, and share of urban schools. Before attributing this pre-2018 dip to anticipation in North Carolina or to unobserved shocks in Georgia that would threaten causal interpretation, I re-estimate the policy effects using a propensity score matching DID and synthetic control DID, which help align pre-policy covariates and trends between North Carolina and Georgia.

To account for potential pretrends in test scores across North Carolina and Georgia, I additionally estimate a synthetic control difference-in-differences model as proposed in Arkhangelsky, Athey, Hirshberg, Imbens, and Wager (2021).

3.1.2 Difference-in-differences Results

Table 3 presents the estimated average treatment of principal incentive pay on North Carolina schools from the difference-in-differences model of Equation 1. Compared to Georgia, principal incentive pay in North Carolina induced a large decline in test scores, with math dropping 0.18 standard deviations (SD). Figure 2 plots the event study coefficients estimated from Equation 2. While there is no difference in test scores conditional on observables between each state in 2014-15 and 2015-16, there is a decline in scores in the year prior to the policy, 2016-17. With the policy being enacted into law June 28, 2017, right before the start of the 2017-18 academic year, it is unlikely this decline in differences is due to an anticipation effect.[8] Based on state-year level unconditional, aggregate trends, Georgia had a slight increase in test scores over time, which could bias the policy treatment effect downward. To account for this trend, I perform two additional

[8] A House Select Committee on school-based administrator pay first convened to discuss changes to principal compensation October 24, 2016, fall semester of the 2016-17 academic year. While principals may have heard rumors of these discussions, the exact pay re-structure was not formalized. In fact, there were numerous pay structures suggested, some of which were later abandoned.

Table 3: Difference in Difference Estimates: Elementary School Students

	Math	Reading
DiD	-0.176***	-0.182***
	(0.023)	(0.018)
School-grade FE	Y	Y
Year FE	Y	Y
Mean	-0.168	-0.170
SD	0.924	0.940
N	28,402	28,402

*** p<0.01, ** p<0.05, * p<0.1. Observations consider school-grade-year units in grades 3-5 during the 2014-15 through 2017-18 academic years. Test scores are standardized using 2016-17 subject-grade means and standard deviations. Standard errors are clustered at the district level. Controls include school-grade-year level shares of ethnicity, race, sex, economic disadvantage status, limited English proficiency status, bins for the number of students who took the exam, and school indicators for being located in an urban region and Title I status.

analyses. First, I implement a synthetic difference-in-differences (SDID) design that constructs a "synthetic" control group from Georgia that more similarly matches pre-treatment trends in North Carolina. Second, I estimate a matching difference-in-differences (MDID) model that balances pre-treatment observable characteristics between the two states. Implementing each of these models raises confidence in the parallel trends assumption being satisfied. While the SDID re-weights observations to match pre-treatment test scores, the MDID controls for compositional differences in student demographics and school characteristics across the two states, which may contribute to differing trends in test scores.

3.2 Synthetic Control Difference-in-Differences

Following Arkhangelsky et al. (2021), I implement a synthetic difference-in-differences (SDID) methodology, which combines the conventional two-way fixed effects difference-in-differences model with the synthetic control model. The SDID estimator first constructs a weighted synthetic control for treated units by choosing weights on control school-grades and on pre-policy years so that the synthetic control (Georgia school-grades) closely follows the pre-policy test score trajectory of North Carolina school-grades. Given these weights, I then estimate a weighted version of Equation (1).

More formally, SDID constructs unit weights $\hat{\omega}_{gs}$ and time weights $\hat{\lambda}_t$ from pre-policy residualized test scores y_{gst}^{res}.[9]

The treatment effect is then computed as a difference-in-differences in weighted mean residualized outcomes between North Carolina and the synthetic control, Georgia. More specifically, the SDID estimate of the average treatment effect, $\hat{\beta}^{\text{SDID}}$, is obtained from the objective function of Equation (3).

$$(\hat{\beta}^{\text{SDID}}, \hat{\alpha}_0, \hat{\delta}, \hat{\gamma}) = \arg \min_{\beta, \alpha_0, \delta, \gamma} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{sgt}^{\text{res}} - \alpha_0 - \delta_{sg} - \gamma_t - \beta W_{it} \right)^2 \hat{\omega}_i \hat{\lambda}_t, \tag{3}$$

where N is the number of school-grades in the balanced panel, T is the number of time periods, and $W_{st} \in \{0, 1\}$ denotes the treatment indicator, $NC_s \times Post_t$. Each school s receives weight $\hat{\omega}_s$ and each year t receives weight $\hat{\lambda}_t$.

3.3 Propensity Score Matching

Next, I implement a difference-in-differences matching strategy following Heckman, Ichimura, and Todd (1997, 1998); Smith and Todd (2005), where I condition on a school's propensity to be treated given their set of observable covariates. Define the propensity score of school s to be in North Carolina as $p(x_s)$ where

$$p(x_s) = \Pr(D_s = 1 \mid x_s) \tag{4}$$

where $D_s \in \{0, 1\}$ indicates treatment status $(1 = NC, 0 = GA)$, and x_s is a vector of pre-treatment covariates in 2016-17 as defined in Equation 1. $p(x_s)$ is estimated via logistic regression. Next, define Δy_{sD} as the change in mean test scores with

$$\Delta y_{sD} = y_{sD,2018} - y_{sD,2017} \tag{5}$$

The DID matching estimator is then

$$\widehat{ATT} = \frac{1}{N_1} \sum_{i \in I_1 \cap S_D} \left[\Delta y_{1i} - \sum_{j \in I_0 \cap S_D} w_{ij} \, \Delta y_{0j} \right] \tag{6}$$

[9] y_{gst}^{res} is the residualized school-grade test score average of grade g in school s after conditioning on the vector of covariates x_{gst} described in Section 3. Formally,

$$y_{gst}^{\text{res}} = y_{gst} - x_{gst}' \hat{\alpha},$$

where the vector $\hat{\alpha}$ is obtained from a regression of test scores on all covariates using pre-policy observations.

where I_1 and I_0 index school-grades in North Carolina and Georgia, respectively, S_D is the region of common support, and N_1 is the total number of school-grades such that $N_1 = |I_1 \cap S_D|$. Weights, w_{ij}, are created from $p(X)$ using the Gaussian kernel.

Identification of $\widehat{ATT}$ requires two assumptions: strong ignorability and overlapping support. Strong ignorability is satisfied given the change in test scores pre- and post-policy is independent of treatment status conditional on observable covariates. Formally, $\Delta y_{s0} \perp\!\!\!\perp D_s \mid X_s$. This assumption would be violated if, for example, parents in Georgia decide to relocate to North Carolina to enjoy the perceived benefit of principal performance pay. Since this would be largely infeasible, especially in the short run, this assumption is likely verified.

Next, common support requires that for every school in North Carolina with pre-policy covariates $X = x$, there exist schools in Georgia with positive probability of $X = x$ (i.e., $0 < p(x) < 1$). To satisfy this assumption, I impose common support when estimating $p(X_s)$, and thus the estimated $\widehat{ATT}$ is for treated North Carolina schools on common support. Visually, based on Figure A4, there is a large overlap in support, and this assumption is satisfied. Additionally, Table 13 provides evidence of covariate balance once conditioning on the propensity score.

3.3.1 SDID and MDID Results

Tables 4 and 5 display the average treatment effect (ATT) estimate from the propensity score matching model of Equation 6 and the synthetic DID, respectively. Under these models, I find a smaller, but still sizable negative effect, with the incentive pay reducing test scores in math and reading by 0.12-0.15 SD.

Table 4: Matching Difference in Difference Estimates: Elementary School Students

	Math		Reading	
	(1)	(2)	(3)	(4)
ATT	-0.124***	-0.131***	-0.130***	-0.134***
	(0.023)	(0.009)	(0.022)	(0.008)
Mean	-0.123	-0.053	-0.107	-0.047
SD	0.936	0.933	0.935	0.940
Weighted	N	Y	N	Y
N Treated	3,654	3,654	3,646	3,646
N Untreated	3,253	3,253	3,274	3,274
N	6,907	6,907	6,920	6,920

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Observations consider school-grade-year units in grades 6-8 during the 2014-15 through 2017-18 academic years. Test scores are standardized using 2016-17 subject-grade means and standard deviations. Standard errors are clustered at the district level. Controls include school-grade-year level shares of ethnicity, race, sex, economic disadvantage status, limited English proficiency status, bins for the number of students who took the exam, and school indicators for being located in an urban region and Title I status. Results in columns (1) and (2) are unweighted, and columns (3) and (4) are weighted by the number of students who took each exam.

Table 5: Synthetic Control Difference in Difference Estimates: Elementary School Students

	Math		Reading	
	(1)	(2)	(3)	(4)
ATT	-0.164***	-0.147***	-0.177***	-0.160***
	(0.012)	(0.014)	(0.010)	(0.012)
School-grade FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Covariates Included	N	Y	N	Y
Mean	-0.174	-0.174	-0.175	-0.175
SD	0.923	0.923	0.927	0.927
N	27,836	27,836	27,836	27,836

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Observations consider school-grade-year units in grades 3-5 during the 2014-15 through 2017-18 academic years. Test scores are standardized using 2016-17 subject-grade means and standard deviations. Bootstrapped standard errors using 500 replications are reported in parentheses. Controls include school-grade-year level shares of ethnicity, race, sex, economic disadvantage status, limited English proficiency status, bins for the number of students who took the exam, and school indicators for being located in an urban region and Title I status.

4 Decomposing Channels: Within North Carolina Analysis

With incentive pay ineffective in raising test scores, I now seek to first understand whether principals more exposed to incentive pay (in terms of salary gains) responded the same as those less exposed, and second, decompose channels of the incentive to understand the cause of the test score declines. Within a school, principals typically do not interact directly with students, aside from disciplining, but rather manage teachers. In a three-step channel decomposition analysis, I first estimate principal effort along four dimensions over time, second, estimate teacher VA pre- and post-policy, and third, relate principal effort and teacher VA. Additionally, I break down these effects by various heterogeneity margins.

4.1 Exposure Analysis

Principals may respond heterogeneously to performance pay depending on their potential salary gain from 2016-17 to 2017-18. The reform to salary reindexed pay from experience, number of teachers, and degree type to school performance (growth bins) and number of students. Under the pre-policy salary schedule, principals are paid about 1.3-2% for each additional year of experience and every additional 10 teachers managed.[10] Under policy years, principals receive an additional 5% above base for each step in student body size and 10% on top of base for each step in school growth status. For each principal k, the schedule-induced salary change Δw_k in 2017-18 summarizes the stakes of the new system. After 2017-18, the expected marginal payoff to extra effort is higher when Δw_k is larger: the dollar gains from moving into (or remaining in) a higher growth bin are larger than when Δw_k is low.By contrast, principals whose pre-policy pay was already high, and thus have a lower Δw_k, face weaker marginal incentives to exert extra effort and will have lower gains (or larger declines) in test scores. To test this empirically, I create an exposure index, W_k, for each principal using imputed 2017-18 salary under the old schedule, $w^O_{ks,2018}$, and the new schedule, $w^N_{ks,2018}$, and estimate the effect on test scores of having a principal with high exposure post-policy.

I impute 2017-18 salary under the 2016-17 experience-based schedule using 2017-18 principal experience, degree type, and number of teachers in the school. I then increase this value by 1.5%

[10] Principals need to pass an experience threshold to earn higher salary, which increases with the number of teachers in a school. For example, in 2016-17, a principal managing 1-10 teachers begins earning a higher salary after the 13th year of experience, whereas a principal managing over 100 teachers will not earn a higher salary until reaching 24 years of experience. Experience includes total years working as an educator.

to account for typical inflation adjustments to salary.[11] I impute 2017-18 salary according to the salary schedule in Figure A2. The median of a principal's prior 3 growth statuses (if available) determines which growth status they are paid according to. For example, in 2017-18 a principal with growth statuses "Not Met" in 2014-15, "Met" in 2015-16, and "Exceeded" in 2016-17 will fall under the "Met Growth" category for 2017-18 salary determinations and will be awarded 10% above "Base" salary. To test the validity of these imputations as approximations for actual salary, I regress 2016-17 actual salary on 2016-17 imputed salary and 2017-18 actual salary on 2017-18 imputed salary. I find both coefficients are statistically significant at the 1% level, with confidence intervals containing 1, so my imputed salary is a sufficient statistic for actual salary.

For each principal k in 2017-18, define raw exposure $\Delta \mathrm{w}_k = w^O_{ks,2018} - w^N_{ks,2018}$ and standardized exposure index, W_k, as follows:

$$W_k = \frac{\Delta \mathrm{Sal}_{k,2018} - \mu_{\Delta w}}{\sigma_{\Delta w}} \tag{7}$$

where $\mu_{\Delta w}$ and $\sigma_{\Delta w}$ are the mean and standard deviation of raw exposure across all principals in 2017-18, respectively. I assume exposure is constant across time for a given principal, and estimate the specification in Equation 8.

$$y_{igkst} = \theta_0 + \theta_1 \left(W_k \times \mathrm{Post}_t \right) + x'_{it}\theta_2 + z'_{st}\theta_2 + \mu_i + \lambda_g + \delta_s + \varepsilon_{igkst} \tag{8}$$

I estimate Equation 8 separately for reading and math outcomes. The vector x_{it} contains student-level time-varying characteristics, including indicators for repeating a grade, being an English learner, disabled, and a new entrant to school s. The vector z_{st} contains school-level time-varying characteristics including indicators for being in an urban location or Title I eligible, as well as average characteristics across students — all those contained in x_{it} in addition to share economically disadvantaged, White, Black, and Hispanic. The inclusion of student fixed effects, μ_i, controls for time-invariant student characteristics, such as baseline ability or motivation. Grade-year fixed effects, λ_g, absorb statewide grade-specific shocks. Finally, school fixed effects, δ_s, control for school-level compositional differences, such as baseline achievement, persistent peer effects, and neighborhood socioeconomic status.[12] Standard errors are clustered at the district level.

[11]From 2015-16 to 2016-17 principal salaries increased 1.5% at each experience and teacher quantity level to adjust for inflation.

[12]For robustness, I include a specification that removes δ_s as schools and principals (and thus principal exposure) are likely colinear.

4.2 Latent Principal Effort

I measure principal effort along four dimensions: administrative tasks, professional development, instructional leadership, and school climate.[13] Principal effort toward administrative work includes tasks such as reducing paperwork burdens for teachers and disciplining students. Effort toward professional development for teachers includes providing resources to or opportunities for teachers to improve teaching methods. Effort toward school instruction captures principal actions towards classroom observations of teachers, communicating feedback to teachers, and making data-driven decisions for school instruction. Finally, efforts toward school climate include actions toward school expectations, responsiveness to concerns, and building an atmosphere of trust and respect.

Denote X^d_{mjt} as the m^{th} ordinal survey response for teacher j in year t that measures latent principal effort dimension $d \in \{ad, pd, in, cl\}$, where $\{ad, pd, in, cl\}$ correspond to administration, professional development, instructional leadership, and school climate, respectively. Following Agostinelli and Wiswall (2025) and Cunha, Heckman, and Schennach (2010), I model each item as a noisy ordinal indicator of an underlying continuous effort factor e^d_{jt}:

$$X^{d*}_{mjt} = \mu^d_m + \lambda^d_m e^d_{jt} + u^d_{mjt}, \tag{9}$$

where X^d_{mjt} is defined as an ordinal system of equations as outlined in Appendix C. The parameters λ^d_m are factor loadings that determine how strongly item m loads on dimension d. The set $\mathcal{L}^d$ collects the survey items used to construct each dimension e^d_{jt}

To identify the latent factors, I impose various normalizations. First, I assume that $e^d_{jt} \sim \mathcal{N}(0,1)$, which fixes the scale of effort on each dimension. Second, I set $\mu^d_m = 0$ and assume that both the factor loadings and the item thresholds are time-invariant, so that changes over time are attributed to changes in effort rather than changes in the survey scale. Third, I assume local independence: conditional on e^d_{jt}, item-specific errors u^d_{mjt} are independent of the factor and independent across items, $u^d_{mjt} \perp e^d_{jt}$ and $u^d_{mjt} \perp u^{d'}_{m'jt}$ for $(d,m) \neq (d',m')$.[14] Under these assumptions, each effort dimension is identified from the ordinal responses, up to a sign normalization, where I assume higher survey ratings correspond to higher effort.

[13] Various literature has emphasized these dimensions of principal leadership central to the role of principals and their impact on schools. see Bloom et al. (2015); Grissom, Loeb, and Master (2013); Grissom, Loeb, and Mitani (2015); Hallinger and Murphy (1985); Sebastian and Allensworth (2012); Witziers, Bosker, and Krüger (2003)

[14] In the baseline specification, I currently do not allow for correlations across effort dimensions or across item-specific errors beyond their dependence on $\mathbf{e}_{jt}$.

Denote the vector of latent factors for teacher j in year t as $\mathbf{e}_{jt} = (e^{ad}_{jt}, e^{pd}_{jt}, e^{in}_{jt}, e^{cl}_{jt})'$. Let $\widehat{\mathbf{e}}_{jt}$ denote the predicted teacher-perceived effort for the principal of teacher j in year t. To obtain principal effort in each year, I aggregate these predicted teacher-level factors to the principal-year level. For each dimension $d \in \{ad, pd, in, cl\}$ and principal k in year t,

$$e^{d}_{kt} = \frac{1}{|J_{kt}|} \sum_{j \in J_{kt}} \widehat{e}^{d}_{jt}, \tag{10}$$

where J_{kt} denotes the set of teachers assigned to principal k in year t.

4.3 Policy Effects on Teacher VA

Next, I estimate teacher value-added on student test scores building on Chetty et al. (2014a), who provide evidence that teacher value-added is an unbiased measure of teacher quality. To account for some teachers having small class sizes, I utilize a shrinkage estimator similar to Chetty et al. (2014a). I estimate teacher value-added from the specification in Equation 11.

$$y_{ijcst} = \beta_0 + \beta_1 y_{i,t-1} + x'_{it}\beta_2 + z'_{cst}\beta_3 + \mu_{j\tau} + \epsilon_{ijcst}, \tag{11}$$

where y_{ijcst} is the test score of student i in class c taught by teacher j in school s in year t, x_{it} is a vector of student characteristics including indicators for economic disadvantage status, disability status, and limited English proficiency, as well as time-invariant demographics, z_{cst} is a vector containing classroom and school characteristics including size and demographic information. $\mu_{j\tau}$ is the estimated policy-variant teacher value-added in policy regime τ, and finally ϵ_{ijcst} is an idiosyncratic error. I construct Empirical Bayes estimates following the procedure outlined in Jackson (2018), which shrinks the dispersions of noisy estimates (smaller classrooms) toward the mean estimate.

4.4 Relating Principal Effort to Teacher Effectiveness

Finally, I regress the estimated teacher VA on teacher and principal effort and characteristics. Specifically, I estimate the specification in Equation 12:

$$\hat{\mu}_{jt} = \alpha_0 + \alpha_1 e^{ad}_{k} + \alpha_1 e^{pd} + \alpha_1 e^{in}_{k} + \alpha_1 e^{cl}_{k} + x'_{k}\alpha_4 + \epsilon_{jks}, \tag{12}$$

where e_k^d is principal effort along dimension d. x'_{kt} is a vector of principal observables, containing degree type, experience as an educator and principal, retirement eligibility, race, ethnicity, sex, and an indicator for working in an urban location.

5 Results

Tab 6 shows the estimated effect of principal exposure on math and reading outcomes specified in Equation 8. Tables 7 and 8 present changes in teacher VA and latent principal effort pre- and post-policy estimated from Equations 11 and 10, respectively. I find that principal effort toward all margins reduced significantly except effort toward the instructional program. Effort toward administrative tasks reduced the most, with a 0.10 SD decline, followed by effort toward school learning climate and professional development reducing 0.08 SD and 0.05 SD, respectively. Based on Figure 1, the distribution of teacher effectiveness in math declined post-policy, but there was no change in reading effectiveness. Compared to math, reading achievement is influenced by more than a student's teacher. For example, parents typically can help more with reading compared to mathematics concepts. Student reading skills are also developed through non-school assigned reading.

Table 6: Teacher Value Added Before and After Pay Reform

	Math		Reading	
	(1)	(2)	(3)	(4)
Exposure	-0.290***	-0.290***	-0.151*	-0.151***
	(0.060)	(0.057)	(0.087)	(0.047)
Mean	0.015	0.015	0.033	0.033
School-grade FE	Y	Y	Y	Y
District Clustering	Y	N	Y	N
School Clustering	N	Y	N	Y
N	347,112	347,112	347,118	347,118

Finally, Table 9 presents the effect of latent principal effort on teacher effectiveness of Equation 12. I find that teacher effectiveness is most responsive to principal effort along administrative tasks, which is where principals reduce effort the most. Teacher effectiveness in math increases with principal effort toward professional development, but not reading. Given students struggle more with math than reading, it is expected that general teacher training will improve effectiveness toward math more than reading.

Table 7: Teacher Value Added Before and After Pay Reform

	Math			Reading		
	Pre	**Post**	**Difference**	**Pre**	**Post**	**Difference**
	$\mu_{j,0}$	$\mu_{j,1}$	$\Delta\mu_j$	$\mu_{j,0}$	$\mu_{j,1}$	$\Delta\mu_j$
Grades 4-5	0.00	-0.05	-0.05***	0.00	0.00	-0.01**
	(0.25)	(0.24)	(0.00)	(0.16)	(0.15)	(0.003)
Grade 4	0.00	-0.05	-0.01**	0.00	0.00	0.02***
	(0.26)	(0.25)	(0.00)	(0.16)	(0.16)	(0.005)
Grade 5	0.00	-0.05	-0.03***	0.00	-0.01	0.00
	(0.24)	(0.23)	(0.00)	(0.16)	(0.15)	(0.004)
$N(gr.4-5)$	14,328	8,374	2,348	11,636	3,569	1,553

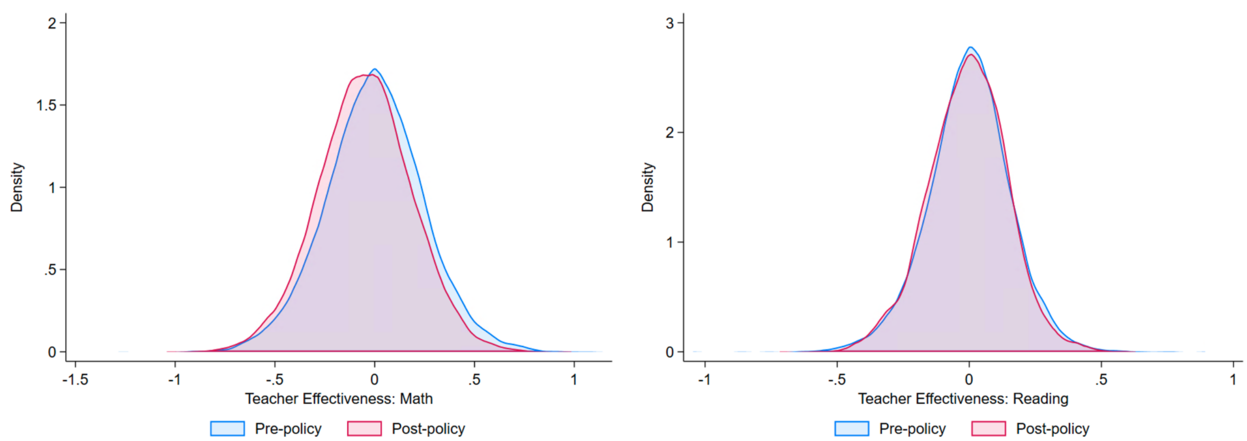

Figure 1: Distribution of Teacher Effectiveness in Math and Reading

Next, principal effort toward the school's instructional program reduces teacher effectiveness in math, but this effect is small and marginally significant. Finally, principal effort toward a school's learning climate reduces teacher effectiveness in math and reading. Principal effort along this margin includes efforts toward giving teachers autonomy in their curriculum. The decline in effectiveness along this margin may indicate teachers would benefit from a standardized curriculum or that when teachers are given more leadership, they have less capacity for teaching.

6 Conclusion

This paper studies the effects of incentive pay on principal effort, teacher effectiveness, and student test scores. Specifically, I use administrative data in North Carolina and Georgia and syn-

Table 8: Latent Principal Effort

	(1) **Pre** $e_{k,0}$	(2) **Post** $e_{k,1}$	(3) **Difference** Δe_k	(4) **Projected Diff.** Δe_k^P
Administrative	0.095	0.035	-0.060***	-0.102***
	(0.016)	(0.017)	(0.017)	(0.014)
Professional Dev.	0.054	0.024	-0.030**	-0.054***
	(0.010)	(0.010)	(0.015)	(0.013)
Instruction	0.059	0.061	-0.003	-0.006
	(0.010)	(0.010)	(0.014)	(0.012)
Climate	0.068	0.042	-0.023*	-0.082*
	(0.012)	(0.012)	(0.017)	(0.014)
N	874	874	874	874

*** p<0.01, ** p<0.05, * p<0.1. Standard errors of the mean are in parenthesis. Columns (1) and (2) display average principal latent effort in 2016 and 2018, respectively. Column (3) presents the raw difference in latent effort between 2018 and 2016 across all principals. Column (4) takes the difference between estimated and predicted latent effort in 2018. Predicted latent effort in 2018 is calculated using estimated coefficients from the following specification fitted on 2012-2016 data: $e_{k,0} = \alpha + \delta_t + \gamma_k + \epsilon_{kt}$, where $\delta_t + \gamma_k$ are year and principal fixed effects, respectively and ϵ_{kt} is an idiosyncratic error term. Standard errors are clustered at the principal level. This approach assumes that principal effort evolves linearly over time.

thetic and matching difference-in-differences models to estimate the effect of incentive pay on student test scores. Overall, I find that the incentive pay scheme implemented in North Carolina reduced student test scores by 0.12-0.15 SD in math and 0.13-0.16 SD in reading. Given this counterintuitive result, I decompose policy channels within North Carolina. Using teacher survey responses, I estimate principal latent effort across four dimensions during each survey year, teacher effectiveness pre- and post-pay policy, and the effect of principals on teachers for survey years. To further understand the impact of incentive pay on principals, I construct an incentive pay exposure measure for each principal, which measures their standardized difference in 2017-18 incentive-based pay and counterfactual experience-based pay. I find that principal effort toward administrative tasks, school climate, and professional development reduced by 0.05-0.10 SD. Teacher effectiveness declined by 0.05 SD in math and 0.01 SD in reading. Teacher effectiveness increases the most from principal effort toward administrative tasks, which declined the most. Furthermore, principals more exposed to incentive pay reduced test scores 0.02 SD below principals less exposed, while both groups still experienced declines in test scores. This effect is quite small, indicating the policy did not generate large differences in performance across principal performance.

Table 9: Relating Teacher Value Added to Latent Principal Effort

	(1)	(2)
	Teacher Effectiveness	
	Math	Reading
$\hat{e}_{ad}$: Administrative	0.084***	0.036***
	(0.008)	(0.005)
$\hat{e}_{pd}$: Professional Dev.	0.034***	-0.008*
	(0.007)	(0.004)
$\hat{e}_{in}$: Instruction	-0.021**	0.008
	(0.010)	(0.006)
$\hat{e}_{cl}$: Climate	-0.051***	-0.014***
	(0.008)	(0.005)
Mean	0.017	0.008
SD	0.218	0.113
N	26,593	20,349

*** p<0.01, ** p<0.05, * p<0.1. Standard errors of the mean are in parenthesis.

Taken together, the decline in student test scores can be attributed to the decline in principal effort. To provide intuition for this unexpected outcome, consider two groups of principals: (1) low-experience principals earning low salary pre-policy, who enjoy large salary jumps post-policy, and [15] (2) high-experience principals who had high salary levels in 2016-17, the year before policy implementation, and who received hold harmless pay once the new salary schedule was implemented in 2017-18. For the first group, incentive pay functioned as a large salary raise ($10,000 on average) regardless of effort. From an effort-consumption tradeoff perspective (adapted from the neoclassical labor-leisure tradeoff model), incentive pay generated a large income effect, reducing effort. Intuitively, while principals can still earn a higher salary from effectively outpacing expected student performance, the additional effort necessary to obtain a salary bonus has a steep marginal cost and low marginal benefit.

For the second group, the hold harmless salary, which prevents principals from being paid less under incentive-based pay, functions as insurance against bad states in terms of student performance. Additionally, with the highest possible salary a principal can receive decreasing by approximately $25,000 before longevity pay, principals in this group, who were already making high salary under experience-based pay, do not have large salary growth potential. Taking both of these observations together, principals in this group also face larger marginal costs of effort than

[15] The average annual salary increase for all principals from 2016-17 to 2017-18 was $9,600.

marginal benefits, and reduce their effort.

These results highlight the complexities of designing incentive pay policies. The particular pay incentive scheme implemented in North Carolina may have been more successful if principals were given a smaller salary raise, regardless of student performance, and could successfully achieve performance bonuses more frequently, with smaller necessary effort increases. In future work I will develop a Stackelberg model of principal and teacher effort decisions to further quantify the channels through which incentive pay works within a school and conduct various counterfactual incentive pay policy schemes, such as the one described above.

References

Agostinelli, F., & Wiswall, M. (2025). Estimating the technology of children's skill formation. *Journal of Political Economy*, *133*(3), 846–887.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009, 04). Large stakes and big mistakes. *The Review of Economic Studies*, *76*(2), 451-469. Retrieved from https://doi.org/10.1111/j.1467-937X.2009.00534.x doi: 10.1111/j.1467-937X.2009.00534.x

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021, December). Synthetic difference-in-differences. *American Economic Review*, *111*(12), 4088–4118. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.20190159 doi: 10.1257/aer.20190159

Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H., & Wilson, D. (2009). Evaluating the impact of performance-related pay for teachers in england. *Labour Economics*, *16*(3), 251-261. Retrieved from https://www.sciencedirect.com/science/article/pii/S0927537108001218 doi: https://doi.org/10.1016/j.labeco.2008.10.003

Bandiera, O., Prat, A., Guiso, L., & Sadun, R. (2011, January). Matching firms, managers and incentives [Working Paper]. (16691). Retrieved from http://www.nber.org/papers/w16691 doi: 10.3386/w16691

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2012, 11). Does management matter? evidence from india *. *The Quarterly Journal of Economics*, *128*(1), 1-51. Retrieved from https://doi.org/10.1093/qje/qjs044 doi: 10.1093/qje/qjs044

Bloom, N., Lemos, R., Sadun, R., & Reenen, J. V. (2015). Does management matter in schools? *The Economic Journal*, *125*(584), 647–674. Retrieved 2025-10-20, from http://www.jstor.org/stable/24737101

Bluestein, S., & Goldschmidt, P. (2021). Principal effects on academic progress over time and the potential effects of school context and principal leadership practices. *Journal of School Administration Research and Development*, *6*(1), 12–23.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (Tech. Rep.). National Bureau of Economic Research.

Brehm, M., Imberman, S. A., & Lovenheim, M. F. (2017). Achievement effects of individual performance incentives in a teacher merit pay tournament. *Labour Economics*, *44*, 133-150. Retrieved from https://www.sciencedirect.com/science/article/pii/

S0927537116304286 doi: https://doi.org/10.1016/j.labeco.2016.12.008

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a, September). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.104.9.2593 doi: 10.1257/aer.104.9.2593

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b, September). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–79. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.104.9.2633 doi: 10.1257/aer.104.9.2633

Chiang, H., Lipscomb, S., & Gill, B. (2016). Is school value added indicative of principal quality? *Education Finance and Policy*, *11*(3), 283–309.

Coelli, M., & Green, D. A. (2012). Leadership effects: school principals and student outcomes. *Economics of Education Review*, *31*(1), 92-109. Retrieved from https://ideas.repec.org/a/eee/ecoedu/v31y2012i1p92-109.html doi: 10.1016/j.econedurev.2011

Cohodes, S., Eren, O., & Ozturk, O. (2023). *The long run effects of a comprehensive teacher performance pay program on student outcomes* (Working Paper No. 31056). Retrieved from https://www.nber.org/system/files/working_papers/w31056/w31056.pdf (NBER Working Paper No. 31056)

Cunha, F., Heckman, J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, *78*(3), 883-931. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA6551 doi: https://doi.org/10.3982/ECTA6551

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from impact. *Journal of Policy Analysis and Management*, *34*(2), 267–297. Retrieved 2025-10-20, from http://www.jstor.org/stable/43866371

Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement? *Canadian Journal of Economics/Revue canadienne d'économique*, *47*(2), 634–663.

Dhuey, E., & Smith, J. (2018). How school principals influence student learning. *Empirical Economics*, *54*, 851–882.

Di Liberto, A., Giua, L., Schivardi, F., Sideri, M., & Sulis, G. (2023). *Managerial practices and student performance: Evidence from changes in school principals* (Tech. Rep.). IZA Discussion Papers.

Figlio, D., & Kenny, L. (2007, June). Individual teacher incentives and student performance. *Jour-*

nal of Public Economics*, *91*(5-6), 901–914. (Funding Information: We have benefitted from comments from John Krieg, Heather Rose and Hanna Skandera and seminar participants at the University of Florida and the American Education Finance Association. We appreciate the helpful comments from Jim Poterba and two anonymous reviewers for this Journal. We are very grateful for Jin Jeon's help in coding the surveys and alerting us to various inconsistencies in the responses. We appreciate the financial support of the National Science Foundation and the Warrington College of Business Administration. All errors are our own.) doi: 10.1016/j.jpubeco.2006.10.001

Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from new york city public schools. *Journal of Labor Economics*, *31*(2), 373–407. Retrieved 2025-10-20, from http://www.jstor.org/stable/10.1086/667757

Glewwe, P., Ilias, N., & Kremer, M. (2010, July). Teacher incentives. *American Economic Journal: Applied Economics*, *2*(3), 205–27. Retrieved from https://www.aeaweb.org/articles?id=10.1257/app.2.3.205 doi: 10.1257/app.2.3.205

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, *42*(8), 433–444. doi: 10.3102/0013189X13510020

Grissom, J. A., Loeb, S., & Mitani, H. (2015). Principal time management skills: Explaining patterns in principals' time use, job stress, and perceived effectiveness. *Journal of Educational Administration*, *53*(6), 773–793. doi: 10.1108/JEA-09-2014-0117

Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *The Elementary School Journal*, *86*(2), 217–247. doi: 10.1086/461445

Heckman, J., Ichimura, H., & Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, *64*(4), 605–654.

Heckman, J., Ichimura, H., & Todd, P. (1998, 04). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, *65*(2), 261-294. Retrieved from https://doi.org/10.1111/1467-937X.00044 doi: 10.1111/1467-937X.00044

Imberman, S. A., & Lovenheim, M. F. (2015). Incentive strength and teacher productivity: Evidence from a group-based teacher incentive pay system. *Review of Economics and Statistics*, *97*(2), 364–386.

Jackson, C. K. (2018). What do test scores miss? the importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), 2072-2107. Retrieved from https://doi.org/10.1086/699018 doi: 10.1086/699018

Jones, M., & Hartney, M. T. (2017). Show who the money? teacher sorting patterns and performance pay across u.s. school districts. *Public Administration Review*, *77*(6), 919–931. Retrieved 2024-05-09, from http://www.jstor.org/stable/26648691

Lavy, V. (2009, December). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, *99*(5), 1979–2011. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.99.5.1979 doi: 10.1257/aer.99.5.1979

Lazear, E. P. (2000, December). Performance pay and productivity. *American Economic Review*, *90*(5), 1346–1361. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.90.5.1346 doi: 10.1257/aer.90.5.1346

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, *9*(2), 508–523. Retrieved 2025-10-20, from http://www.jstor.org/stable/3003596

Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review*, *36*, 60-72. Retrieved from https://www.sciencedirect.com/science/article/pii/S0272775713000800 doi: https://doi.org/10.1016/j.econedurev.2013.05.004

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from india. *Journal of Political Economy*, *119*(1), 39–77. Retrieved 2025-10-20, from http://www.jstor.org/stable/10.1086/659655

National Center for Education Statistics. (2025). *Ccd reference library: Glossary (document #7).* Retrieved 2025-10-16, from https://nces.ed.gov/ccd/doc/7_CCD_Glossary.docx (Common Core of Data (CCD))

North Carolina General Assembly. (2016, December). *Final report to the 2017 session of the 2017 general assembly of north carolina* (No. st12201). Raleigh, NC. Retrieved from https://www.ncleg.gov/Files/Library/studies/2017/st12201.pdf (Prepared by the Joint Legislative Study Committee on School-Based Administrator Pay)

Roland G. Fryer, J. (2017, May). *Management and student achievement: Evidence from a randomized field experiment* (NBER Working Papers No. 23437). National Bureau of Economic Research, Inc. Retrieved from https://ideas.repec.org/p/nbr/nberwo/23437.html doi: None

Rosen, S. (1982). Authority, control, and the distribution of earnings. *The Bell Journal of Economics*, *13*(2), 311–323. Retrieved 2024-04-24, from http://www.jstor.org/stable/3003456

SAS Institute Inc. (2023). *Sas® evaas: Statistical models and business rules* (Tech. Rep.). Cary, NC: Prepared for the North Carolina Department of Public Instruction. Retrieved from https://ncdpi.sas.com/ (Includes gain, predictive, and projection models; business rules)

Sebastian, J., & Allensworth, E. (2012). The influence of principal leadership on classroom instruction and student learning: A study of mediated pathways to learning. *Educational Administration Quarterly*, *48*(4), 626–663. doi: 10.1177/0013161X11436273

Smith, J., & Todd, P. (2005). Does matching overcome lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, *125*(1), 305-353. Retrieved from https://www.sciencedirect.com/science/article/pii/S030440760400082X doi: https://doi.org/10.1016/j.jeconom.2004.04.011

Westall, J. (2022). The labor market consequences of principal performance pay. *SSRN 4213350*.

Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, *39*(3), 398–425. doi: 10.1177/0013161X03253411

Wu, Y. (2017, 04). Incentive contracts and the allocation of talent. *The Economic Journal*, *127*(607), 2744-2783. Retrieved from https://doi.org/10.1111/ecoj.12397 doi: 10.1111/ecoj.12397

Table 10: Principal Characteristics

	Pre-Policy			Post-Policy		
	All	Bottom 25%	Top 25%	All	Bottom 25%	Top 25%
Total Experience	21.03	23.99	18.19	20.88	25.08	19.75
	(6.85)	(5.66)	(5.17)	(5.94)	(6.24)	(5.59)
Experience as Principal	6.18	6.81	5.15	6.24	7.53	5.69
	(4.43)	(4.84)	(3.55)	(4.36)	(5.14)	(4.07)
Base Monthly Nominal Salary	5,142	5,343	4,958	6,015	5,830	6,087
	(569)	(543)	(401)	(581)	(621)	(634)
Base Monthly Real Salary	5,230	5,414	5,037	5,832	5,661	5,903
	(582)	(550)	(405)	(545)	(582)	(598)
Base Annual Nominal Salary	61,707	64,113	59,494	72,183	69,960	73,044
	(6,830)	(6,518)	(4,806)	(6,972)	(7,446)	(7,605)
Base Annual Real Salary	62,761	64,964	60,445	69,990	67,934	70,833
	(6,985)	(6,598)	(4,858)	(6,546)	(6,986)	(7,173)
Has Hold Harmless Supplement	0.13	0.17	0.13	0.23	0.46	0.17
	(0.33)	(0.37)	(0.34)	(0.42)	(0.50)	(0.38)
Number of Teachers	36.68	36.46	37.68	36.59	35.57	37.38
	(14.07)	(13.67)	(14.27)	(14.14)	(13.61)	(14.64)
Growth Expectation Not Met	0.24	0.33	0.15	0.25	0.23	0.25
	(0.43)	(0.47)	(0.36)	(0.43)	(0.42)	(0.43)
Growth Expectation Met	0.49	0.47	0.48	0.48	0.50	0.47
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Growth Expectation Exceeded	0.28	0.20	0.37	0.27	0.27	0.28
	(0.45)	(0.40)	(0.48)	(0.44)	(0.45)	(0.45)
No Growth Expectation Score	0.00	0.00	0.00	0.00	0.00	0.00
	(0.03)	(0.04)	(0.00)	(0.04)	(0.04)	(0.00)
Elementary School	0.73	0.74	0.77	0.73	0.72	0.74
	(0.44)	(0.44)	(0.42)	(0.44)	(0.45)	(0.44)
Middle School	0.27	0.26	0.23	0.27	0.28	0.26
	(0.44)	(0.44)	(0.42)	(0.44)	(0.45)	(0.44)
Turnover 1	0.09	0.09	0.09	0.10	0.05	0.12
	(0.28)	(0.29)	(0.29)	(0.31)	(0.22)	(0.32)
Turnover 2	0.13	0.00	0.00	0.11	0.13	0.11
	(0.33)	(0.00)	(0.00)	(0.31)	(0.34)	(0.31)
Turnover 2, Retirement Eligible	0.28	0.00	0.00	0.14	0.15	0.11
	(0.45)	(0.00)	(0.00)	(0.35)	(0.36)	(0.31)
Turnover 2, Not Retirement Eligible	0.09	0.00	0.00	0.10	0.11	0.11
	(0.29)	(0.00)	(0.00)	(0.30)	(0.31)	(0.32)
Same School	0.78	0.73	0.75	0.78	0.95	0.75
	(0.41)	(0.45)	(0.43)	(0.41)	(0.22)	(0.43)
Retirement Eligible	0.14	0.20	0.04	0.10	0.29	0.06
	(0.35)	(0.40)	(0.20)	(0.30)	(0.46)	(0.24)
Age	45.93	48.80	43.00	45.61	49.65	44.37
	(7.77)	(6.55)	(6.61)	(7.10)	(7.08)	(7.00)
Sixth Year Advanced Degree	0.11	0.14	0.09	0.10	0.14	0.09
	(0.32)	(0.35)	(0.28)	(0.29)	(0.35)	(0.29)
Doctorate Degree	0.09	0.15	0.06	0.11	0.17	0.10
	(0.28)	(0.36)	(0.24)	(0.31)	(0.37)	(0.30)
Master's Degree	0.79	0.71	0.84	0.79	0.68	0.80
	(0.40)	(0.45)	(0.36)	(0.41)	(0.47)	(0.40)
White	0.74	0.72	0.72	0.73	0.71	0.71
	(0.44)	(0.45)	(0.45)	(0.44)	(0.45)	(0.45)
Black	0.24	0.25	0.26	0.24	0.25	0.26
	(0.43)	(0.43)	(0.44)	(0.43)	(0.44)	(0.44)
Other Race	0.01	0.02	0.02	0.01	0.01	0.01
	(0.12)	(0.14)	(0.13)	(0.11)	(0.11)	(0.11)
Asian	0.00	0.00	0.00	0.00	0.01	0.00
	(0.03)	(0.06)	(0.00)	(0.06)	(0.11)	(0.06)
Female	0.66	0.70	0.65	0.67	0.70	0.66
	(0.47)	(0.46)	(0.48)	(0.47)	(0.46)	(0.47)
N	6,770	797	1,799	3,357	538	1,539

Table 11: Difference in Difference Estimates by Grade: Elementary School Students

	Grade 3		Grade 4		Grade 5	
	Math	Reading	Math	Reading	Math	Reading
DiD	-0.227***	-0.166***	-0.103***	-0.236***	-0.011	-0.071**
	(0.029)	(024)	(0.026)	(0.026)	(0.034)	(0.029)
Mean	-0.188	-0.177	-0.141	-0.165	-0.279	-0.274
SD	0.936	0.933	0.935	0.953	0.972	1.038
N	9,461	9,461	9,324	9,324	9,313	9,313

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Observations consider school-grade-year units in grades 3-5 during the 2014-15 through 2017-18 academic years. Test scores are standardized using 2016-17 subject-grade means and standard deviations. Standard errors are clustered at the district level. Controls include school-grade-year level shares of ethnicity, race, sex, economic disadvantage status, limited English proficiency status, bins for the number of students who took the exam, and school indicators for being located in an urban region and Title I status.

Table 12: Difference in Difference Estimates: Middle School Students

	Math	Reading
DiD	-0.060***	0.029
	(0.024)	(0.024)
School-grade FE	Y	Y
Year FE	Y	Y
Mean	-0.318	-0.261
SD	0.968	1.031
N	12,940	12,940

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Observations consider school-grade-year units in grades 6-8 during the 2014-15 through 2017-18 academic years. Test scores are standardized using 2016-17 subject-grade means and standard deviations. Standard errors are clustered at the district level. Controls include school-grade-year level shares of ethnicity, race, sex, economic disadvantage status, limited English proficiency status, bins for the number of students who took the exam, and school indicators for being located in an urban region and Title I status.

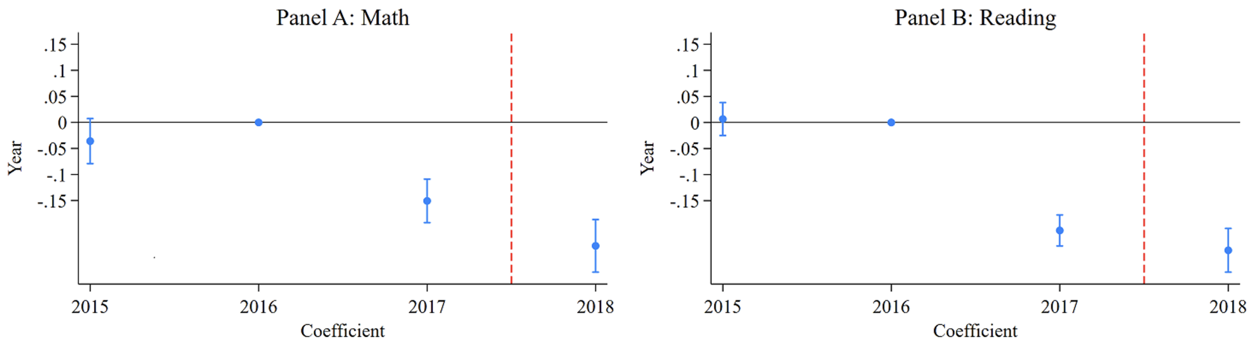

Figure 2: Event Study Results

Table 13: Balancing Testing: Math

	NC	GA
White	0.089***	-0.000
	(0.007)	(0.007)
Black	-0.130***	0.000
	(0.007)	(0.006)
Hispanic	0.025***	0.001
	(0.004)	(0.005)
Female	0.003**	0.000
	(0.001)	(0.001)
SWD	0.003**	-0.000
	(0.001)	(0.001)
EDS	0.010*	0.001
	(0.006)	(0.006)
Urban School	-0.136***	0.000
	(0.012)	(0.013)
Title 1 Eligible School	0.075***	0.001
	(0.009)	(0.010)
Avg. Number of Students/Grade	-0.526***	-0.007
	(0.025)	(0.028)
N	7,004	7,004

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Observations consider school-grade-year units in grades 3-5 during the 2016-17 through 2017-18 academic years. Column (1) presents estimates from an uncontrolled, univariate regression, and column (2) additionally controls for propensity scores. Robust standard errors are reported in parentheses.

Table 14: Teacher Value Added Before and After Pay Reform: Heterogeneity

	Pre $\mu_{j,0}$	**Post** $\mu_{j,1}$	**Difference** $\Delta\mu_j$
Low Principal Exp.	0.00	-0.01	-0.02***
	(0.25)	(0.22)	(0.007)
High Principal Exp.	0.00	-0.02	-0.02***
	(0.22)	(0.22)	(0.007)
Low EDS	0.02	-0.01	-0.02***
	(0.22)	(0.21)	(0.006)
High EDS	0.00	-0.02	0.02***
	(0.27)	(0.24)	(0.008)
Low Student Perf.	0.01	-0.05	-0.06***
	(0.27)	(0.24)	(0.011)
Hight Student Perf.	0.03	0.01	-0.02***
	(0.23)	(0.22)	(0.210)

Table 15: Relating Teacher Value Added to Latent Principal Effort

	Teacher Effectiveness	
	Math	Reading
$\hat{e}_{ad}$: Administrative	0.084***	0.036***
	(0.008)	(0.005)
$\hat{e}_{pd}$: Professional Dev.	0.034***	-0.008*
	(0.007)	(0.004)
$\hat{e}_{in}$: Instruction	-0.021**	0.008
	(0.010)	(0.006)
$\hat{e}_{cl}$: Climate	-0.021**	0.008
	(0.010)	(0.006)
Principal Yrs. Educ. Experience	-0.001***	-0.000
	(0.000)	(0.000)
Principal Yrs. Prin. Experience	0.001***	-0.000
	(0.000)	(0.000)
Prin. Doct. Degree	-0.044***	-0.011*
	(0.010)	(0.006)
Prin. Mast. Degree	-0.050***	-0.019***
	(0.009)	(0.005)
Prin. 6th Yr Adv. Degree	-0.051***	-0.015***
	(0.010)	(0.006)
Retirement Eligible	0.017***	0.011***
	(0.005)	(0.003)
Prin. White	0.028***	0.017**
	(0.011)	(0.007)
Prin. Black	0.040***	0.025***
	(0.011)	(0.007)
Prin. Hispanic	0.011	-0.007
	(0.011)	(0.012)
Prin. Female	0.010***	0.005***
	(0.003)	(0.002)
Urban School	0.033***	0.021***
	(0.003)	(0.002)
Constant	0.013	-0.011
	(0.015)	(0.009)
Mean	0.017	0.008
SD	0.218	0.113
N	26,593	20,349

Appendix A Additional Figures and Tables

Table A1: Mean Characteristics of Students and Schools

	North Carolina	Georgia
White	0.49	0.40
	(0.26)	(0.29)
Black	0.26	0.37
	(0.21)	(0.29)
Hispanic	0.18	0.16
	(0.13)	(0.17)
Other Race	0.06	0.04
	(0.07)	(0.02)
Female	0.49	0.49
	(0.05)	(0.03)
Economically Disadvantaged	0.68	0.66
	(0.21)	(0.27)
Students with Disabilities	0.12	0.12
	(0.05)	(0.03)
English Learner	0.06	0.08
	(0.08)	(0.12)
Urban School	0.49	0.59
	(0.50)	(0.49)
Title I School	0.80	0.73
	(0.40)	(0.45)
Number of Students	169.76	209.88
	(105.83)	(130.28)
N	21,348	18,944

Observations consider school-grade-year units in grades 3-5 during the 2014-15 through 2017-18 academic years in North Carolina and Georgia. Standard deviations of the mean are reported in parentheses.

Table A2: Survey items by principal effort dimension

Teacher leadership and autonomy - Teachers have input into school decisions - Teachers are trusted to make sound decisions - Teachers are recognized as educational experts - Teachers are encouraged to lead - School leadership shares responsibility with teachers - Teachers are encouraged to try new ideas - Teachers are comfortable raising concerns - Teachers help set school priorities - Teachers help select instructional materials - Teachers help determine professional development - Teachers help develop school policies - Teachers have a meaningful role on improvement teams - Staff collaborate on problem solving - Staff have opportunities for leadership roles - Staff feel ownership over school initiatives - Teachers feel empowered to influence outcomes	Student conduct and climate - Students understand school behavior expectations - Administrators consistently enforce rules - Administrators support classroom discipline - Administrators address teacher concerns - Staff treat each other with respect - There is an atmosphere of trust - Teachers feel comfortable raising issues - Administrators are responsive to feedback - Staff concerns are resolved in a timely way - Communication between staff and administration is open - Staff feel safe reporting problems - The school is a good place to work and learn - Staff share responsibility for school climate - Expectations for student conduct are clear and fair
Instructional leadership and performance management - Administrators set clear expectations for teaching - Classroom observations provide useful feedback - Teacher evaluation is fair and consistent - Administrators support improving instruction - Administrators communicate a clear vision for instruction - Administrators follow up on instructional goals - Teachers receive feedback that helps them improve - School uses data to guide instruction - State assessment data are available and used - Results from assessments are used for improvement - Administrators remove barriers to effective teaching	Professional development and enabling conditions - Professional development is aligned with priorities - Professional development is followed up in classrooms - Professional development meets teachers' needs - Professional development is differentiated - Time is provided for professional learning - Teachers have sufficient instructional time - Non-instructional duties are reasonable - Paperwork requirements are manageable - Class interruptions are minimized - Materials, technology, and facilities are adequate - Support personnel are available when needed

Source: NCERDC

FY 2016-17 PRINCIPAL SALARY SCHEDULES
PRINCIPAL III
22 - 32 Teachers
Effective July 1, 2016

Combined Years of Exp	Schedule/ Pay Level	Base Monthly Salary	Base + 1% Monthly Salary	Base + 2% Monthly Salary	Base + 3% Monthly Salary	Base + 4% Monthly Salary	Base + 5% Monthly Salary	Base + 6% Monthly Salary
0-17	0-11	$4,735	$4,782	$4,830	$4,877	$4,924	$4,972	$5,019
18	0-12	$4,797	$4,845	$4,893	$4,941	$4,989	$5,037	$5,085
19	0-13	$4,860	$4,909	$4,957	$5,006	$5,054	$5,103	$5,152
20	0-14	$4,924	$4,973	$5,022	$5,072	$5,121	$5,170	$5,219
21	0-15	$4,992	$5,042	$5,092	$5,142	$5,192	$5,242	$5,292
22	0-16	$5,058	$5,109	$5,159	$5,210	$5,260	$5,311	$5,361
23	0-17	$5,126	$5,177	$5,229	$5,280	$5,331	$5,382	$5,434
24	0-18	$5,196	$5,248	$5,300	$5,352	$5,404	$5,456	$5,508
25	0-19	$5,266	$5,319	$5,371	$5,424	$5,477	$5,529	$5,582
26	0-20	$5,342	$5,395	$5,449	$5,502	$5,556	$5,609	$5,663
27	0-21	$5,415	$5,469	$5,523	$5,577	$5,632	$5,686	$5,740
28	0-22	$5,490	$5,545	$5,600	$5,655	$5,710	$5,765	$5,819
29	0-23	$5,565	$5,621	$5,676	$5,732	$5,788	$5,843	$5,899
30	0-24	$5,644	$5,700	$5,757	$5,813	$5,870	$5,926	$5,983
31	0-25	$5,726	$5,783	$5,841	$5,898	$5,955	$6,012	$6,070
32	0-26	$5,808	$5,866	$5,924	$5,982	$6,040	$6,098	$6,156
33	0-27	$5,881	$5,940	$5,999	$6,057	$6,116	$6,175	$6,234
34	0-28	$5,998	$6,058	$6,118	$6,178	$6,238	$6,298	$6,358
35	0-29	$6,118	$6,179	$6,240	$6,302	$6,363	$6,424	$6,485
36	0-30	$6,240	$6,302	$6,365	$6,427	$6,490	$6,552	$6,614
37	0-31	$6,365	$6,429	$6,492	$6,556	$6,620	$6,683	$6,747
38	0-32	$6,492	$6,557	$6,622	$6,687	$6,752	$6,817	$6,882
39	0-33	$6,622	$6,688	$6,754	$6,821	$6,887	$6,953	$7,019
40	0-34	$6,754	$6,822	$6,889	$6,957	$7,024	$7,092	$7,159
41+	0-35	$6,889	$6,958	$7,027	$7,096	$7,165	$7,233	$7,302

NOTES:

1. ADD $126 per month for an earned advanced degree.
2. ADD $253 per month for an earned doctorate degree.
3. Placement on one of the Base through Base + 6% salary schedules is determined by 1997-98, 1998-99 and 1999-2000 only ABCs and School Safety accomplishments.
4. A principal shall be placed on the step on the salary schedule that reflects the total number of years of experience as a certified employee of the public schools and an additional step for every 3 years serving as a principal on or beofre June 30, 2009.

Figure A1: Principal Salary Schedule 2016-17

PRINCIPAL SALARY
(Monthly Schedules)
FY 2017-18
Effective July 1, 2017

ADM Range	Schedule/ Pay Level	Base Monthly Salary	Schedule/ Pay Level	Growth Met Monthly Salary	Schedule/ Pay Level	Growth Exceeded Monthly Salary
up to 400	B1	$5,145.92	G1	$5,660.50	E1	$6,175.08
401 to 700	B2	$5,403.25	G2	$5,943.50	E2	$6,483.83
701 to 1,000	B3	$5,660.50	G3	$6,226.58	E3	$6,792.58
1,001 to 1,300	B4	$5,917.83	G4	$6,509.58	E4	$7,101.33
over 1,300	B5	$6,175.08	G5	$6,792.58	E5	$7,410.08

Figure A2: Principal Salary Schedule 2017-18

<table>
<tr><th>Base</th><th>Met</th><th>Exceeds</th></tr>
<tr><td>Not met +
Not met +</td><td>Met +
Met+
Exceeded/Not Met</td><td>Exceeded +
Exceeded +
Not Met/Met</td></tr>
<tr><td>Principal has not supervised a school for 2 of the last 3 years</td><td>Exceeded +
Met +
Not Met</td><td></td></tr>
<tr><td></td><td>Principal for 2 of the last 3 years of a school not eligible to receive a school growth status</td><td></td></tr>
</table>

Figure A3: Principal Growth Status Determination

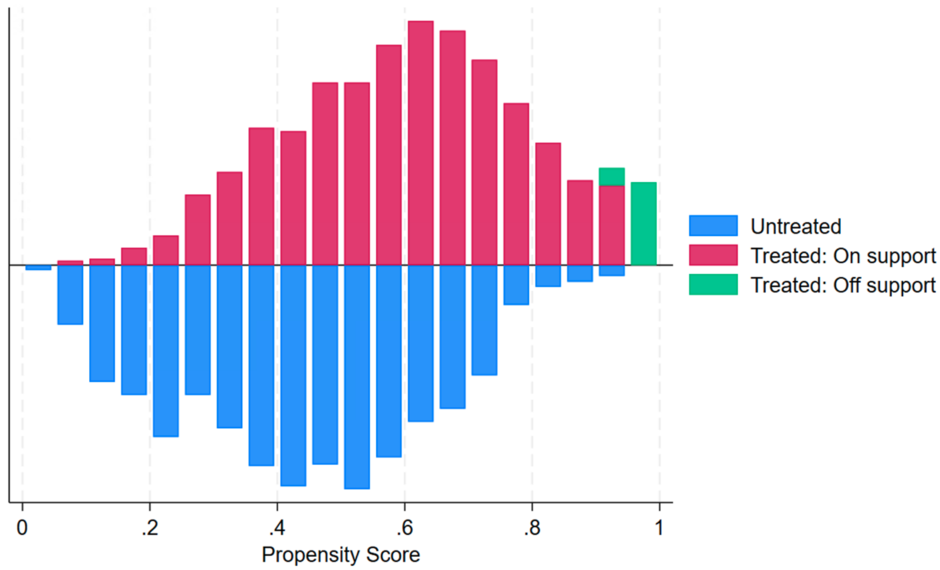

Figure A4: Propensity Score Matching: Overlap of Common Support

Appendix B Teacher Working Conditions Survey

Table B3 displays an example set of questions give to teachers on school leadership. Responses coded as "Don't Know" are not included, as they do not provide meaningful information information on principal perceptions. Such responses are small, typically representing less than 5% of responses to a questions across survey periods. I select survey questions targeted at school leadership based on the set of questions common across the 2012, 2014, 2016, and 2018 survey waves. All responses are coded such that a higher response value corresponds to a more favorable view of the survey subject.

Table B3: Questions from the North Carolina Teacher Working Conditions Survey

School Leadership

Please rate how strongly you agree or disagree with the following statements about school leadership in your school.

1 = Strongly Disagree
1 = Disagree
1 = Agree
1 = Strongly Agree
1 = Don't Know

There is an atmosphere of trust and mutual respect in this school.

The school leadership consistently supports teachers.

The school improvement team provides effective leadership at this school.

Teachers are held to high professional standards for delivering instruction.

Teacher performance is assessed objectively.

The procedures for teacher evaluation are consistent.

Teachers receive feedback that can help them improve teaching.

Teachers feel comfortable raising issues and concerns that are important to them.

The school leadership facilitates using data to improve student learning.

The faculty are recognized for accomplishments.

NCERDC Teacher Working Conditions Survey, 2018

Appendix C Principal Latent Effort

This appendix describes the measurement model used to construct the four dimensions of principal effort discussed in Section 3: administration, professional development, instructional leadership, and school climate. For each dimension $d \in \{ad, pd, in, cl\}$ and survey item $m \in \mathcal{L}^d$, let X^d_{mjt} denote the ordered response of teacher j in year t about the principal of her school.

I assume that each observed ordinal response X^d_{mjt} is generated by an underlying continuous latent variable

$$X^{d*}_{mjt} = \mu^d_m + \lambda^d_m e^d_{jt} + u^d_{mjt}, \tag{13}$$

where e^d_{jt} is the latent principal effort factor perceived by teacher j in year t on dimension d, λ^d_m is the factor loading for item m, and u^d_{mjt} is an idiosyncratic error term. Following the standard latent factor literature, I impose the normalizations $e^d_{jt} \sim \mathcal{N}(0,1)$ and $u^d_{mjt} \sim \mathcal{N}(0,1)$ and set $\mu^d_m = 0$ so that the location of each item is absorbed into the thresholds.[16] Observed responses take on ordered values $c \in \{1,2,3,4\}$, where higher values correspond to more favorable views of the principal on that item. Let

$$-\infty = \tau^d_{m,0} < \tau^d_{m,1} < \tau^d_{m,2} < \tau^d_{m,3} < \tau^d_{m,4} = +\infty$$

denote item-specific thresholds. The observed ordinal response is then defined by

$$X^d_{mjt} = c \quad \text{if} \quad \tau^d_{m,c-1} < X^{d*}_{mjt} \leq \tau^d_{m,c}, \qquad c = 1,2,3,4.$$

Conditional on latent effort e^d_{jt}, the probability of observing response c on item m follows an ordered probit structure:

$$\Pr\big(X^d_{mjt} = c \mid e^d_{jt}\big) = \Phi\big(\tau^d_{m,c} - \lambda^d_m e^d_{jt}\big) - \Phi\big(\tau^d_{m,c-1} - \lambda^d_m e^d_{jt}\big), \qquad c = 1,2,3,4, \tag{14}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

Let θ^d collect the parameters for dimension d (loadings $\{\lambda^d_m\}_{m\in\mathcal{L}^d}$ and thresholds $\{\tau^d_{m,c}\}$). For each dimension d, teacher j, and year t, the likelihood contribution integrates over the latent effort

[16] I assume local independence: conditional on $\mathbf{e}_{jt}$, item-specific errors are independent of the factors and of each other. In the baseline specification I do not allow for additional correlations across effort dimensions or across item errors.

factor:

$$L_{jt}^{d}(\theta^{d}) = \int \prod_{m \in \mathcal{L}^{d}} \Pr\big(X_{mjt}^{d} \mid e_{jt}^{d}; \theta^{d}\big)\, \varphi(e_{jt}^{d})\, de_{jt}^{d}, \tag{15}$$

where $\varphi(\cdot)$ is the standard normal density. The log-likelihood across dimensions, teachers, and years is

$$\log L(\theta) = \sum_{d} \sum_{j,t} \log L_{jt}^{d}(\theta^{d}), \tag{16}$$

and I estimate the parameter vector $\hat{\theta}$ by maximum likelihood. Predicted teacher-level effort $\widehat{e}_{jt}^{d}$ is then aggregated to the principal-year level as in Equation (10) in the main text.