Pension Reform and Labor Supply: Retention and Productivity under a Pension Cut

Andrew C. Johnston[*] Jonah Rockoff James Harrington[†]

September 27, 2024

Abstract

We examine the effect of a representative pension reform on the retention and productivity of public-school teachers. The reform reduced pension annuities and increased penalties for early retirement—projected to save eight percent of revenues. We leverage detailed administrative records and a discontinuity in the reform to estimate its effect. The reform increased worker retention by 2%, discouraging early retirement. Using idiosyncratic within-school variation in reform exposure, we find the reform increased student achievement by 0.03σ, mediated by retention and effort. Thus, the reform maintained or improved both the extensive and intensive margins of labor supply.

JEL: H75, J26, M52

[*]University of California at Merced, 5200 Lake Road, Merced CA, 95340. Email: acjohnston@ucmerced.edu. The authors are grateful to the Texas Education Agency for providing the data, and we thank Chad Aldeman, Adam Bloomfield, Mark Duggan, Maria Fitzpatrick, Dillon Fuchsman, Todd Gormley, Robert Jensen, Cory Koedel, Alex Mas, Olivia S. Mitchell, Josh McGee, and Laura Quinby for constructive comments. Ryan Santos provided outstanding research assistance

[†]Columbia Business School, 3022 Broadway, New York NY, 10027. Email: jonah.rockoff@columbia.edu

1 Introduction

Pension debt poses a significant challenge across the modern world. In 20 OECD countries, these obligations now exceed official government debt (Mitchell, 2020; Hanif et al., 2016). And, according to the World Economic Forum, liabilities are projected to grow sixfold by 2050, to $400 trillion (2017).

The challenge is especially pronounced in the United States, where governments have accrued the world's largest shortfall (Hanif et al., 2016; Mitchell, 2020; Equable Institute, 2023).[1] The resulting fiscal pressure has prompted governments to cut education and infrastructure spending, arguably reducing economic growth and upward mobility (Anzia, 2017; Koedel, 2019; Barro, 1991; Duflo, 2001; Czernich et al., 2011; Chetty et al., 2014b).[2] To stabilize pensions and protect other public goods, policymakers across the country have introduced reforms that curb the cost of pensions (e.g., Goldhaber et al., 2015; Gohl et al., 2023). This paper examines the effect of one such reform.

Opposing these reforms are advocates of the public workforce. They warn that pension cuts—now implemented in some form in all states—risk undermining, or even "crippling," public services (Woods, 2010; Van Hollen, 2017; Sabatosa, 2021).[3] They contend that cutting pensions will reduce morale, effort, and retention, which harms frontline services and exacerbates staffing shortages in vital functions like education and public safety (Keefe, 2014; Kenneally and Bond, 2020; Early, 2023).

Given these competing concerns—with mounting debt on one side and vital services on the other—this paper examines the apparent tradeoff of reform. Namely, *how do pension cuts affect labor supply, worker output, and the quality of public services?* To address these questions, we use detailed administrative data and a discontinuity that allows us to estimate the effects of a representative reform.

[1] American unfunded liabilities are expected to grow from $28 trillion in 2015 to $137 trillion in 2050, all in 2015 dollars (World Economic Forum, 2017).

[2] See also Illinois Policy Institute (2024)

[3] Critics of pension cuts sometimes use stark language to describe the potential impacts of reform. U.S. Senator Chris Van Hollen (D-MD) writes, for instance, that a reform similar to the one we consider would "decimate" the public workforce, especially their ability to "retain the best and brightest" (Van Hollen, 2017). For their part, *HR Magazine* warns that cuts affecting government workers will "cripple front line services" and "decimate staff wellbeing" (Woods, 2010).

The reform we examine cut costs on two fronts. First, it reduced the annuities workers would receive in retirement by 3–11 percent, depending on a worker's earning history. Second, the reform multiplied the penalties for retiring early. Before the reform, a worker could retire at age 55 with 90 percent of the benefits she would have received under the full-retirement formula. After the reform, her annuity would be cut to *half* that. The congressional budget office estimated the reform would save the state $250 million each year, approximately 8 percent of yearly pension contributions (Legislative Budget Board, 2005).

Economic theory offers conflicting predictions about how pension cuts like these would affect the retention and productivity of workers. Given conflicting theoretical predictions, empirical evidence becomes essential for understanding its actual impact.

The reform we examine has a few unique features that lend themselves to empirical insight. First, the reform's cutoffs allow us to compare essentially identical workers in the same setting who are subject to different pension regimes. This provides for an especially careful comparison. Second, Texas's early adoption allows us to observe the long-term impact of reform. Third, unlike most pension reforms which only affect new hires, the reform we examine affected *current* workers. This unique aspect allows us to avoid selection issues when estimating the causal effect.[4] Fourth, since public schools employ over half of government workers, the setting provides substantial statistical power. And fifth, schools collect annual achievement records, offering psychometrically validated measures of output. Together, these features provide an exceptional setting for understanding the implications of reform.

To measure its effect, we exploit the fact that workers beyond age and experience thresholds were *grandfathered* into the old pension, while those who didn't were placed into the new one. The cutoffs enable a quasi-experimental discontinuity design.[5]

Theory offers conflicting predictions on retention: On one side, reform may reduce retention by effectively cutting pay (Stock and Wise, 1990; Coile and Gruber, 2007); on

[4]Almost every state since has reduced pension parameters for new hires while preserving benefits for existing workers (Brainard and Brown, 2018). Over half of all pension programs in the country raised retirement ages or vesting requirements. The vast majority only affected new hires (Quinby et al., 2018).

[5]A worker was grandfathered if, in the fall of 2005, she was at least 50 years old, had at least 25 years of experience, *or* the sum of her age and experience exceeded 70.

the other, reform may increase retention by penalizing early retirement or by income effects (Lazear, 1986; Camerer et al., 1997).

We measure the reform's effect on retention using a regression discontinuity design (RDD), comparing similar workers on either side of the threshold. One year after implementation, the reform slightly *reduced* retention, by 0.2 percentage points, consistent with a morale effect. However, the effect reversed substantially in subsequent years. Five years after implementation, the reform increased retention by 2.9 percentage points, and this positive effect persisted through the ten-year mark (both significant at the 0.001 level). Fifteen years post-reform, retention rates converged. Notably, the retention effects were concentrated in years the reform reduced early retirement benefits, suggesting it improved retention mostly by discouraging early retirement.

To further validate our findings, we employ a difference-in-discontinuity design (DDRD), comparing the main discontinuity at the reform threshold to a placebo discontinuity from five years prior. This approach allows us to control for any pre-existing differences or specification errors endemic to the threshold, providing a more demanding test of the reform's causal effect. These results confirm the robustness of our main findings and, if anything, suggest that the reform's impact may be somewhat larger than our main estimates indicated. The difference-in-discontinuity estimates, for instance, recover a retention effect that is 30 percent larger than our baseline estimate at the ten-year mark.

We next examine whether the reform affected teacher productivity, as measured by unexpected gains in math and reading skills. Theory again predicts conflicting effects. Gift-exchange models predict the reform (an effective pay cut) will depress morale and thereby reduce output (Akerlof, 1982; Weakliem and Frenkel, 2006; Kube et al., 2013). Models by Becker and Lazear, however, predict workers increase effort when early retirement is less attractive (Lazear, 1979a; Gibbons and Murphy, 1992). This is because a longer work horizon encourages skill investment and discourages shirking (Becker, 1962; Ben-Porath, 1967; Lazear, 1979a).

To estimate the reform's effect, we employ a difference-in-differences approach. Our identification strategy leverages both the timing of the reform and the variation in exposure across school-grade cells. The approach reveals that the reform increased math achievement by 0.03 student standard deviations (significant at 0.001) and increase reading achievement by 0.01 (significant at 0.010). To address potential endogeneity concerns, we also estimate models that instrument exposure with exposure predicted from teaching assignments the year before the reform. This approach generates similar, somewhat larger, estimates in both math and reading.

When we control for the mediating influence of teacher experience, the effects attenuate by 7 to 30 percent, suggesting that while retention responses to the reform explain some of the effect, most is attributable to changes in effort.[6]

The 2005 Texas reform foreshadowed a wave of similar measures nationwide. Between 2009 and 2014, three quarters of states implemented cost-cutting pension reforms—most sharing elements of the Texas model. And since 2009, all 50 states have implemented cost-saving measures in their pension programs (Aubry and Crawford, 2017; Quinby et al., 2018). Mirroring Texas, nearly two-thirds of states reduced pension annuities by calculating final average salary over a longer period. And, like Texas, approximately 40 percent of states cut early-retirement benefits (Brainard and Brown, 2018). Consequently, our results speak to a wider set of reforms across the country.

This paper contributes to a literature examining the influence of pensions on worker turnover and performance (Friedberg and Webb, 2005; Costrell and McGee, 2010; Goldhaber et al., 2017; Fitzpatrick, 2019; Ni et al., 2020).[7] Brown (2013) uses a differencing-and-bunching approach to evaluate the effect of pensions on labor supply. Manoli and Weber (2016) use retirement-eligibility notches in Austria for the same purpose. Koedel and Xiang (2017) and Ni et al. (2021) examine the effect of a local pension enhancement on retention.

What is unique about this paper is its ability to examine the effect of a cost-cutting reform on both the retention and the output of workers. In particular, the environment

[6]An alternative explanation—improved selection in retention—is unlikely for two reasons. One, teacher exit from pensions typically doesn't correlate with effectiveness (Koedel et al., 2013; Bates and Johnston, 2024), and, if anything, retirement programs tend to retire less effective workers (Fitzpatrick and Lovenheim, 2014). This would predict the opposite of what we find. Two, the small changes in retention, moreover, would require unrealistically large differences in selection to explain the achievement increase.

[7]A related literature carefully examines how workers value their retirement benefits (Fitzpatrick, 2015; Biasi, 2019; Fuchsman et al., 2020; Ni and Podgursky, 2016). See also Johnston (2021)

provides a rare opportunity to examine the consequences of pensions on worker productivity, which is not normally observable. The cutoffs provide us with simple treatment-control comparisons, where similar workers working in the same setting make choices under different pension regimes. We thus can shed light on the effect of reform, both on the extensive and intensive margins of labor supply.

2 The Pension Reform

This section details the 2005 Texas pension reform. We describe the bill's legislative history, its key provisions, its rationale, and how it was communicated to teachers. We then explain the grandfathering provisions and provide a quantitative assessment of the reform's impact on pension benefits. This overview lays the foundation for our empirical analysis.

The Texas reform bill was introduced in March of 2005 by bipartisan coauthors, Craig Eiland (a House Democrat from Galveston) and Robert Duncan (a Senate Republican from Lubbock). It left its legislative committees without opposition, passed in the state House and Senate by wide margins in May, and was signed into law by the governor that June. As will become clear, the bill's swift passage, occurring at the end of a school year, is helpful for our empirical strategy.

The bill had two primary features affecting current workers. First, it expanded the basis for computing retirement annuities from the highest three years of salary to the highest five. In effect, this reduces a worker's pension annuity by the equivalent of an experience step in the pay schedule for a typical teacher, usually a permanent three-percent reduction. For workers serving stints in leadership (or other highly paid, temporary roles), the cut was larger. For instance, a teacher serving three years as principal would see her annuity fall by 11 percent under the new regime. This is because her high salary years would be diluted when calculating Final Average Salary (FAS) in the new calculation.[8,9] Second, the reform

[8]Calculating FAS based on more years mitigates the cost of pension "spiking." Pension spiking is the practice of having a few years of very high salary to substantially alter a worker's pension annuity (Fitzpatrick, 2017). The canonical example is for workers to strategically time leadership to maximize annuities. Incorporating more years to compute a worker's final average salary (FAS) makes spiking harder and less effective at raising annuities above typical compensation.

[9]Teachers who served short stints in leadership experienced larger annuity cuts. Leadership pay in our

multiplied the penalties for retiring early. Annuities for those claiming early at age 60 would be cut by a third. Annuities for those claiming early at age 55 would be cut in half.[10]

The bill's sponsors argued that these cuts were a necessary response to ballooning liabilities and rising costs for schools and taxpayers. Rather than cuts, opponents of the bill proposed counter legislation *raising* annuities for current retirees and increasing taxpayer contributions from the legislature by 42 percent, from 6 percent of payroll to 8.5 (Legislative Budget Board, 2005). The legislature passed the cost-reducing measure rather than the cost-increasing one.

The reform was well-communicated to teachers. In the summer the bill passed, the union's magazine, *Advocate*, featured the pension reform as its cover story with a fullpage picture of a browning apple core. In her quarterly letter to members, the union president wrote of the recent pension reform: "*Sine die*...it feels just that way—that someone or something has *died*. Could it be our hopes and dreams...? This was the most disappointing and disheartening legislative session we have ever seen" (emphasis original; Haschke, 2005).[11] To its hundreds of thousands of members, the union provided detailed, accurate information about the reform and the grandfathering cutoff (Texas State Teachers' Association, 2005; National Center for Education Statistics, 2008).[12]

To soften the blow, the reform grandfathered workers nearer retirement, leaving their pension benefits unaffected. A worker would be grandfathered if, on Sept 1, 2005, she were at least 50 years old, had at least 25 years of experience, or had age and experience summing to at least 70. The relevant threshold for just over half of workers was the age cutoff. These are workers who began their career for Texas public schools sometime after their 30th birthday. The relevant threshold for the substantial remainder was the rule of 70 which grandfathered workers who started careers in public schools relatively early in life.[13]

data is typically an increase of 30 percent *ceteris paribus*. Because the final average salary was calculated on a larger base, FAS falls more for employees who had short stints in leadership. For example, a teacher who served as principal for the last three years of her career would see her yearly FAS fall by by 11 percent under the new regime, reducing her pension annuity in proportion.

[10]Teachers could retire with 20 years of service at age 55 in the old regime and receive 90 percent of their full benefits. After the reform, the same person would receive 47 percent of their full benefits.

[11]*Sine die* has the same orthography as the English verb "to die," but the two are unrelated. The Latin phrase adjourn a proceeding with no appointed date to resume. It means "without a date".

[12]Texas had at the time 291,000 unionized teachers.

[13]The experience cutoff was not relevant for most workers because a worker would have had to begin

Employees who hadn't met one of those thresholds were subject to the new regime while those meeting at least one cutoff were shielded from any change in their pension benefits.

To understand the impact of these changes, we compare pension benefits under the two regimes. We calculate final average salary (FAS) for each worker in the administrative records using a three- and five-year basis. On average, the FAS is 4 percent lower using five years instead of three, implying a 4 percent cut to retirement annuities for the average affected worker. Workers see larger cuts when they have a few years of high salary from a leadership position, since a larger base reduces the influence of irregular high years.

To examine the effect of greater penalties for early retirement, we calculate the present value of retirement wealth under both regimes for various archetypal workers.[14] We calculate the present value of benefits assuming workers (i) live to age 84 (the mean for college-educated women), (ii) use an annual discount factor of 4 percent (Ericson and Laibson, 2018; Johnston, 2021), and (iii) have final average salary of $94,500 in 2021 dollars (the sample mean in 2005).

For workers with 10 years of experience at age 50, the annuity reduction comes only from the change to calculating final average salary, usually reducing the annuity by 4 percent. The early-retirement cut never affects this group because they are eligible for full retirement before becoming eligible for early retirement.

Before the reform, a worker with 20 years of experience at age 50 would maximize the present value of her pension wealth by claiming at age 55, generating a stream of benefits valued at $1.2 million, approximately $700,000 in present discounted value at age 55. After the reform, however, the penalties for early retirement cut the annuity stream in half when the teacher claims at age 55. Thus under the new regime, the worker would maximize the present value of her annuity by *delaying* retirement for five years, which reduces the present value of her retirement stream by 18 percent, compared to the old regime. Assuming a longer life or lower discount factor shrinks the perceived reduction in value from reform.[15]

continuous employment before age 20 for experience to be the relevant cutoff.

[14] We hold constant the years of service to show the effect of the reform on pension wealth.

[15] To demonstrate the size of the cut, these calculations fix years of service and FAS to compare the two regimes with the same inputs. If we allow years of service to respond endogenously to maximize the present value of retirement income, the effect of the reform actually raises pension value by 2 percent over the status quo, but it comes at the cost of working an additional five years. When we also account for changes in FAS

The cut is smaller for teachers who become early-retirement eligible between ages 55 and 60. For workers who are already eligible for full retirement at age 50, the cut is 4 percent from the reform through the expanded FAS base.

In the study of pensions, examining the employees of public schools presents several important advantages. First, the majority of public-sector workers are in education, providing large administrative records and accompanying statistical power. Second, because of the profession's size, reach, and influence on long-run outcomes, the results are valuable in themselves (Chetty et al., 2014b; Papay and Kraft, 2015a). Third, data on private-sector workers are quite hard to collect. Since teachers and other workers in public schools are government employees, their records are often made available to researchers. Fourth and finally, output measures are available in education in the form of standardized skill measures, allowing analysts to examine how reform affects teacher performance.

3 Data Description

We collect administrative staffing records on all public school employees from the Texas Education Agency covering 2000 through 2021. The data include yearly employment records for some 2,046,975 individual workers. Many are teachers (50 percent), but the data also include nurses, therapists, librarians, bus drivers, custodial staff, and other pension-eligible employees. We focus our analysis on teachers where we can estimate both retention and output effects.[16] In each year, the records indicate a worker's unique identifier, her district, campus, age (measured on September 1 of each year), professional role, grade assignment, subject taught, years of experience, base pay, total pay, full-time or part-time status, and whether she works on contract.[17] Yearly employment records allow us to observe *when* a worker leaves a school, a district, or public education altogether.

While these staffing records provide detailed information about employee retention, we

due to extended years of service, the new regime's present value is 6 percent greater than the old.

[16]We also present retention estimates for all workers in the online Appendix.

[17]We do not have workers' exact dates of birth which would suggest an alternative approach using the discontinuity in age, as measured in days around the 50-year threshold and the cutoff date of September 1. The state is concerned about safeguarding the privacy of employees.

also aim to measure the reform's impact on productivity. We use student-level achievement data on math and reading collected by the Texas Education Agency and accessed through a secure research data center. The achievement data cover grades three to eight, from 2000 to 2011. Although we can't directly match students to specific teachers, we know which teachers instruct math and reading in each grade. We therefore can construct reliable measures of reform exposure that vary *across* grades within each school, providing careful within-school comparisons for causal evaluation.

4 The Texas Teacher Pension System

To fully understand the implication of the pension reform, it is helpful to know the retirement eligibility rules in Texas. There, workers are eligible for normal retirement if they are 65 years old with at least 5 years of service, or the sum of their age and experience totals 80 or more (e.g., a 50 year old worker with 30 years of experience would qualify).[18] When she meets one of these requirements, a teacher is eligible to claim her pension and start receiving an annuity, which is calculated:

$$A_i = YOS_i \times 2.3\% \times FAS_i \qquad (1)$$

The annuity received by an eligible worker i is the product of three factors: the worker's years of service (YOS_i); the state's benefit multiplier (2.3% in Texas); and the worker's "final average salary" (FAS_i). The FAS_i is calculated as the average salary of the worker's highest earning three years before the reform and the highest five after the reform. (The term is somewhat imprecise, since the highest earning years need not be "final.") A worker retiring with 30 years of credit and a final average salary of \$94,500 (the sample mean) would receive a yearly annuity of \$65,205 for the rest of her life, adjusted for inflation each year (\$94,500 × 30 × 0.023 = \$65,205).

A worker can claim a reduced annuity if she is at least 55 years old with at least

[18] Since 2005, the state pension has undergone further reform, but these reforms have not affected the cohorts we study since they grandfather essentially all incumbent workers.

20 years of experience. However, early retirement penalties reduce her annuity for each credit (year of age or experience) she falls short of normal retirement eligibility. The reform significantly changed these early retirement penalties. Before the reform, early-retirement penalties were approximately 2 percent for every credit short of full-retirement eligibility.[19] In the new regime, the penalties were raised to 10–12 percent per unit short of normal retirement. These penalties brought down annuities for early retirees by up to half of their pre-reform levels.[20]

5 The Effect of Reform on Worker Retention

This section examines the effect of the 2005 Texas pension reform on worker retention, a key outcome of interest for policymakers and a primary rationale for state pension systems. We begin by explaining how we measure retention using our panel data. We then detail our methodological approach, including how we determine workers' eligibility for grandfathering and our regression discontinuity design. Finally, we discuss our model specification and bandwidth selection when estimating the reform's impact on worker retention.

One of the primary rationales for state pensions is to retain workers in the state's public schools (Lazear, 1979a; Gustman et al., 1994). Policymakers have reason to be concerned with retention since worker productivity improves with experience (Wiswall, 2013; Papay and Kraft, 2015b), and because hiring is costly in terms of both time and money spent recruiting, screening, and onboarding (Barnes et al., 2007; Watlington et al., 2010). The panel nature of our data enables us to observe when a worker departs public-school teaching in Texas, precisely the outcome of interest to policymakers. If a worker leaves public school for a private school, or moves to a public school in Baton Rouge, she has not been retained in Texas public schools and thus disappears from our records.

Using the age and experience a worker has in the fall of 2005, we calculate each worker's

[19] If she retired at age 55 with 20 years of service, she would be 5 years short of normal eligibility, according to the rule of 80, and the annuity would be reduced by $5 \times 2\% = 10\%$ of the annuity calculated in equation (1).

[20] A range is provided because the penalties aren't strictly formulaic. They are encoded, not in a simple formula, but a complicated table whose penalties do not follow a simple pattern (Teacher Retirement System of Texas, 2019).

distance to the grandfathering cutoff. This calculation involves three values:

- Age cutoff: We compute the employee's distance beyond the age cutoff (age on September 1, 2005, minus 50)
- Experience cutoff: We compute the employee's distance beyond the experience cutoff (experience accrued by September 2005 minus 25).
- Rule-of-70 cutoff: We compute the employee's distance beyond the rule-of-70 cutoff (experience plus age in September 2005, minus 70).

A worker needs to meet only one of these criteria to be grandfathered. Her effective distance to grandfathering, therefore, is the most positive distance beyond any of these cutoffs. Those with distance greater than or equal to zero are grandfathered and those with negative values are subject to the reform.

We model the outcome variable R_i (usually retention for y years) as a continuous function of distance to the grandfathering cutoff, and we estimate the outcome discontinuity that occurs at the threshold:

$$R_i = \beta T_i + f(x_i - x') + u_i \tag{2}$$

Here, $x_i - x'$ is a worker's distance to the eligibility cutoff, x', and T_i equals one if worker i was subject to the reform, and zero if not. Thus, $f(x_i - x^{'})$ is a function of the running variable that controls for the continuous relationship between distance to the cutoff and retention. This allows us to isolate the causal effect of the reform on retention.

We present both linear and quadratic specifications to accommodate possible curvature in the relationship between the running variable and the outcomes, serving as a robustness check on our results. The estimates from the two models are never statistically distinguishable, but the point estimates are less precise when using a quadratic specification.

In regression discontinuity designs, limiting the bandwidth—the range of data around the cutoff used in the analysis—allows the analyst to more closely approximate experimental conditions. We use a triangular kernel and calculate Imbens-Kalyanaraman bandwidths to balance the tradoff between bias and precision (Imbens and Kalyanaraman, 2012). Because the quadratic specification requires additional degrees of freedom, we sometimes have too few observations to estimate standard errors within the optimal bandwidth. We therefore use twice the IK bandwidth when estimating quadratic models, and we also present estimates with a wide range of alternative bandwidths to assess robustness. Each alternative bandwidth produces very similar estimates throughout the support with quadratic models, and the estimates are quite stable within a large range for linear models. At much larger bandwidths, linear models produce large estimates suggesting the reform substantially improved retention. Though these are directionally in line with the main estimates, we view their size as the result of specification error, rather than capturing the true effect.

This design allows us to carefully estimate the causal effect of the pension reform on worker retention, addressing a key question in our study: How do pension cuts affect labor supply on the extensive margin?

5.1 Evaluating the Regression Discontinuity

The regression discontinuity design produces an unbiased estimate of the causal effect of reform, if two assumptions hold.

First, the assignment variable (distance to the threshold) must have a continuous effect on the outcome at the cutoff. This assumption ensures that any discontinuities in the outcome can be attributed to discontinuous treatment rather than to inherent characteristics of the assignment variable. While not directly testable, this assumption is supported in our case by the fact that outcomes evolve smoothly along the assignment variable away from the cutoff.

The second assumption is that no other determinants of the outcome are discontinuous at the cutoff. This assumption guarantees that the treatment is the only factor affecting outcomes discontinuously at the threshold. In examining education and pension legislation in Texas, we find no other laws or reforms that rely on the same or similar cutoff rules (i.e., age 50, 25 years of service, or the rule of 70) to determine benefits during work or in retirement. The primary remaining concern in this context is the possibility of workers manipulating their placement around the threshold. Manipulation of the assignment variable can threaten the validity of the design because it could introduce a discontinuity in *unobservable* characteristics driven by selection around the cutoff.

But manipulation in our setting is especially unlikely. First, it would represent deliberate and discoverable fraud. Age and year of experience are easily verifiable through records already available to the employer, making falsification both difficult and risky for workers. Second, we find empirical evidence against manipulation. If workers were able to manipulate their assignment variable to avoid the reform, we would expect to see an excess mass of workers on the favorable side of the threshold, and a deficit on the unfavorable side (McCrary, 2008).[21] We present the density of workers around the cutoff in Figure 2. The distribution is smooth, including at the threshold, suggesting there was no measurable manipulation of the running variable into treatment or control.

Together, these facts suggest our setting can produce unbiased estimates of the reform's effect.

5.2 Retention Estimates from the Discontinuity

This subsection presents our estimates of the reform's effect on worker retention, leveraging the regression discontinuity design. We begin by visualizing retention rates around the cutoff, then provide formal estimates for various retention horizons, and conclude with robustness checks.

Figure 3 presents worker retention as a function of a worker's distance to the threshold. We display four series, each indicating the share of workers retained over different time periods after the reform: 1 year, 5 years, 10 years, and 15 years post-reform.

To quantify the effects and assess their statistical significance, we turn to formal RDD estimation in Table 2. Our analysis considers multiple measures of retention: whether a worker is retained for at least 1 year, at least 5 years, at least 10 years, and at least 15 years (the longest period allowed by the time elapsed since the reform). To generate a

[21] Alternatively, the Texas legislature could have chosen the exact cutoffs to benefit particular subpopulations of education workers. The fact that they simply chose round numbers (50, 25, 70) suggests this was not the case.

cumulative measure of retention rates, we calculate the *mean* retention rate over 15 years for each worker we observe in the reform year.[22] These varied measures allow us to examine both short- and long-term effects of the reform on worker retention.

The discontinuity estimates reflect the effect of the reform for workers near the cutoff. The point estimates are positive, suggesting the reform increased retention rates. Local linear estimates suggest that the reform increased five-year retention by 1.9 percentage points (2.3 percent) and ten-year retention by 1.9 percentage points (3.6 percent), both significant at the 0.001 level. To further explore the dynamics of the impact, we examine year-by-year retention effects in Appendix Figure A.3 where we present estimated retention effects each year after the reform. Effects are close to zero in the first few years, after which the reform increases retention and the effect remains at about 2 percentage points from five to eleven years post reform. These are also the years in which early retirement was heavily penalized for teachers near the threshold.

The average worker at the cutoff was 49.4 years old in the year of the reform, so five to eleven years after is when affected workers concentrated around ages 54.4 to 60.4. These are the ages at which early retirement was attractive in the old regime, but heavily penalized in the new one. After eleven years, the effects fall, eventually becoming insignificant. Over the 15-year window, the reform is associated with an average 1.2 pp increase in retention (2.2 percent), significant at the 0.001 level. This result implies that the reform extended careers by 2.4 months on average. In the quadratic specification, the point estimate on the effect of the reform on five- and ten-year retention suggests the reform increased retention by 1.2–1.6 percentage points with 95-percent confidence intervals ruling out small attrition effects. Specifically, the quadratic estimates rule out attrition effects five years after reform, and rule out attrition effects of more than 0.02 percentage points ten years after reform).

To further assess the validity of our results, we conduct placebo tests and use them to estimate a difference-in-discontinuity design. The placebo estimates help evaluate whether specification error or unobservables may introduce bias in the design. We estimate the RDD as if the reform took place in 2000, five years before the actual implementation.

[22]Specifically, we count how many years they work after the reform and divide by the number of possible years they could have worked, which is 15.

These placebo estimates reveal that unobservables or specification error tend to bias the estimates *against* our findings, with opposite-signed retention estimates at 5 and 10 years. The opposite sign estimates indicate that any bias would lead us to underestimate the true positive effect of the reform on retention, making our main results conservative.

The difference-in-discontinuity estimates (DDRD), present the difference between the baseline estimates and the placebos, presented in Appendix Table A1. They indicate that our baseline estimates are robust, and if anything the true effect may be also larger than found in the RDD. The estimates from the DDRD at 5 years are larger than the baseline estimates (7.7 percentage points instead of 2.9), and the effects at 10 years are also larger than baseline (2.5 percentage points instead of 1.9). Notably, our baseline estimates showed an initial reduction in retention associated with the reform. This also appears in the placebo, suggesting that the initial decrease may be an artifact of specification error rather than a true reform effect.

To explore the sensitivity of our results to bandwidth selection, we present estimates at a range of bandwidths—from two to twenty around the threshold—in Appendix Figures A.5 and A.4. The linear estimates are robust to reductions or expansions in bandwidth. The optimal bandwidth tends to be conservative, with larger estimates at both narrower and broader bandwidths. For all bandwidth selections, estimates are positive and highly significant in the linear model. For quadratic models, the estimates are also stable across a range of bandwidths, with each estimate being small. In the sample of teachers, quadratic estimates are usually not significant. In the sample of all workers, quadratic estimates are all positive and half are significant at the five percent level.

6 The Effect of Reform on Worker Output

We next examine the reform's effect on worker output, measured as unexpected gains in student achievement.[23] Understanding the impact of pension reform on worker output is helpful for policymakers who are attempting to balance fiscal health with the quality of

[23]That is, gains not predicted by past performance and demographics. This approach allows us to isolate the impact of the reform from other factors that influence student performance.

public services. This is especially true in schools where teacher performance significantly affects human capital formation in children, with far-reaching consequences (Chetty et al., 2011; Petek and Pope, 2023).

Theory offers conflicting predictions on how such a reform will affect worker performance. Two models predict that pay cuts will reduce effort. First, the classic Akerlof model proposes that the labor markets can be understood as a system of gift exchange between employers and employees (Akerlof, 1982). Because of reciprocity motives or notions of fairness, workers respond to compensation with proportional effort. Morale models have the same basic prediction. Generous payments improve worker morale which in turn elicits greater effort and dedication. The inverse is that, when compensation is cut for any reason, workers are predicted to scale back their effort accordingly (Akerlof and Yellen, 1990; Fehr et al., 1993; Mas, 2006). In fact, fear of cratering morale and effort is the most commonly cited rationale for nominal wage rigidity (Campbell III and Kamlani, 1997; Bewley, 1998).

Conversely, some models predict the reform may increase effort in schools by penalizing early retirement. First, the human capital investment model suggests that workers facing a longer career horizon may invest more in their skills and knowledge, leading to increased productivity (Becker, 1962; Ben-Porath, 1967). Second, shirking models argue that by increasing the cost of job loss or workplace conflict, the reform may reduce shirking and encourage greater effort (Shapiro and Stiglitz, 1984; Lazear, 1979b; Stock and Wise, 1990; Coile and Gruber, 2007). These two models suggest that, despite effectively reducing compensation, the pension reform could paradoxically increase worker effort and improve outcomes. Unrelated to effort, the reform could improve output by increasing teacher retention which builds the stock of experience (Wiswall, 2013; Papay and Kraft, 2015b).

To empirically test these competing theoretical predictions and estimate the empirical impact of the pension reform on worker output, we employ a quasi-experimental research design that leverages variation in reform exposure across grades within a school.

6.1 Exploiting Idiosyncratic Variation in Reform Exposure

Our identification strategy leverages quasi-experimental variation in student exposure to reform-affected teachers. We exploit idiosyncratic differences in exposure across grades within the same school. This variation naturally arises from differences in teacher age and experience across grades. This variation is plausibly exogenous, as it is unlikely that students sort into specific grades within a school based on teachers' reform exposure, age, or experience.

This allows us to compare the performance of grades with different exposures to the reform in the same school, using a difference-in-differences design.

Our approach enables causal inference of the reform's effect while controlling for various potential confounding factors. Leveraging variation across school, grade, and time, we implement an exacting triple-difference specification. This method allows for precise within-school comparisons while carefully controlling for secular trends that differ by school. We formalize this approach in equation 3:

$$A_{isgt} = \pi T_{sgt} + \beta(T_{sgt} \times post) + g(A_{isg,t-1}) + \Gamma X_{isgt} + \alpha_{st} + \gamma_{gt} + \delta_{gs} + \varepsilon_{sgt} \qquad (3)$$

In this equation, A_{isgt} represents the achievement of student i in school s and grade g at time t. We normalize A_{isgt} to have mean zero and standard deviation one, allowing coefficients to describe effects in easily interpretable student standard deviations. The treatment exposure, measured by T_{sgt}, ranges from 0 to 1 and reflects the proportion of teachers in each school-grade cell exposed to the reform. A value of 0 indicates all the teachers in the cell were grandfathered (not subject to the reform), while 1 indicates all teachers were exposed to the reform.

To calculate this exposure measure, we use detailed staffing data that capture the time allocation of teachers across different grades and subjects. This granular approach is necessary because teaching assignments vary: some teachers specialize in math or reading, while others cover both subjects or spread their time across multiple grades. We first compute the reform exposure for each individual teacher. We then aggregate teacher exposure to the school-grade level, weighted each teacher by the proportion of time each spends in that specific grade and subject.[24] In our analytic sample, the average student has an exposure of 53 percent in math (standard deviation: 37 percent) and 49 percent in reading (standard deviation: 36 percent). The support in each ranges from 0 percent to 100 percent.

The coefficient of interest is β, which captures the effect of the reform on student achievement in terms of student standard deviations. The coefficient π serves as a useful test of the parallel trends assumption. Specifically, π indicates whether reform exposure correlates with outcomes *before* the reform was enacted. Reassuringly, we find no such correlation between eventual reform exposure and pre-reform student outcomes. This lack of pre-reform correlation suggests that the control group provides a suitable counterfactual time trend, as omitted factors appear to be orthogonal to treatment.

The equation includes several other components. The term $g(A_{isg,t-1})$ represents a cubic polynomial of students' prior achievement, which explains a substantial portion of the variation in current achievement and significantly improves the precision of our estimates. The component X_{isgt} is a vector of student-level controls, including indicators for sex, ethnicity, ESL (English as a Second Language), and low-income status, all interacted with grade. We also incorporate a comprehensive set of fixed effects: α_{st}, γ_{gt}, and δ_{gs} represent school-year, grade-year, and grade-school fixed effects, respectively. These allow us to account not just for differences across schools, grades, and years individually, but also for the unique combinations of school-year, school-grade, and grade-year (see Hendricks, 2014).

Our approach allows us to control for stable unobservable differences across treatment groups while simultaneously adjusting for changes in common unobservable factors over time. We achieve this by using control cells—those completely shielded from the reform—as a benchmark. The coefficient β thus captures the differential change in achievement post-reform between treatment and control cells. Essentially, it quantifies how much more (or less) achievement changed in reform-exposed cells compared to unexposed cells.

The treatment share of a cell T_{sgt} may change endogenously over time due to teacher

[24] For instance, a full-time third-grade math teacher exposed to the reform contributes more to the cell's treatment intensity than a math teacher who splits her time evenly between two grades or one who also teaches reading.

turnover or strategic staffing decisions in response to the reform. To address this potential endogeneity, we instrument the post-reform exposure share with the predicted exposure share based on teacher compositions in 2004, the year before the reform was announced and implemented. We denote this instrument as $T_{sg,2004}$. Intuitively, the policy was not known in 2004, making the allocation of reform exposure across school-grade cells based on 2004 staffing plausibly exogenous. This method isolates the exogenous portion of the variation in post-reform exposure.

Our instrument demonstrates considerable strength in the first stage. When predicting reform exposure post-reform, we observe F-statistics of 2,488 for math and 1,702 for reading, both well above the conventional threshold of 10 or 20 for a sufficiently strong instrument (Stock and Yogo, 2005). We implement the instrument using two approaches: a reduced form approach, to estimate an intent-to-treat effect of the reform on student achievement, and a two-stage least squares approach to estimate the local average treatment effect.

The reduced form is estimated:

$$A_{isgt} = \beta(T_{sg,2004} \times post) + g(A_{isg,t-1}) + \Gamma X_{isgt} + \alpha_{st} + \gamma_{gt} + \delta_{gs} + \varepsilon_{sgt} \tag{4}$$

In this equation, $T_{sg,2004}$ represents the school-grade's exposure to the reform based on its teacher composition in 2004. The coefficient β captures the intent-to-treat effect, measuring the impact of being assigned treatment in 2004. It's important to note that because the instrument doesn't vary over time, we cannot estimate the placebo effect of the reform prior to its implementation when using school-grade fixed effects. All other components of the equation remain as previously defined.

We also implement the instrument using a two-stage least squares (2SLS) estimator. The first and second stages are represented by the following equations:

$$\widehat{T_{sgt} \times post} = \rho T_{sg,2004} + \theta(T_{sg,2004} \times post) + f(A_{isg,t-1}) + \Lambda X_{isgt} + \mu_{st} + \nu_{gt} + \xi_{gs} + \eta_{sgt} \tag{5}$$

$$A_{isgt} = \pi T_{sgt} + \beta(\widehat{T_{sgt} \times post}) + g(A_{isg,t-1}) + \Gamma X_{isgt} + \alpha_{st} + \gamma_{gt} + \delta_{gs} + \varepsilon_{sgt} \quad (6)$$

Equations 5 and 6 formalize our two-stage least squares approach. In the first stage, we use exposure share in 2004 interacted with the post-reform indicator as an instrument for the potentially endogenous treatment shares observed after the reform. The second stage then uses these predicted values to estimate the causal effect of treatment shares. In this framework, the coefficient β captures the local average treatment effect (LATE) of reform exposure, providing a causal estimate of the reform's impact on student achievement. This approach allows us to address potential endogeneity concerns and isolate the causal effect of reform.

6.2 Estimated Effect on Output

In Table 4, we present the estimates of how reform affected math achievement, where math achievement is usually found to be more responsive to intervention. We present results from three estimation strategies: Ordinary least squares (OLS, equation 3); reduced form (RF, equation 4); and two-stage least squares (2SLS, equation 6)

OLS Estimates:

Columns (1) and (2) present OLS estimates based on equation (3). Full exposure to the reform is associated with an increase in student math achievement of 0.033 student standard deviations, significant at the 0.001 level. Importantly, the relationship between reform exposure and achievement gains *before* the reform is small, insignificant, and of the opposite sign. This implies that reform exposure is not correlated with unobserved factors before the reform and the parallel trends assumption holds. When controlling for experience, the estimate decreases by 30% to 0.024 (still significant at 0.001), suggesting that up to 30 percent of the OLS estimate is mediated by changes in teacher experience composition.

Reduced Form Estimates:

Columns (3) and (4) present the intent-to-treat estimates from the reduced form. Full treatment of a grade is associated with an improvement in math achievement of 0.014 student standard deviations after the reform (significant at 0.010). Here, we cannot estimate the relationship between the instrument and the outcome before the reform because the instrument is constant over time. When we control for experience, the estimate falls by 7 percent, indicating that changes in experience composition mediate at most a small portion of the reduced form effect.

Two-Stage Least Squares Estimates:

Columns (5) and (6) present the 2SLS estimates. The first stage shows that a fully treated school-grade would have, on average, 23 percentage points more treatment exposure in the post-reform period. The 2SLS estimates indicate that full treatment would improve math achievement by 0.043 student standard deviations after the reform (significant at 0.001). These estimates are not statistically different from the OLS estimates. Reassuringly, we find small, opposite signed effects of the reform before it is implemented, implying that the reform was orthogonal to unobserved factors of achievement. Controlling for experience reduces the estimated effect, by about 10 percent, suggesting once again a modest role of experience in driving the effects.

Overall, across all specifications, we find consistent evidence that the reform positively impacted math achievement, with experience composition playing a minor role in mediating the effect.

We next examine its effect on reading achievement. The estimated effects of the reform on reading skills can be found in table A4. These estimates follow a similar pattern to those for math—they are consistently positive—but they are smaller in magnitude. OLS implies a 0.013 student standard deviation effect in reading (significant at the 1 percent level). The RF estimates generate a similar 0.010 student standard deviation effect (significant at 0.010) 2SLS yields a larger but less precise estimate than OLS. These results align with prior work finding that achievement in math is more affected by school interventions than that in reading (Rivkin et al., 2005; Chetty et al., 2014a).

7 Conclusion

In response to mounting pension liabilities, every state in the U.S. has implemented pension reforms to reduce costs by cutting benefits. We examine the effect of a representative reform on worker retention and output. This reform grandfathered some workers based on their age and experience, while subjecting others to a new regime that cut annuities and heavily penalized early retirement. Our study leverages the threshold determining reform exposure to provide careful treatment and control comparisons. Our setting and empirical design offers several advantages.

First, the variation in pension regimes is credibly exogenous, as the policy altered workers' pension based on attributes they could not control. Second, the setting provides a tidy comparison group of similar workers unaffected by the reform, allowing us to compare workers in a shared setting where the only difference is their pension scheme.[25] Third, the public-school context provides large administrative data on retention through staffing records. Fourth, unlike many settings where worker productivity is difficult to observe, schools offer some of the best large-scale productivity measures available in any professional setting.

Theory offers mixed predictions for both the effect of the reform on retention and on output. The reform reduces the continuation value of employment, potentially decreasing retention, but also penalizes early retirement, encouraging retention. Our discontinuity estimates provide an answer, revealing small positive effects on worker retention, particularly in years when the reform makes early retirement less attractive.

Theoretical predictions about the reform's effect on worker output are similarly ambiguous. While classic predictions suggest that effective pay cuts harm morale and depresses effort, alternative theories propose two mechanisms for *increased* worker effort. One, the reform extends the expected career horizon by making early retirement less attractive, encouraging greater investment in workplace skills. Two, the reform reduces the outside options for teachers, which may discourage shirking.

[25]We would encourage future work to examine how pension reforms, like the one considered here, affect the choice of individuals to enter the teaching profession.

We test these hypotheses by exploiting variation in reform exposure across grades within a school. We find that students more exposed to the reform perform better, with fully treated grades scoring 0.03 student standard deviations higher in math and 0.01 student standard deviations higher in reading. Controlling for the mediating influence of experience makes little difference in the estimates, implying the effects stem from increased effort rather than changes to the stock of teacher experience.

The results offer a rare opportunity to examine similar workers working in the same setting who yet were part of different pension regimes. They suggest that carefully designed pension reforms can potentially improve both retention and productivity.

References

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *The Quarterly Journal of Economics*, 97(4):543–569.

Akerlof, G. A. and Yellen, J. L. (1990). The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics*, 105(2):255–283.

Anzia, S. F. (2017). "Pensions in the Trenches: Are Rising City Pension Costs Crowding Out Public Services?". *Working Paper*, pages 1–50.

Aubry, J.-P. and Crawford, C. V. (2017). State and local pension reform since the financial crisis. *State and Local Pension Plans*, (54).

Barnes, G., Crowe, E., and Schaefer, B. (2007). The cost of teacher turnover in five school districts: A pilot study. Technical report, National Commission on Teaching and America's Future.

Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 106(2):407–443.

Bates, M. and Johnston, A. C. (2024). Do pensions enhance worker effort and selection? evidence from public schools. Available at SSRN: `https://ssrn.com/abstract=4811526`.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5, Part 2):9–49.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4, Part 1):352–365.

Bewley, T. F. (1998). Why not cut pay? *European Economic Review*, 42(3-5):459–490.

Biasi, B. (2019). "Higher Salaries or Higher Pensions? Inferring Preferences from Teachers' Retirement Behavior". *Working Paper*.

Brainard, K. and Brown, A. (2018). Significant reforms to state retirement systems. *National Association of State Retirement Administrators*, pages 1–98.

Brown, K. (2013). "The Link Between Pensions and Retirement Timing: Lessons from California Teachers". *Journal of Public Economics*, 98:1–14.

Camerer, C., Babcock, L., Loewenstein, G., and Thaler, R. (1997). Labor supply of New York City cabdrivers: One day at a time. *The Quarterly Journal of Economics*, 112(2):407–441.

Campbell III, C. M. and Kamlani, K. S. (1997). The reasons for wage rigidity: evidence from a survey of firms. *The Quarterly Journal of Economics*, 112(3):759–789.

Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *Quarterly Journal of Economics*, 126(2):749–804.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014b). "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States". *The Quarterly Journal of Economics*, 129(4):1553–1623.

Coile, C. and Gruber, J. (2007). Future social security entitlements and the retirement decision. *The Review of Economics and Statistics*, 89(2):234–246.

Costrell, R. M. and McGee, J. B. (2010). Teacher pension incentives, retirement behavior, and potential for reform in arkansas. *Education Finance and Policy*, 5(4):492–518.

Czernich, N., Falck, O., Krestchmer, T., and Woessmann, L. (2011). "Broadband Infrastructure and Economic Growth". *The Economic Journal*, 212(552):505–532.

Duflo, E. (2001). "Schooling and the Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment". *American Economic Review*, 91:795–813.

Early, B. (2023). Are pensions the answer to the public-sector worker shortage? *Route Fifty*. Director of Membership and Strategic Alliances, NCPERS.

Equable Institute (2023). Unfunded liabilities for state pension plans in 2023. *Equable Original*.

Ericson, K. M. and Laibson, D. (2018). Intertemporal choice. *NBER Working Paper 25358*, pages 1–67.

Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? an experimental investigation. *The Quarterly Journal of Economics*, 108(2):437–459.

Fitzpatrick, M. D. (2015). "How much are public school teachers willing to pay for their retirement benefits?". *American Economic Journal: Economic Policy*, 7(4):165–188.

Fitzpatrick, M. D. (2017). Pension-spiking, free-riding, and the effects of pension reform on teachers' earnings. *Journal of Public Economics*, 148:57–74.

Fitzpatrick, M. D. (2019). Pension reform and return-to-work policies. *Journal of Pension Economics Finance*, 18(4):500–514.

Fitzpatrick, M. D. and Lovenheim, M. F. (2014). Early retirement incentives and student achievement. *American Economic Journal: Economic Policy*, 6(3):120–154.

Friedberg, L. and Webb, A. (2005). "Retirement and the Evolution of Pension Structure". *Journal of Human Resources*, 40(2):281–308.

Fuchsman, D., McGee, J. B., and Zamarro, G. (2020). "Teachers' Willingness to Pay for Retirement Benefits: A National Stated Preferences Experiment.". *Working Paper*.

Gibbons, R. and Murphy, K. J. (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy*, 100(3):468–505.

Gohl, N., Haan, P., Kurz, E., and Weinhardt, F. (2023). Working life and human capital investment: Causal evidence from a pension reform. *Labour Economics*, page 102426.

Goldhaber, D., Grout, C., and Holden, K. (2015). Public pension reform and teacher turnover: Evidence from washington state. National Center for Analysis of Longitudinal Data in Education Research.

Goldhaber, D., Grout, C., and Holden, K. L. (2017). Pension structure and employee turnover: Evidence from a large public pension system. *ILR Review*, 70(4):976–1007.

Gustman, A. L., Mitchell, O. S., and Steinmeier, T. L. (1994). The role of pensions in the labor market: A survey of the literature. *ILR Review*, 47(3):417–438.

Hanif, F., Millard, C., Bass, E. J., Curmi, E., Lam, D., and Pittaway, N. (2016). "The Coming Pensions Crisis: Recommendations for Keeping the Global Pensions System Afloat". pages 1–127.

Haschke, D. N. (2005). "Reflections on the 79th Legislature". *Advocate*, (Summer):2.

Hendricks, M. D. (2014). Does it pay to pay teachers more? evidence from texas. *Journal of Public Economics*, 109:50–63.

Illinois Policy Institute (2024). Illinois public services being cut to pay unsustainable pension cost. Accessed: 2024-07-22.

Imbens, G. and Kalyanaraman, K. (2012). "Optimal Bandwidth Choice for the Regression Discontinuity Estimator". *The Review of Economic Studies*, 79(3):933–959.

Johnston, A. C. (2021). Preferences, selection, and the structure of teacher pay. *Working Paper*.

Keefe, A. (2014). Teachers union president vows to fight cuts to pensions. *WBEZ Chicago*. Chicago Teachers Union president Karen Lewis. AP/File.

Kenneally, K. and Bond, T. (2020). Millennial state & local government employee views on their jobs, compensation & retirement. *Issue Briefs*.

Koedel, C. (2019). "California's Pension Debt is Harming Teachers and Students Now–And It's Going to Get Worse".

Koedel, C., Podgursky, M., and Shi, S. (2013). Teacher pension systems, the composition of the teaching workforce, and teacher quality. *Journal of Policy Analysis and Management*, 32(3):574–596. Working paper version.

Koedel, C. and Xiang, P. B. (2017). Pension enhancements and the retention of public employees. *ILR Review*, 70(2):519–551.

Kube, S., Marechal, M. A., and Puppe, C. (2013). "Do Wage Cuts Damage Work Morale?". *Journal of the European Economic Association*, 11(4):853–870.

Lazear, E. P. (1979a). Why is there mandatory retirement? *Journal of Political Economy*, 87(6):1261–1284.

Lazear, E. P. (1979b). Why is there mandatory retirement? *Journal of Political Economy*, 87(6):1261–1284.

Lazear, E. P. (1986). Retirement from the labor force. *Handbook of labor economics*, 1:305–355.

Legislative Budget Board (2005). Fiscal note, 79th legislative regular session: Sb1691 by duncan. pages 1–4.

Manoli, D. and Weber, A. (2016). "Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions". *American Economic Journal: Economic Policy*, 8(4):160–182.

Mas, A. (2006). Pay, reference points, and police performance. *The Quarterly Journal of Economics*, 121(3):783–821.

McCrary, J. (2008). "Manipulation of the Running Variable in the Regression Discontinuity Design". *Journal of Econometrics*, 142(2):698–714.

Mitchell, O. S. (2020). "Building Better Retirement Systems in the Wake of the Global Pandemic". *No. w27261 National Bureau of Economic Research.*

National Center for Education Statistics (2008). Total number of public school teachers and percentage of public school teachers in a union or employees' association, by state: 1999-2000, 2003-04, and 2007-08.

Ni, S. and Podgursky, M. (2016). How teachers respond to pension system incentives: New estimates and policy applications. *Journal of Labor Economics*, 34(4):1075–1104.

Ni, S., Podgursky, M., and Wang, X. (2020). Teacher pension plan incentives, retirement decisions, and workforce quality. *Journal of Human Resources*, pages 1218–1232.

Ni, S., Podgursky, M., and Wang, X. (2021). Teacher pension enhancements and staffing in an urban school district. *Working Paper 21-01*.

Papay, J. P. and Kraft, M. A. (2015a). "Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement". *Journal of Public Economics*, 130:105–119.

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

Petek, N. and Pope, N. G. (2023). The multidimensional impact of teachers on students. *Journal of Political Economy*, 131(4):1057–1107.

Quinby, L. D., Sanzenbacher, G. T., and Aubry, J.-P. (2018). How have pension cuts affected public sector competitiveness? *Center for Retirement Research*, (59).

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458.

Sabatosa, J. (2021). Teachers oppose pension plan. *Rutland Herald*.

Shapiro, C. and Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *The American Economic Review*, 74(3):433–444.

Stock, J. H. and Wise, D. A. (1990). Pensions, the option value of work, and retirement. *Econometrica*, 58(5):1151–1180.

Stock, J. H. and Yogo, M. (2005). Testing for weak instruments in linear iv regression. In Andrews, D. W. K. and Stock, J. H., editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, chapter 5, pages 80–108. Cambridge University Press, Cambridge.

Teacher Retirement System of Texas (2019). Percentages to be applied to amount of standard service retirement benefit.

Texas State Teachers' Association (2005). "Changes in Retirement". *Advocate*, (Summer):10–11.

Van Hollen, C. (2017). Van hollen urges senate leadership to protect federal employee retirement benefits. Accessed: 2024-07-22.

Watlington, E., Shockley, R., Guglielmino, P., and Felsher, R. (2010). The high cost of leaving: An analysis of the cost of teacher turnover. *Journal of Education Finance*, pages 22–37.

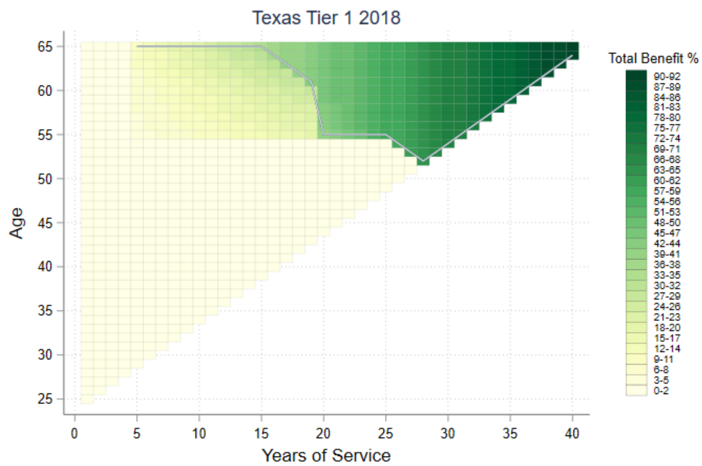
Weakliem, D. L. and Frenkel, S. J. (2006). "Morale and Workplace Performance". *Work and Occupations*, 33(3):335–361.

Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, 100:61–78.

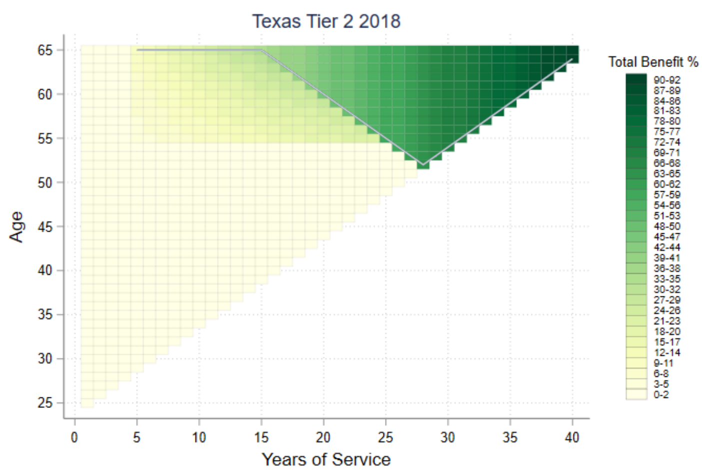
Woods, D. (2010). Public sector spending cuts will decimate staff wellbeing and morale. *HR Magazine*. Accessed: 2024-07-22.

World Economic Forum (2017). "We'll Live to 100–How Can We Afford It?". *White Paper*, pages 1–24.

Figure 1: Pension Generosity by Age and Years of Service, Pre- and Post-Reform



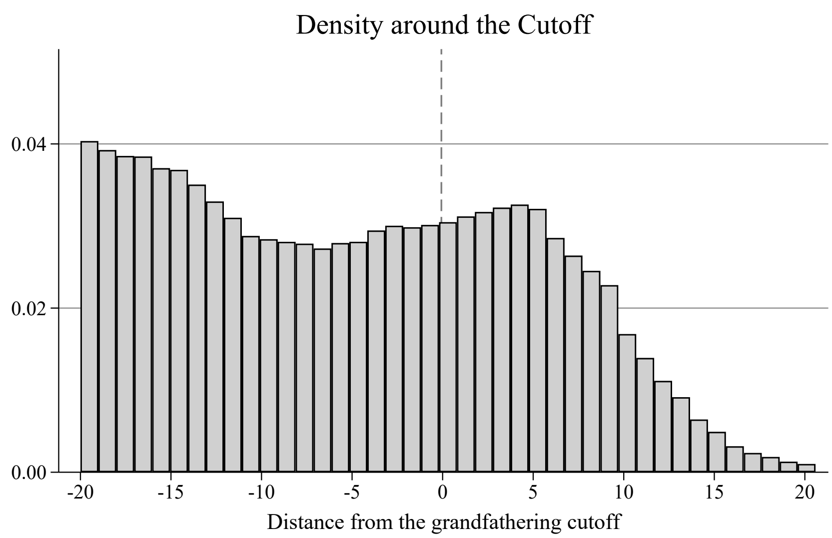
(a) Pension Generosity Pre-Reform



(b) Pension Generosity Post-Reform

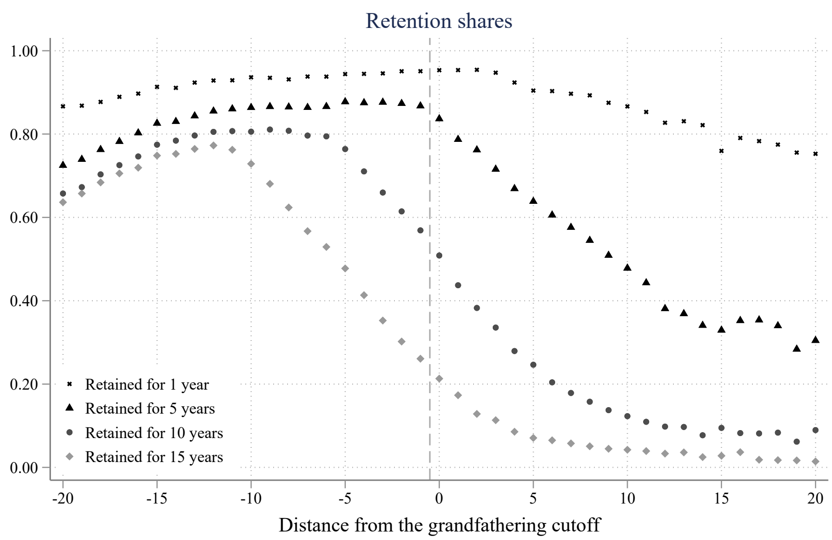
Notes: These heat maps illustrate pension annuity generosity before and after the reform. The x-axis represents years of service (0-40), the y-axis represents age (25-65), and the color intensity indicates the pension annuity generosity as a share of final salary (0-89%). Darker green areas represent age-experience combinations with higher pension annuities. Before the reform, early retirement was generous with minimal penalties for claiming early at age 55 with 20 years of experience. After the reform, early retirements were heavily penalized, as evident from the reduced green. The grey line represents the optimal retirement age to maximize the present value of the lifetime pension, conditional on years of service. Pre-reform, teachers with 20–25 years of experience maximized benefits by claiming at age 55. After, they maximized by waiting until full-retirement. The upside-down pyramid shape of pension generosity results from the rule of 80, where age plus years of service must equal 80 for full retirement. This creates a diagonal boundary where more years of service reduce the required retirement age.

Figure 2: Teacher Density around the Pension Grandfathering Cutoff



Note: This figure presents the density of teachers around the grandfathering cutoff for the pension reform. The x-axis represents the distance from the cutoff in years, with negative values indicating teachers below the cutoff (not grandfathered) and positive values above (grandfathered). The y-axis shows the density of teachers at each point. An excess mass on the favorable side of the discontinuity would signal manipulation of the running variable, potentially compromising the validity of the regression discontinuity design. We detect no such manipulation, as evidenced by the smooth distribution around the cutoff. Source: Administrative data from Texas Education Agency.

Figure 3: Teacher Retention Rates Relative to Grandfathering Cutoff



Note: This figure illustrates the retention rates of public school teachers in Texas over various time horizons following the reform. The x-axis represents the distance from the cutoff in years, with negative values indicating teachers below the cutoff (not grandfathered) and positive values above (grandfathered). The vertical dashed line at 0 marks the cutoff point. The y-axis shows the share of teachers retained. Each dot represents the average retention rate for teachers at a specific distance from the cutoff. The four series correspond to different retention periods: 1 year (top series), 5 years, 10 years, and 15 years (bottom line), showing lower retention rates over longer periods. Source: Administrative data from Texas Education Agency.

Table 1: Summary Statistics of Teacher-Level Variables for Regression Discontinuity Analysis of Teacher Retention

	(1) **Mean**	(2) **SD**	(3) **N**
Age	42.19	11.19	309,853
Experience	11.45	9.663	309,853
Here in 1yr	0.907	0.291	309,853
Here in 5yrs	0.746	0.435	309,853
Here in 10yrs	0.565	0.496	309,853
Here in 15yrs	0.439	0.496	309,853
Distance	-6.596	10.69	309,853
Grandfathered	0.327	0.469	309,853
Teacher	1.000	0.000	309,853
Base Pay	$41,503	$8,433	309,853
Other Pay	$1,366	$2,510	309,853
Bachelors	0.942	0.234	309,853
Masters	0.232	0.422	309,853

Notes: This table presents summary statistics for the analytic sample, focusing on public school teachers in Texas observed in the year of the pension reform. Column (1) shows the mean values, column (2) the standard deviations, and column (3) the number of observations. The sample includes 309,853 teachers, with variables covering demographics, retention rates over various time horizons, pension reform status, and educational qualifications. "Distance" refers to the teacher's position relative to the grandfathering cutoff, with negative values indicating teachers not grandfathered into the old pension system. Source: Administrative data from Texas Education Agency.

Table 2: Regression Discontinuity Estimates of Retention Effects among Teachers

	(1) Retained ≥1 year	(2) Retained ≥5 years	(3) Retained ≥10 years	(4) Retained ≥15 years	(5) Average Ret. Rate
Linear Controls					
Reform (RDD)	-0.002***	0.029***	0.019**	0.004	0.0102***
	(0.000)	(0.005)	(0.006)	(0.004)	(0.0014)
Bandwidth	2.76	3.72	3.67	3.91	3.43
Quadratic Controls					
Reform (RDD)	0.001	0.024*	0.007	-0.002	0.0034
	(0.006)	(0.008)	(0.010)	(0.010)	(0.0060)
Bandwidth	5.52	7.44	7.33	7.83	6.86
Mean DV (at cutoff)	0.953	0.859	0.513	0.211	0.623
Observations	309,860	309,860	309,860	309,860	309,860

Notes: This table presents the regression discontinuity design (RDD) estimates of the effect of pension reform on teacher retention. The analysis examines retention rates for periods of ≥1, ≥5, ≥10, and ≥15 years, as well as the average retention rate overall. We use a triangular kernel. Results are shown for both linear and quadratic control functions, using IK-optimal bandwidth for linear and twice-optimal for quadratic specifications. We present the constant (Mean DV at cutoff) so that the reader can gauge the size of each effect relative to the counterfactual. + $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 3: Summary Statistics of Student-Level Variables for Difference-in-Differences Analysis of Teacher Productivity

Variable	Mean	SD	N
observation year	2006.6	3.335	11,300,719
math percentile	50.87	28.54	11,300,719
read percentile	49.64	28.76	8,199,421
math percentile, z-score	0.000	1.000	11,300,719
read percentile, z-score	-0.00	1.000	8,199,421
reform exposure - math	0.530	0.331	11,300,719
reform exposure - reading	0.518	0.314	8,199,421
exposure instrument - math	0.530	0.369	11,300,719
exposure instrument - reading	0.489	0.356	8,199,421
student grade	6.421	1.369	11,300,719
female student share	0.503	0.500	11,300,719
black student share	0.140	0.347	11,300,719
white student share	0.392	0.488	11,300,719
Hispanic student share	0.429	0.495	11,300,719
low-income share	0.592	0.492	11,300,719
English as a Second Language share	0.054	0.226	11,300,719

Note: This table presents summary statistics for student-level variables used in the difference-in-differences analysis of teacher productivity following the 2005 Texas pension reform. "Reform exposure" refers to the share teachers in a student's school-grade affected by the reform. "Exposure instrument" is based on pre-reform staff compositions in the student's school-grade. Percentile scores are within-grade ranks based on state-wide test performance. Z-scores are standardized versions of these percentiles. Demographic variables refer to individual student characteristics. N represents the number of student-year observations. The smaller sample size for reading scores is due to some grades not administering standardized reading tests.

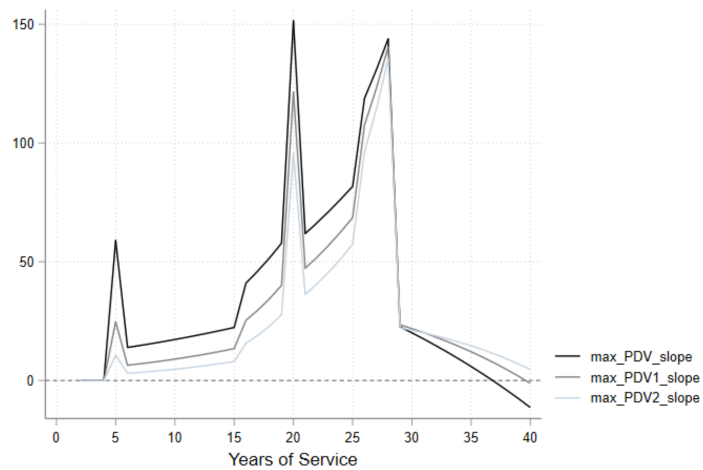
Table 4: Triple-Difference Estimates of Worker Output – Math Achievement

	Math achievement (student standard deviations)					
	OLS (1)	OLS (2)	RF (3)	RF (4)	2SLS (5)	2SLS (6)
Reform exposure x post	0.033***	0.024***			0.043***	0.039***
	(0.005)	(0.005)			(0.009)	(0.010)
Reform exposure	-0.005	-0.005			-0.011	-0.014*
	(0.004)	(0.004)			(0.006)	(0.007)
Instrument x post			0.014**	0.014**		
			(0.005)	(0.005)		
First stage					0.232***	0.227***
					(0.0004)	(0.0004)
Adjusted R-squared	0.658	0.658	0.658	0.658	0.658	0.658
Observations	11,300,719	11,300,719	11,300,719	11,300,719	11,300,719	11,300,719
Lag-score polynomial	X	X	X	X	X	X
School-year FE	X	X	X	X	X	X
School-grade FE	X	X	X	X	X	X
Grade-year FE	X	X	X	X	X	X
Experience controls		X		X		X

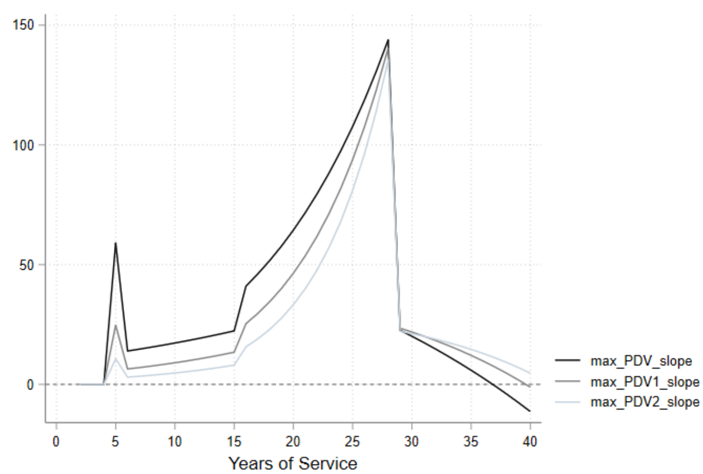
Notes: This table presents triple-difference estimates of the pension reform's effect on worker output, measured by student gains in math achievement. (Estimates for reading can be found in Appendix Table A4.) The outcome variable is normalized student math scores (mean zero, standard deviation one), and the treatment variable describes what share of a school-grade cell was exposed to the reform. The controls in these specifications include a cubic polynomial of the student's past math scores and demanding triple-difference controls (school-year FE, school-grade FE, and grade-year FE). We also present estimates controlling for experience, to show how much of the output effect is explained by retention. Columns (1) and (2) show OLS estimates, regressing student achievement on student's reform exposure × post. Due to potential endogeneity in teacher grade assignments and retention, we instrument exposure using each school-grade's exposure in 2004-2005, the year before the reform's announcement and implementation. Columns (3) and (4) present reduced form estimates, using the instrument as the treatment measure, providing intent-to-treat estimates. Columns (5) and (6) present two-stage least squares estimates, instrumenting post-reform exposure with pre-reform exposure. Columns (2), (4), and (6) additionally control for experience shares in each school-grade cell over time to investigate the role of retention in mediating achievement effects. These experience controls indicate what fraction of the school-grade's teachers are in each experience decile each year. Standard errors are clustered by school-grade cells. The results suggest positive effects of the reform on math achievement, with the 2SLS estimates showing a significant increase of 0.039–0.043 student standard deviations in student math scores post-reform. These effects persist after controlling for teacher experience, indicating that improved retention does not explain most of the positive impact. Significance levels: * p<0.05, ** p<0.010, *** p<0.001.

A Additional Exhibits

Figure A.1: Pension Wealth Returns to an Additional Year of Service, Pre- and Post-Reform



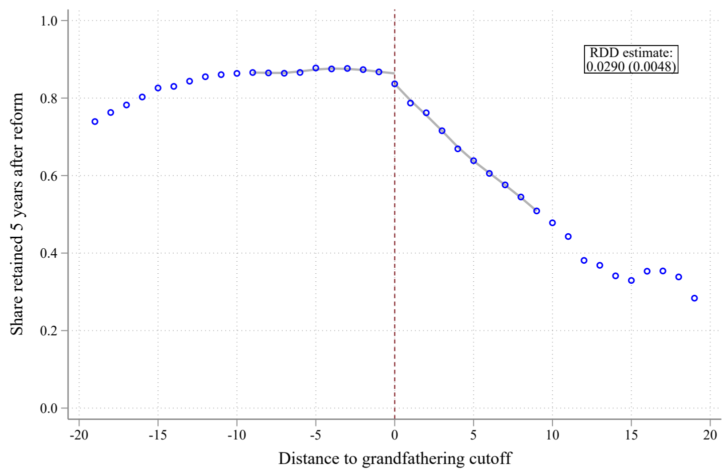
(a) Pension Wealth Returns to Employment Pre-Reform



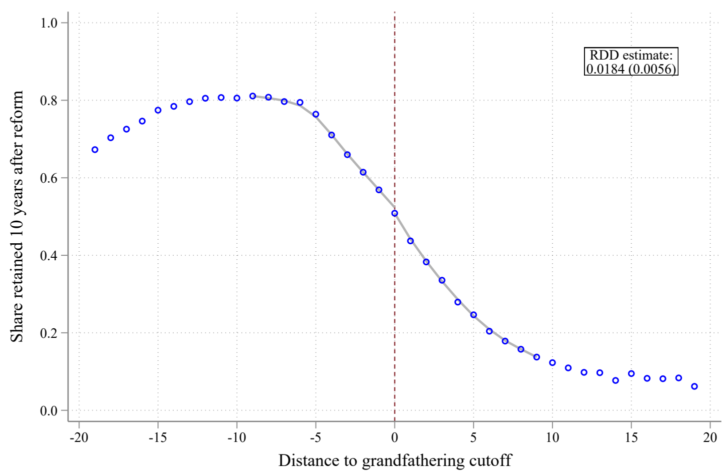
(b) Pension Wealth Returns to Employment Post-Reform

Notes: These figures illustrate the incremental pension wealth accrual for workers from an additional year of service, before and after the pension reform. The x-axis represents years of service (0-40), while the y-axis shows the percentage increase in pension wealth for working from an additional year of employment, denominated as a share of final salary. Three lines are presented, each representing different discount rates used in calculating the present value of pension wealth: the black line represents the returns assuming a 3 percent discount rate (max_PDV_slope); the slate represents a 5 percent rate; the light grey represents a 7 percent rate. The first figure shows the return-to-experience profile pre-reform, where there is a significant spike in pension wealth accrual at 20 years, when workers qualify for early retirement. The second figure shows the return profile after the reform. The early retirement spike has been eliminated.

Figure A.2: Regression Discontinuity Plots of Teacher Retention Rates



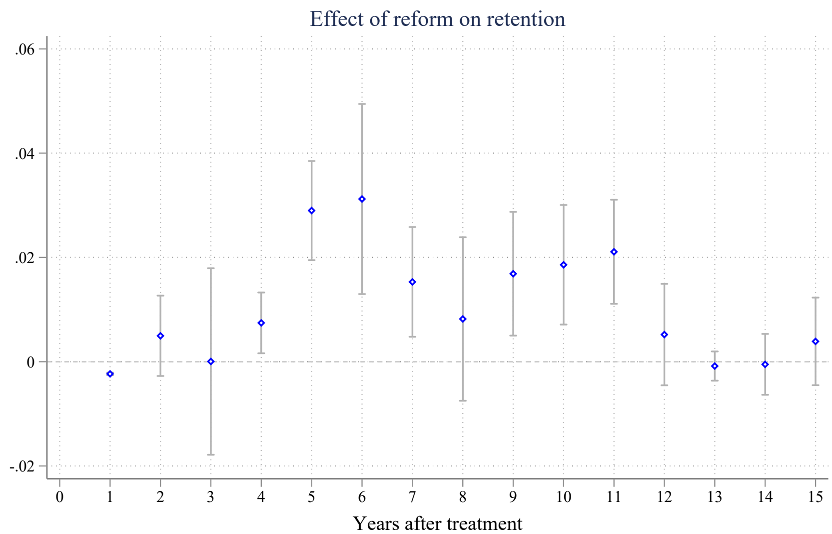
(a) Five-Year Retention



(b) Ten-Year Retention

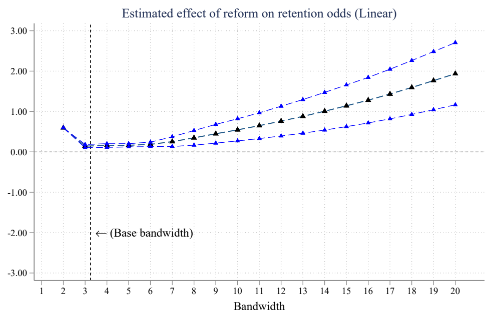
Notes: These figures show the results of a regression discontinuity design (RDD) analysis examining teacher retention rates 5 and 10 years after the reform. The x-axis represents the distance to the grandfathering cutoff, with negative values indicating teachers below the cutoff (not grandfathered) and positive values above (grandfathered). The vertical dashed line at zero marks the cutoff point. The y-axis shows the share of teachers retained at different time intervals. Panel A shows five-year retention rates around the grandfathering cutoff, and Panel B shows the same for ten-year retention.

Figure A.3: Effect over Time

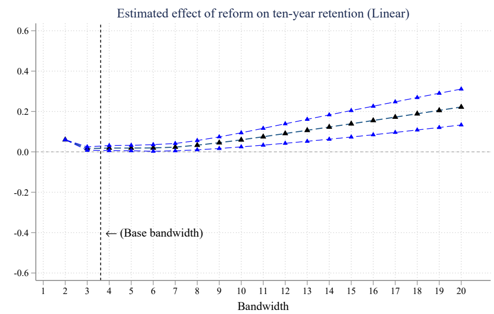


Note: This figure shows the estimated effect of the reform on retention (measured in percentage points) at various years after implementation. We estimate the RDD with a triangular kernel and independent linear terms on either side of the threshold within the optimal bandwidth. Points represent point estimates, with vertical bars indicating 95% confidence intervals. The effects are indistinguishable from zero immediately after the reform, grow to approximately 2 percentage points from five years after the reform through eleven years after the reform, and then shrink back to approximately zero. Source: Administrative data from Texas Education Agency.

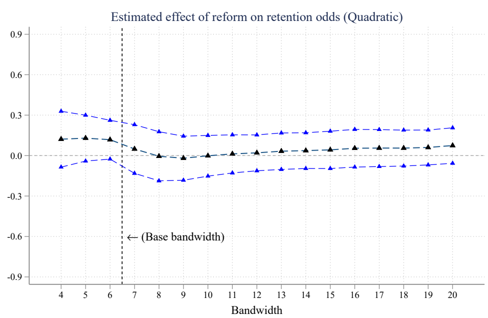
Figure A.4: Assessing Bandwidth Selection among Teachers



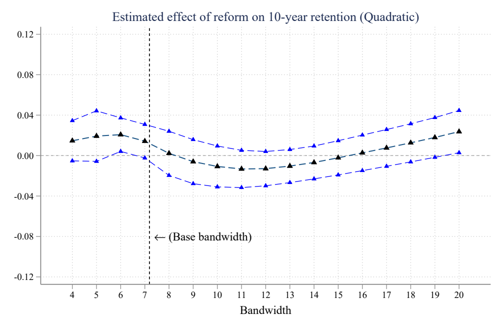
(a) Linear Model, Average Retention



(b) Linear Model, Ten-Year Retention



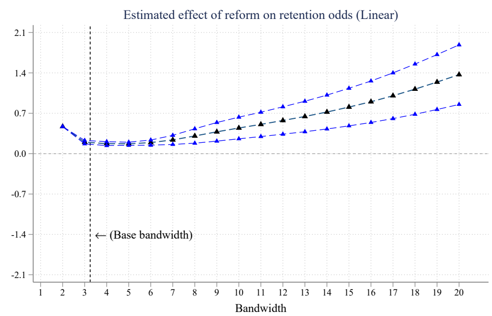
(c) Quadratic Model, Average Retention



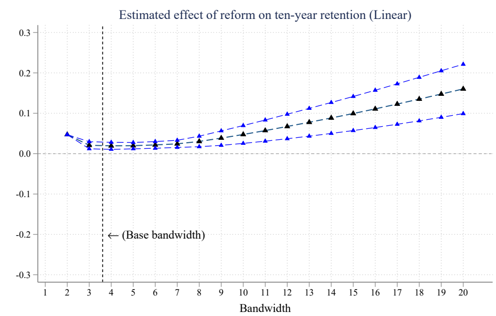
(d) Quadratic Model, Ten-Year Retention

Note: These figures show how the estimated effect of the reform on teacher retention vary with bandwidth choice. The base bandwidth choice is indicated by a vertical dashed line. The black triangles represent point estimates and the blue triangles show confidence intervals. All estimates use a triangular kernel. Panels (a) and (b) use linear terms on either side of the threshold, while (c) and (d) use quadratic. Panels (a) and (c) show effects on average retention odds, while (b) and (d) show effects on ten-year retention rates. The y-axis represents the estimated effect of the reform on the respective retention measure. Linear models show wider variation in estimates across bandwidths. Quadratic models display more consistent estimates, fluctuating around zero for both retention measures. Source: Administrative data from Texas Education Agency.

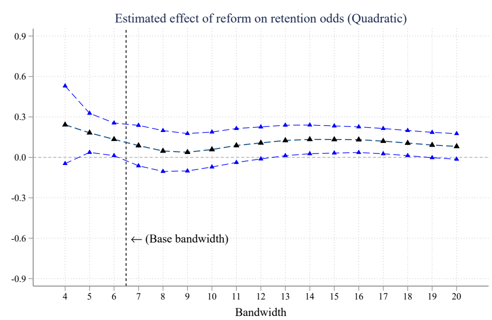
Figure A.5: Assessing Bandwidth Selection among All Workers



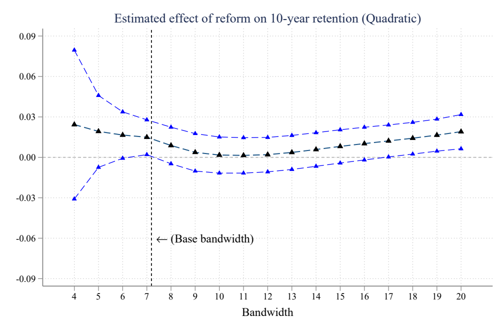
(a) Linear Model, Average Retention



(b) Linear Model, Ten-Year Retention



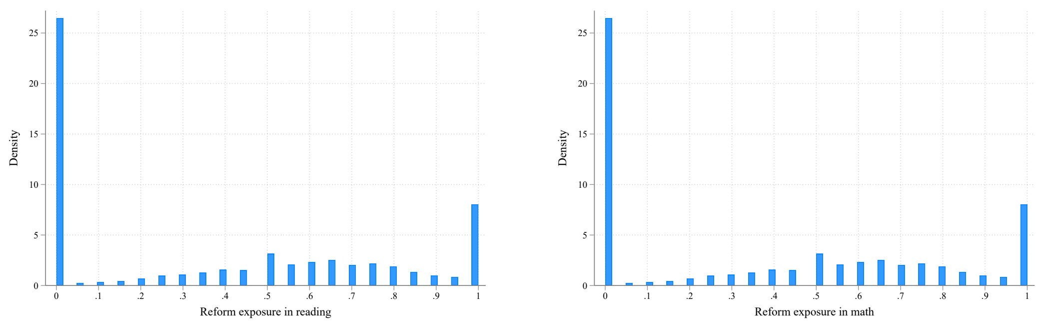
(c) Quadratic Model, Average Retention



(d) Quadratic Model, Ten-Year Retention

Note: These figures show how the estimated effect of the reform on all-worker retention vary with bandwidth choice. The base bandwidth choice is indicated by a vertical dashed line. The black triangles represent point estimates and the blue triangles show confidence intervals. All estimates use a triangular kernel. Panels (a) and (b) use linear terms on either side of the threshold, while (c) and (d) use quadratic. Panels (a) and (c) show effects on average retention odds, while (b) and (d) show effects on ten-year retention rates. The y-axis represents the estimated effect of the reform on the respective retention measure. Linear models show wider variation in estimates across bandwidths. Quadratic models display more consistent estimates, about half of which are significant at the 5% level. For linear models, estimates are consistent for bandwidths 3–7 (average retention) and 3–8 (ten-year retention). Source: Administrative data from Texas Education Agency.

Figure A.6: Distribution of Reform Exposure in Reading and Mathematics



(a) Reading (b) Mathematics

Note: These histograms display the distribution of reform exposure for teachers in reading (left panel) and mathematics (right panel). The x-axis represents the share of a school-grade's teachers exposed to the reform based on 2004 teacher assignments. The values range from 0 (no exposure) to 1 (full career exposure). The y-axis shows the density of teachers at each exposure level. Values are rounded to the nearest 0.05 to prevent small-cell disclosures. Both distributions exhibit similar patterns. Source: Administrative data from Texas Education Agency.

Table A1: Placebo RD and Difference-in-Discontinuities Estimates among Teachers

	(1) Retained ≥1 year	(2) Retained ≥5 years	(3) Retained ≥10 years	(4) Retained ≥15 years	(5) Average Ret. Rate
Treatment RD	-0.002***	0.029***	0.019**	0.004	0.010***
	(0.0001)	(0.005)	(0.006)	(0.004)	(0.001)
Placebo RD	-0.003**	-0.048***	-0.006*	-0.006	0.005**
	(0.001)	(0.010)	(0.003)	(0.004)	(0.002)
Difference-in-RD	0.001	0.077***	0.025***	0.009*	0.005*
	(0.001)	(0.011)	(0.007)	(0.006)	(0.002)

Notes: This table presents regression discontinuity design (RDD) estimates of the effect of pension reform on teacher retention. The columns provide estimates for different periods of retention. The estimates indicate the increased probability associated with reform at the cutoff. The first row shows the baseline estimates. The second row shows placebo estimates when pretending the reform took place in 2000, five years before the policy. In the third row of each panel, we present the difference-in-discontinuity design (DDRD) estimates which is the first row subtracted by the second row. All estimates use a local-linear specification with a triangular kernel. + $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table A2: Summary Statistics of School Employees for Regression Discontinuity Analysis of Worker Retention

	(1) Mean	(2) SD	(3) N
Age	44.20	11.41	620,508
Experience	8.035	9.808	620,508
Here in 1yr	0.883	0.322	620,508
Here in 5yrs	0.696	0.460	620,508
Here in 10yrs	0.508	0.500	620,508
Here in 15yrs	0.377	0.485	620,508
Distance	-4.842	10.83	620,508
Grandfathered	0.373	0.484	620,508
Teacher	0.499	0.500	620,508
Base pay	$33,492	$16,876	620,508
Other pay	$949	$2,076	620,508
Bachelors	0.568	0.495	620,508
Masters	0.182	0.386	620,508

Notes: This table presents summary statistics for the analytic sample, which includes all workers observed in the year of the pension reform (N = 620,508). "Distance" reflects years until pension eligibility, with negative values indicating years before eligibility. "Grandfathered" indicates the share of workers exempted from the new pension regime by grandfathering. "Here in X yrs" describes the retention rate after X years. Pay is reported in nominal dollars. Education levels (Bachelors, Masters) represent the highest degree attained.

Table A3: Regression Discontinuity Estimates of Retention Effects including All Workers

	(1) Retained ≥1 year	(2) Retained ≥5 years	(3) Retained ≥10 years	(4) Retained ≥15 years	(5) Average Ret. Rate
Linear Controls					
Reform (RDD)	-0.001	0.019***	0.019***	0.007	0.0120***
	(0.002)	(0.002)	(0.004)	(0.006)	(0.0011)
Bandwidth	2.77	3.70	3.61	3.60	3.25
Quadratic Controls					
Reform (RDD)	0.000	0.016*	0.012^{+}	0.000	0.0063
	(0.006)	(0.005)	(0.006)	(0.007)	(0.0037)
Bandwidth	5.54	7.39	7.22	7.19	6.49
Mean DV (at cutoff)	0.929	0.809	0.529	0.248	0.549
Observations	620,508	620,508	620,508	620,508	620,508

Notes: This table presents the regression discontinuity design (RDD) estimates of the effect of pension reform on worker retention. Columns (1)-(4) show the effect of being retained for at least 1, 5, 10, and 15 years, respectively. Column (5) shows the effect on average retention rates. We use the IK-optimal bandwidth and a triangular kernel in the linear specification, and twice the optimal bandwidth for the quadratic specification. We present the constant so the reader can gauge the size of each effect relative to the counterfactual. Other coefficients are omitted to spare clutter. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A4: Difference-in-Differences Estimates of Output Effects – Reading Achievement

	Reading achievement (student standard deviations)					
	OLS (1)	OLS (2)	RF (3)	RF (4)	2SLS (5)	2SLS (6)
Reform exposure x post	0.013** (0.004)	0.012** (0.004)			0.026 (0.009)	0.027 (0.009)
Reform exposure	-0.013 (0.003)	-0.007* (0.003)			-0.015 (0.006)	-0.015* (0.006)
Instrument x post			0.010** (0.004)	0.011** (0.004)		
First stage					0.239*** (0.0005)	0.239*** (0.0005)
Adjusted R-squared	0.581	0.582	0.581	0.582	0.581	0.582
Observations	8,199,421	8,199,421	8,199,421	8,199,421	8,199,421	8,199,421
Lag-score polynomial	X	X	X	X	X	X
School-year FE	X	X	X	X	X	X
School-grade FE	X	X	X	X	X	X
Grade-year FE	X	X	X	X	X	X
Experience controls		X		X		X

Notes: This table presents triple-difference estimates of the pension reform's effect on worker output, measured by student gains in reading achievement. (Estimates for math can be found in Table 4.) The outcome variable is normalized student reading scores (mean zero, standard deviation one), and the treatment variable describes what share of a school-grade cell was exposed to the reform. The controls in these specifications include a cubic polynomial of the student's past reading scores and demanding triple-difference controls (school-year FE, school-grade FE, and grade-year FE). We also present estimates controlling for experience, to show how much of the output effect is explained by retention. Columns (1) and (2) show OLS estimates, regressing student achievement on student's reform exposure × post. Due to potential endogeneity in teacher grade assignments and retention, we instrument exposure using each school-grade's exposure in 2004-2005, the year before the reform's announcement and implementation. Columns (3) and (4) present reduced form estimates, using the instrument as the treatment measure, providing intent-to-treat estimates. Columns (5) and (6) present two-stage least squares estimates, instrumenting post-reform exposure with pre-reform exposure. Columns (2), (4), and (6) additionally control for experience shares in each school-grade cell over time to investigate the role of retention in mediating achievement effects. These experience controls indicate what fraction of the school-grade's teachers are in each experience decile each year. Standard errors are clustered by school-grade cells. The results suggest usually significant effects of the reform on reading achievement, with the OLS estimates showing a significant increase of 0.012-0.013 student standard deviations in student reading scores post-reform. These effects change little when controlling for teacher experience, indicating that improved retention does not explain most of the positive impact. Significance levels: * p<0.05, ** p<0.010, *** p<0.001.