Pensions and Teacher Quality:

Evidence from a Return-to-Work Policy in North Carolina

Rachel Jarrold-Grapes*
Syracuse University

Patten Priestley Mahler†
Centre College

July 7, 2022

Abstract

We examine one impact of pension incentives on teacher quality by analyzing a return-to-work policy in North Carolina that effectively removed the "push" incentives embedded in teacher pensions by allowing them to tap into their pension while teaching. Using administrative public-school data from the North Carolina Research Data Center, we estimate the impact of teachers who returned to work after retirement on student outcomes. We develop an instrumental variable identification strategy centered on the cancellation of the policy. We find small improvements in both reading and math achievement (2 percent of a standard deviation for the former and 3.6 percent of a standard deviation for the latter) for students in the same school who had one of these teachers in their grade during the policy relative to students who did not. We also find that RTW teachers are particularly helpful for students in the top quartile of the ability distribution in math and students in grades 4-6 in reading. Additionally, RTW teachers appear to be good at managing student behavior, which could help explain the positive effects on achievement. Overall, the results suggest that schools are losing effective teachers because of pension incentives and that return-to-work policies may be a way to retain them.

---

* Written for completion of the Doctor of Philosophy in Economics at Syracuse University.
† Associate Professor of Economics and Business.
Acknowledgements: We would like to thank Gary Engelhardt, Amy Ellen Schwartz, Maria Zhu, and Ying Shi (Syracuse University), and Jim Wyckoff (University of Virginia) for their helpful comments and suggestions. We would also like to thank participants at the Education and Social Policy Workshop at Syracuse University and the Association for Education Finance and Policy conferences in 2017 and 2019 for their feedback.

1 Introduction

It is well-established in the retirement literature that teachers respond predictably to pension incentives. Most teacher pensions are defined-benefit (DB) plans in which teachers are paid a percentage of their salary each year in retirement once they reach eligibility (pass age and experience thresholds). This structure incentivizes teachers to work until they are eligible for retirement, "pulling" them to stay, and retire soon after eligibility by "pushing" them out. While actual teacher retirement patterns generally align with this theory, we know less about how the incentives generated by pensions affect teacher quality and subsequently student outcomes. For example, pull incentives may lead to the retention of lower-quality teachers, while push incentives may induce higher-quality teachers to exit sooner than they would have otherwise—or vice versa. The impact of pensions on workforce quality may affect student outcomes and, thus, be an important contribution to discussions of pension reform.

The main difficulty in unraveling the impact of pension incentives is determining which teachers would have exited sooner or later in their absence. Only the timing of retirement is observed, not teacher preferences that could predict their behavior under alternative systems. Some existing papers infer preferences from retirement behavior, while others analyze changes in behavior as a result of pension changes. Most of these papers examine policies that manipulate pull incentives, like early retirement incentive programs, and see if take-up patterns are different by teacher quality. In this paper, we analyze a policy that effectively removed the *push* incentives and determine whether high- or low-quality teachers prefer a later retirement.

Specifically, we study a return-to-work (RTW) policy in North Carolina that allowed retired teachers to return full-time and receive unreduced pension benefits *and* their full-time salary concurrently. Since teachers could return almost immediately after retirement (a short break was required), there was effectively no longer a pension push, and the teachers who chose to return were likely those who would have kept teaching in absence of push incentives. In order to determine whether these teachers were high- or low-quality, we estimate their impact on student outcomes. To do so, we use rich, administrative data on teachers and students from North Carolina and identify our effects using exogenous variation in the timing of the policy.

The primary challenge to identification is that the assignment of RTW teachers to schools and classrooms was not random. There is likely unobserved heterogeneity in the types of schools that chose to hire RTW teachers, such as administrator preferences or school policies related to hiring or student achievement. It is also possible that administrators may have assigned RTW teachers to classrooms in a way that is endogenous to student outcomes within schools, such as putting them in classes with low-performing students to improve achievement. To address this, we instrument for the assignment of RTW teachers to grades within schools over time. We predict the probability of having a RTW teacher using observable school and grade characteristics, as well as fixed effects to control for unobserved heterogeneity across schools and grades. We use this predicted probability as our instrument for actual assignment and leverage exogenous variation from the discontinuation of the policy to estimate the impact of RTW teachers on test scores. We compare test scores of students in the same school with a high probability of having a RTW teacher in their grade to those with a low probability before versus after the end of the policy.

We find that RTW teachers had a statistically significant, but small positive effect on reading and math achievement. Within schools, students who had a RTW teacher in their grade had reading score gains that were 2 percent of a standard deviation higher than those who did not during the policy. We also find that math achievement increased by 3.6 percent of a standard deviation. Additionally, we look at heterogeneity by student ability and grade level. We find that students in the top ability quartile in math (as measured by the distribution of the previous test score) who had a RTW teacher performed better than students in the bottom three quartiles, and that RTW teachers had larger effects in reading for grades 4-6 compared to 7 and 8. Overall, these results suggest that pensions incentivize high-quality teachers to exit earlier than they otherwise would, and that student achievement could improve slightly in absence of push incentives. It also suggests that RTW policies can incentivize effective teachers back into the teaching workforce and modestly improve average workforce quality at the schools in which they are hired. These results are of particular interest because North Carolina adopted a similar RTW policy in 2019. Though we do not directly study this new policy, our results suggest that it could have small positive effects on the quality of the teaching workforce.

The rest of our paper is organized as follows. We summarize the previous research on teacher retirement and quality in the following section. In section 3, we provide information on North Carolina's RTW policy. We discuss the data in section 4, including our sample selection process and descriptive statistics comparing students who did and did not have RTW teachers. We then present our empirical strategy, results, and robustness checks in sections 5 and 6. We discuss heterogeneity and possible mechanisms in section 7 and then end with a brief conclusion.

2 Previous Literature

Our paper is related to the existing literature on teacher retirement, which primarily focuses on how DB plans influence teacher retirement decisions. DB plans generate peaks in the pension-accrual profile. As accruals climb towards the peak, the value of working one more year is higher than the benefit of retirement, which incentivizes teachers to remain employed. In the teacher retirement literature these are called "pull" incentives because they pull teachers toward staying. After accruals peak and start to decline, particularly when they become negative, the value of working another year is lower than the benefit of retiring immediately, leading teachers to retire. These are called “push” incentives since they push teachers to exit. Previous research shows that teachers generally stay until their pension accruals peak and then retire soon after. A summary of this research is in Koedel and Podgursky (2016) and includes Costrell and Podgursky (2009), Costrell and McGee (2010), Friedberg and Turner (2010), and Ni and Podgursky (2016), among others.[3] More recent research looks at the impact of changes in return-to-work policies and pension benefit formulas on retirement behavior. For example, Fitzpatrick (2019) examines how state employees in Illinois, including teachers, responded to an increase in the number of hours they could work post-retirement and still receive their full pension benefits. This likely increased the incentive to exit, but she does not find a change in retirement behavior. Ni, Podgursky, and Wang (2021) describes how retirement behavior in St. Louis public schools changed after the replacement rate increased and a cap on annual benefits was introduced in 1999. They show that pension wealth increased and that accruals both peaked and fell earlier, creating a stronger pension push and leading to earlier retirements.

A subset of the teacher retirement literature looks at the impact of retirement on workforce composition and quality. Of particular concern is whether high- and low-quality teachers respond differently to pension incentives in a way that affects the overall quality of the teaching workforce.[4]

[3] The option value model of retirement was proposed by Stock and Wise (1990). Samwick (1998) showed that higher pension accruals and higher option values decrease the probability of retirement. Coile and Gruber (2000 and 2001) introduced the peak value model of retirement, where peak value is the difference between the pension wealth in the current year and the maximum expected value of pension wealth. Like the option value model, workers continue working if their peak values are high and retire if they are low. Asch, et al. (2005) apply both option and peak value models and both show that the probability of retirement falls when expected pension wealth rises.

[4] Ippolito (1997) discusses pensions and workforce quality in the broader labor market. He proposes that firms with DB plans attract forward-looking workers with low internal discount rates, and that these workers are higher quality than those with high discount rates. He argues that low discounters are better workers because they value future benefits and thus perform well in the present to maximize their pensions. Though firms with DB plans are more attractive to low discounters, they can still attract high discounters who may be incentivized to exit the firm later than they would a firm with an alternative pension plan, decreasing workforce quality.

Koedel, Podgursky, and Shi (2013) analyze the impact of DB pension plans on workforce quality using variation from a positive exogenous shock to pension wealth in Missouri. They identify teachers who were likely incentivized by the pension "pull," had a "regular" retirement, or were incentivized by the pension "push." Teachers influenced by the "pull" incentives were those who retired immediately upon reaching retirement age. "Regular" retirees were teachers who, upon reaching retirement, worked only a couple more years and then retired. Those who had to be "pushed out" kept teaching for several years after becoming eligible for retirement. They compare these groups of teachers using value-added models of student achievement gains and find little difference in the quality of teachers across groups. One exception is that teachers who were likely incentivized by the pension "pull" were less effective than "regular" retirees and just as effective as novices in math.

Ni, Podgursky, and Wang (2020) and Kim, et al. (2021) use structural models of retirement to simulate teacher responses to pension changes, paying particular attention to the implications for workforce quality. Using data from Tennessee, Ni, Podgursky, and Wang (2020) find that high-quality teachers are less likely to retire than low-quality teachers at the same age and experience levels. Teacher quality is determined by classroom evaluations, student test-score growth as measured by value added, and student achievement. They simulate how high- and low- quality teachers would react to different pension changes, including late-career bonuses. They find that bonuses given to high-quality teachers in high-poverty schools would incentivize these teachers to postpone retirement, which would benefit high-need students at a relatively low cost. Kim, et al. (2021) simulate the effects of late-career bonuses and deferred retirement plans on teacher retirement decisions. Their findings suggest that both policies would increase the number of years senior teachers work. The authors argue that the benefits of delayed retirement outweigh the costs if these teachers work in STEM classes or low-performing schools.

Two studies look at the impact of pension incentives on workforce quality indirectly by analyzing the effects of retirements on student outcomes. First, Fitzpatrick and Lovenheim (2014) analyze the impact of an early retirement incentive (ERI) program on student achievement. They use school-level data from Illinois and a difference-in-differences strategy to estimate how ERIs impact test scores by comparing schools with many highly experienced teachers to those with few highly experienced teachers before and after program implementation. They find little change in test scores overall, and some evidence of an increase in test scores in disadvantaged and low-performing schools, especially in reading. They show that this positive impact is driven, at least in part, by the replacement of teachers who left with other experienced teachers rather than novices. Second, Williams (2015) studies whether student achievement was affected by ERIs in California and finds that test scores improved, particularly for high school students. These papers reach similar conclusions that teachers who respond to ERIs are lower quality than other highly experienced teachers, meaning that ERIs increase overall workforce quality by incentivizing lower quality teachers to leave. Another way to interpret this is that traditional pensions are incentivizing lower quality teachers to stay longer than they otherwise would, lowering the overall quality of the teaching workforce.

3 North Carolina Context

3.1 Retirement Benefits

We contribute to this literature by assessing the impact of a policy that removed push incentives on student test scores. We study North Carolina where teachers are incentivized to retire at a relatively young age because of the state's DB plan. Those who are 65 years old with five years of membership service (i.e., five years with the Teachers' and State Employees' Retirement System), 60 years old with 25 years of service, or those with 30 years of service (at any age) can receive their full pension benefits immediately upon retirement. Early retirement with reduced benefits is also an option.[5] Annual pension benefits are calculated by multiplying the average salary during the four highest-paying consecutive years of teaching ($\bar{S}$) by the number of years and months of creditable service (Y) and a retirement factor set by the North Carolina General Assembly (1.82%), as shown in equation (1) (Folwell, 2019).

$$Annual\ Benefit = .0182 * \bar{S} * Y \tag{1}$$

Retirement decisions are typically based not on the annual benefit but on the value of the entire stream of benefits one will receive after retirement, or pension wealth. A teacher's pension wealth is defined as the discounted expected value of her annuities from the year she exits teaching to the year she dies. The pension wealth for a teacher who exits in year *t* is shown below in equation (2), where β^{L-t} is the discount rate of time, $\theta_{L|t}$ is the probability that the teacher is alive in the current year given she was living when she exited teaching, and *Annual Benefit* is the annual pension benefit the teacher receives after exiting in year *t* from equation (1).

$$Pension\ Wealth_t = \sum_{L=t}^{T} \beta^{L-t} \theta_{L|t} (Annual\ Benefit)_t \tag{2}$$

Figure 1 shows the pension wealth at different exit ages for a non-Hispanic white female with a bachelor's degree who began teaching at age 22 and earned the median salary of teachers with her

[5] A teacher qualifies for early retirement either at age 50 and 20 years of experience, or age 60 and 5 years of experience. Early retirement benefits are calculated using equation (1) but are further multiplied by a reduction percentage that is related to a teacher's age and years of experience at retirement. For example, a teacher who is 60 years old and has worked less than 25 years (and at least 5 years) when she retires receives 85% of her annual benefit (Folwell, 2019).

level of experience in her last year of teaching, which we assume is the 2006-07 school year. We let β be 0.95 and calculate θ using the probability of a non-Hispanic white female dying between different ages in 2007.[6] We also assume that the teacher will live until 81 years old, the average life expectancy for a non-Hispanic white female born in 2007 (Arias, 2011). As this figure shows, pension wealth increases rapidly as the teacher nears retirement eligibility at age 52, then slows and starts declining around age 60.[7] Figure 2 shows this teacher's accrual profile, i.e., how much her pension wealth would change if she were to work another year as a percentage of her salary. Her pension accrual increases until its peak at age 52 and then rapidly declines and becomes negative when she reaches age 60. In other words, the benefit of working another year increases until she is eligible for retirement, incentivizing her to keep teaching until age 52. After this point, the benefit of working additional years drops dramatically, incentivizing her to retire.

3.2 Return-to-Work (RTW) Policy

North Carolina implemented a RTW Policy in 1999 to combat a potential shortage of teachers in the labor market caused by the retirement of the large cohort of Baby Boomers. Before and after RTW, if retired teachers returned to a full-time teaching position, their pension benefits and health insurance coverage from the retirement system would be suspended.[8] If instead they returned to a part-time position, they could keep collecting health and retirement benefits as long

---

[6] $\beta < 1$ means that the teacher weights the benefits received sooner more than the benefits received later.

[7] According to teacher salary schedules, a teacher's salary peaks at 30 years of experience and stays constant for the remainder of her tenure. Thus, the decrease in pension wealth is a result of $\bar{S}$ growing at a slower rate and then plateauing when a teacher reaches the last step in the salary schedule.

[8] A second retirement account would be opened for retired teachers who returned full-time (except during the policy). If they worked less than three years before retiring a second time, their first retirement account would be reinstated and they could choose to leave their second account open, withdraw their contributions, or receive a second benefit payment. If they worked at least three years before their second retirement, they could either combine their years of service from both employment spells into one monthly payment or reinstate their first retirement account and withdraw their contributions from the second one. During the RTW policy, retired teachers who returned full-time did not earn retirement benefits for their additional years of service, i.e., their annual benefit remained the same after their time as a RTW teacher.

as their earnings did not exceed a cap of half of their previous full-time salary. RTW raised this salary cap by allowing retirees to receive both their full-time salary and pension benefits concurrently, incentivizing retirees back to the full-time workforce.

The policy was originally set to expire in 2003 but was extended multiple times (to 2004, 2005, 2007, and 2009) until it ultimately expired in the fall of 2009. During this time, the policy underwent several revisions, as seen in Figure 3. For the first year, teachers could only return to low-performing schools in places with shortages of teachers in their certification areas. They were also only allowed to return as interim instructors or substitutes, not permanent teachers. These restrictions were lifted in June of 2000. Additionally, for the first two years, teachers were required to take a one-year break in full-time employment before coming back in order to comply with the IRS's definition of retirement. This was reduced to only six months beginning in 2001. Lastly, after October 2007, retirees could only return if they were eligible for normal retirement, meaning they could not retire with reduced benefits just to return under the policy.[9] See Table A1 in the appendix for a more detailed timeline.

The policy incentivized teachers to return by eliminating the pension push. Take the teacher from section 3.1, for example. In absence of the policy, at 60 years old, she could keep working and earn her full-time salary, but at the cost of a decline in her pension wealth. In this case, it is better for her to retire and claim her annuity than it is for her to keep teaching. During the policy, however, it is actually better for her to keep teaching past 60 years old. She can earn both her full-time salary and collect her annuity after she retires and returns at no cost to her pension wealth.[10]

[9] Information about the policy is found in North Carolina General Assembly Legislation: S.L. 1998-212, S.L. 1998-217, S.L. 2000-67, S.L. 2001-424, S.L. 2002-126, SL. 2004-124, S.L. 2005-144, S.L. 2005-276, S.L. 2005-345, S.L. 2007-145, S.L. 2007-326.

[10] She could retire and return if the policy was not in place, but she would not be able to work full-time without giving up her annuity, making it a less-desirable option than when the policy was in place. North Carolina also discourages teachers from coming back to full-time work after retirement in non-policy years.

For concreteness, say she decided to retire at 62 rather than 60. Prior to the policy, her pension wealth would fall by $2,558.90. If she chose to retire at 60 and then return for a year at 61 during the policy, however, then her pension wealth would only be $0.67 less if she retired at 62 instead of 60. This is because her annuity does not change with additional years of service during the policy. The $0.67 difference is driven by the rate of time preference, β, and the survival probability, θ.

Indeed, we see that the policy brought a significant number of teachers back to work after retirement. Figure 4 shows policy take-up as a proportion of two different groups along with the proportion of retirement-eligible teachers. The gray line is the proportion of RTW teachers out of all teachers for each year between 1996 and 2012. It steadily increases and peaks just under 2% in 2008. The black line shows take-up relative to the number of teachers who were eligible for retirement in the prior year. This is zero before the policy begins in 1999 and increases throughout the policy period until 2009, when just over 35% of previously retirement eligible teachers return to full-time work. The number of RTW teachers drops to zero in 2010 corresponding with the expiration of the policy.[11] The proportion of retirement eligible teachers increases from 4% to 6% during this period, as shown by the dashed line. Because it stays relatively constant, the increase in RTW teachers is likely not being driven by just an increase in those eligible for retirement.

Not only did the policy induce teachers to return after retirement, but it also shifted the timing of retirements. Mahler (2013) shows that teachers were 16% more likely to retire right at eligibility during the policy period than before. She also finds that the number of teachers who worked at least one additional year after becoming eligible for retirement fell by 23% while the

[11] As I describe in the next section, we identify RTW teachers based on their budget codes. While teachers are no longer marked as RTW in the budget codes after the policy ends, there are some who keep working full time. In our analysis sample, of the teachers who came back between 2007 and 2009, 94 kept teaching in 2010, 45 in 2011, and 35 in 2012. This is down from about 400 RTW teachers in the sample for each of 2007-2009.

policy was in place. This suggests that at least some teachers responded strategically to the policy, i.e., retired earlier to collect better benefits.

4 Data and Descriptive Statistics

We use statewide administrative data from the North Carolina Education Research Data Center (NCERDC). The main advantage of these data is that students are linked to their teachers. In our analysis, we use data from the 2006-07 through 2011-12 school years and focus on students in grades 3-8.[12]

The data include scores on end-of-grade (EOG) tests, student characteristics at the time of testing, and course membership. Student characteristics include race, ethnicity, gender, economic disadvantage, and gifted, disability, and limited English proficiency (LEP) status. The course membership data, where students are linked teachers, are at the student-by-class-by-year level. We only use math and reading classes because our outcomes of interest are math and reading test scores.

There are three things to note about our class selection process. First, on average, across all schools and years between 2006-07 and 2011-12, about 20% of elementary school classes are self-contained, where we assume instruction in both math and reading occurs. Second, there are block classes, where multiple subjects are taught together. We know the subject breakdown of these classes and keep only those that include math and/or reading. On average, around 3% and 1% of elementary and middle school classes across all schools in the sample period are in blocks, respectively. Lastly, students can be in more than one reading or math class a year. Students may

[12] We observe students and teachers back through the 1994-95 school year but limit our analysis to years after and including 2006-07 because that is when students can be linked to their actual classroom teachers. Prior to 2006-07, students can be linked to the person who proctored their end-of-grade test, but not their classroom teacher.

take their grade-level math class along with an upper-level one, such as 8th graders taking algebra and geometry simultaneously. For reading, this could be students taking a language arts class and an English elective, like literature or composition. Some students are in their grade-level classes as well as gifted or ESL classes, to name a couple. About 14% of elementary school students took multiple reading classes on average during the sample period. For middle schoolers, this fraction is about 11%. In math, roughly 10% of elementary and 5% of middle schoolers took multiple classes on average between 2006-07 and 2011-12. Since any of these classes can contribute to a student's EOG test scores, we keep them all in our sample, meaning that students can appear multiple times in the sample if they take more than one math or reading class.

We restrict our sample of students to those with both math and reading EOG test scores, as well as test scores from the previous year. This way we know that any differences in the math and reading results are not driven by differences in the sample of students. Also, since testing begins in grade 3, our sample only includes students in grades 4-8 because we need the students' prior year score as a control in the empirical specification.

The teacher data include demographic characteristics, information on their schooling (including the selectivity of their colleges based on the Barron's Admissions Competitiveness Index and the highest degree they earned), their years of teaching experience, and snapshots of their pay each year. Importantly, the pay data includes budget codes that allow us to identify who retired and returned during the policy period. We limit our sample to full-time teachers, since they are the ones who can be influenced by the policy. We also observe school characteristics from the Common Core of Data (CCD). They include enrollment, urbanicity, and student characteristics.

Overall, our sample includes over 350,000 students, 12,000 teachers, and 1,700 schools each year. There are about 400 RTW teachers in each year from 2006-07 through 2008-09. On average, these teachers are 57 years old with 32 years of experience. The non-RTW teachers, in contrast, are 36 years old with 11 years of experience on average. RTW teachers are more likely to have advanced degrees but are also more likely to have gone to less competitive colleges than non-RTW teachers. Additionally, while most teachers in the sample are White women, the RTW teachers are even more likely to be women and more likely to be Black. Over half of RTW teachers return to the school they taught in prior to retirement. Mahler (2013) finds that those who did not go back to the same school went to schools with higher poverty rates.

Figures 5 and 6 compare descriptive statistics by subject for the students in our sample who did and did not have a RTW teacher during the 2008-09 school year, the last year the policy was in place. Students are only included one time in these calculations, even if they take more than one math or reading class, so the summary statistics are at the student-level. The stars indicate statistical significance at the standard levels. Students with RTW teachers had lower prior standardized math and reading test scores, were more likely to be Black and economically disadvantaged, and were less likely to be academically or intellectually gifted. [13] On average, RTW teachers appear to have taught students who would likely benefit from having a highly experienced teacher.[14] Indeed, we find that they had a positive impact on students and explain how we identify this effect in the next section.

5 Empirical Strategy

---

[13] We do not exclude students with missing values in covariates from the summary statistics. If we do, the differences we see remain statistically different from zero. However, some shrink in magnitude, such as math and reading test scores, Black, and economic disadvantage. Also, the average student characteristics in the 2006-07 and 2007-08 school years are similar to these, except for the following: students with a RTW reading teacher were more likely to have changed schools from the previous year and less likely to be categorized as a student with a learning or other disability compared to students who did not have a RTW reading teacher.

[14] There is clear evidence that experienced teachers are more effective than novices. The evidence on whether experience gained after the first five years leads to additional improvement in effectiveness is mixed. See Rockoff (2004); Rivkin, Hanushek, and Kain (2005); Harris and Sass (2011); Wiswall (2013); and Papay and Kraft (2015).

If we conducted an experiment with unconditional random assignment of RTW teachers to students, then the difference in test scores between students who were taught by a RTW teacher and those who were not would be the average causal effect of RTW teachers. We could estimate a simple linear model like the one below:

$$Y = \beta_0 + \beta_1 RTW + \varepsilon \tag{3}$$

where Y is a student's standardized test score, RTW is a binary variable indicating whether the student was taught by a RTW teacher, and ε is a random error term. Since assignment is random and not conditional on covariates, the error term is uncorrelated with RTW, i.e., $cov(\varepsilon, RTW) = 0$. This means that the OLS estimate, $\widehat{\beta_1}$, is the average casual effect of RTW teachers on test scores. However, in our setting, RTW teachers were not randomly assigned to students, meaning that there is possibly something else driving the estimated relationship between RTW teachers and test scores. In other words, there is possible selection bias. If this is the case, then the $cov(\varepsilon, RTW) \neq 0$ and the OLS estimate of β_1 is no longer the average causal effect of RTW teachers.

There are a couple of things that could be creating selection bias in equation (3). First, the schools that hired RTW teachers did not do so randomly. Teachers chose to apply to work at certain schools and school administrators decided whether to hire them. For instance, teachers might want to return to a school with good working conditions or students who are relatively easy to teach. Whether teachers get hired at their preferred schools depends both on whether there are vacancies and the hiring preferences of the school administrators. Principals might want to hire a high-quality, highly experienced teacher, or they may want to hire someone who previously worked in their school regardless of their quality. It is also possible that their hiring decisions are swayed by influential parents. We do not observe teacher preferences on where they would like to work, nor do we observe the preferences of administrators on who they would like to hire, ergo, they are captured by the error term and confound the estimate of β_1. Second, after RTW teachers were hired, they were likely not assigned to classrooms randomly. Principals might put RTW teachers in classrooms where students are struggling academically or behaviorally, with the thought that their experience could help boost performance. Instead, they could assign RTW teachers to students who are doing well in order to keep performance high. Also, parents, and the teachers themselves, might request a particular classroom assignment. Like the hiring preferences, we do not observe the preferences of administrators, parents, and teachers that are potentially driving the assignment of RTW teachers to classrooms. Thus, they also may confound the estimate of β_1.

We could mitigate these biases in a few different ways. We could exploit variation within students and compare a student who had a RTW teacher in one year to herself in a different year when she did not have a RTW teacher. This would eliminate the bias of student-teacher sorting, and, if we controlled for school characteristics, we would no longer need to be worried about the sorting of teachers into schools. Another possibility would be to compare across students who did and did not have a RTW teacher, controlling for student and school characteristics. However, neither of these methods is feasible in our case because there is not a lot of variation in RTW within students over time and the sample of students who had a RTW teacher is quite small. Therefore, we use a different strategy that has the flavor of a combination of propensity score matching and a difference-in-differences design.

Our first step is to define our comparison groups. We use the assignment of RTW teachers to grades rather than classrooms to remove the bias from student-teacher sorting within grades. This is reasonable to do because RTW teachers were more likely to be in rural, town, or suburban schools anyway.[15] These schools are generally smaller than city schools and have a single class per grade, implying that the treatment is already at the grade level and we are simply making this definition uniform across schools. We then predict the probability that a RTW teacher is assigned to a particular school and grade based on observable, time-varying school and grade characteristics and a set of fixed effects. We do this because the number of school-grades with a RTW teacher is small and this allows us to compare school-grades with different probabilities of having a RTW teacher instead of those that either did or did not have one.

We use the variation in the probability of being assigned a RTW teacher over time to estimate the effect of these teachers on achievement. Specifically, we exploit the fact that the policy was discontinued in 2009. We instrument for the actual assignment of a RTW teacher with the predicted probability of assignment over time and rely on school fixed effects to remove any biases from the sorting of teachers to schools.

We construct our instrument in two steps. First, we use a binary variable that is 1 for school-grade-years that have a RTW teacher and 0 otherwise. This can be 0 or 1 for the years the policy is in place but is always 0 after the policy expires. Using a probit model, we then regress this indicator on school and school-by-grade characteristics that are potentially related to the probability of a school having a RTW teacher in a grade and year. We estimate the probit only during the policy years (2006-07 through 2008-09). Specifically, we estimate the model below:

$$RTW_{gst} = \delta_0 + \delta_1 X_{st} + \delta_2 W_{gst} + \alpha_s + \theta_g + \gamma_t + \mu_{gst} \qquad (4)$$

where g, s, and t represent grades, schools, and academic years, respectively. The outcome variable, *RTW*, is a binary variable equal to 1 if a RTW teacher works in grade g in school s during year t, and 0 otherwise. The vector X includes time-varying school characteristics that are

[15] On average, 20% of schools with a RTW teacher between 2006-07 and 2008-09 were in rural areas; 21% were in towns and suburbs; and 16% were in cities.

potentially related to whether a RTW teacher works in school *s* during year *t*. These characteristics include enrollment and the proportions of students who are economically disadvantaged or LEP. W is a vector of time-varying grade-by-school characteristics potentially related to whether a teacher works in grade g and school *s* during year *t*, including the proportions of students who are female, Black, Hispanic, Asian, or Native American. α, θ, and γ are fixed effects to control for idiosyncrasies across schools, grades, and years, respectively. μ is a grade-by-school-by-year random error term.

We estimate equation (4) only for the years the policy is in place. Second, we predict RTW for *all* years in the sample, both during the policy period and after, and multiply these predicted probabilities by a *Post* variable equal to 1 for years after the policy expired (i.e., the 2009-10 through 2011-12 school years). These values, $\widehat{RTW_{gst}} * Post_t$, are our instrument for RTW_{gst}.

We use this instrument to estimate the following model:

$$Y_{igst} = \beta_0 + \beta_1 RTW_{gst} + \beta_2 Y_{igs,t-1} + \beta_3 X_{igst} + \beta_4 Z_{gst} + \alpha_s + \theta_g + \gamma_t + \varepsilon_{igst} \qquad (5)$$

where i, g, s, and *t* represent students, grades, schools, and academic years, respectively. The dependent variable, *Y*, is the math or reading score from EOG tests, which are standardized within the population by grade and year to have a mean of zero and a standard deviation equal to one. Lagged test scores are included on the righthand side as a proxy for unobserved student ability, effort, and family background. α is a school fixed effect that controls for unobserved differences in teacher work preferences and administrator hiring preferences across schools. The grade fixed effect, θ, controls for unobserved heterogeneity in the grades to which RTW teachers are assigned. γ is a year fixed effect, which captures idiosyncrasies over time related to hiring a RTW teacher, like the Great Recession. *X* is a vector of student characteristics, including race, ethnicity, gender, economic disadvantage, disability status, LEP status, gifted status, and indicators for whether a student is repeating the previous grade or changed schools from the previous year. Z is the predicted probability of a RTW teacher being in school *s* during year *t* and working in grade g (i.e., $\widehat{RTW}_{gst}$ from the probit model). ε is a random student-by-grade-by-school-by-year error term. Standard errors are clustered by school-grade. We are identifying β_1 off plausibly random variation in RTW_{gst} that is generated by the instrument. Specifically, we are using within-school variation in the probability of having a RTW teacher in grade g over time. We compare students in the same school with a high probability of having a RTW in their grade to those with a low probability before and after the end of the policy.

We estimate this model using 2SLS. The first stage is an OLS regression of RTW_{gst} on the instrument, $\widehat{RTW}_{gst} * Post_t$, and the second stage is an OLS regression of Y_{igst} on the predicted values from the first stage. The exclusion restriction is satisfied because the second stage does not include $\widehat{RTW}_{gst} * Post_t$. One concern with using probit fitted values is that identification of the second stage relies on variation induced by the difference in functional forms. To quell this concern, we show that our estimates do not change in a meaningful way when we use a linear probability model to estimate equation (4) instead of a probit. The results are discussed with our other robustness checks in section 6.3.

6 Results

6.1 Probit Estimation

Marginal effects from the estimation of equation (4) are presented in Table 1. The first column shows results from the estimation of our core probit model that includes the racial and ethnic composition of school-grades, the proportions of economically disadvantaged and LEP students by school, and school enrollment, as well as school, grade, and year fixed effects. The proportions of Black, Native American, and economically disadvantaged students, as well as school enrollment, are positively related to the probability of a RTW teacher working in a school-grade. The proportions of female, Hispanic, Asian, and LEP students are negatively correlated with having a RTW teacher. We add school-level proportions of students categorized as having different disabilities to the core model in column (2). The proportions of students categorized as having an emotional, learning, or speech-language disability are positively related to the probability of a RTW teacher working in a school and grade, while the proportions of students categorized as having a physical or mental disability are negatively correlated with having a RTW teacher. In column (3), we add the average experience of non-RTW teachers in the previous year and the proportion of teachers eligible for retirement in the previous year, both of which are at the school-level and are positively correlated with the probability of a RTW teacher working in a school-grade. In the remainder of this section, we only discuss results that use the predicted values from the core probit specification. We check the robustness of our results to the other specifications later on.

6.2 Main Results

Equation (5) is identified by variation in the predicted values of the first stage regression within school-grades over time. An analysis of variance shows that much of the variation in the first stage is across schools, but that there is still a sizeable amount of variation from the mean within school-grades. For both reading and math, the partial sum of squares for school-grade interactions is about 600, which is statistically greater than zero at the 1% level (Table 2).

OLS and IV estimation results for reading are presented in columns (1)-(3) and (4)-(6) of Table 3, respectively. Columns (1) and (4) include school, grade, and year fixed effects. We add the previous standardized reading test score as a regressor in columns (2) and (5) and include student characteristics in columns (3) and (6). The OLS estimates in all three specifications are very small, and, except for the estimates in column (3), we cannot reject the null hypothesis that they are equal to zero. For example, in column (3) the coefficient on *RTW* is 0.0064 and has a standard error of 0.0034. Looking at the OLS estimates, it appears that RTW teachers did not have a significant impact on reading achievement. However, these estimates are likely biased because of the endogeneity of test scores and the assignment of RTW teachers. We address this endogeneity by instrumenting *RTW* with the predicted probability of a RTW teacher working in school *s* and grade *g* in year *t* multiplied by *Post*. In column (4), the estimated coefficient on *RTW* using this estimation strategy is 0.0267 and has a standard error of 0.0131. When we add the student's previous standardized reading test score on the right-hand side to control for unobserved ability, effort, and family background, *RTW* becomes significant. The coefficient is 0.0162 and has a standard error of 0.0067, and we can reject the null hypothesis that it is equal to zero at the 5% level of significance. The coefficient becomes slightly larger in column (6) with the addition of student characteristics (0.0198) and is statistically different from zero at the 1% significance level. We perform a Hausman test and can reject the null hypothesis that the OLS and IV estimates of *RTW* are equal at the 2% significance level. These results indicate that reading test scores increased by 1.98 percent of a standard deviation for students in the same school who had a RTW teacher in their grade during the policy compared to students who did not, conditional on covariates. Though this is a small effect, it is economically meaningful because reading achievement is traditionally thought to be influenced by learning done at home more than changes in school policies (Cronin, et al. (2005), Figlio and Ladd (2008)).

Table 4 shows OLS and IV estimation results for math and is organized like Table 3. Similar to the OLS estimates for reading, the ones for math are very small and not statistically different from zero. In column (3), the coefficient on *RTW* is 0.0075 and has a standard error of 0.0055. The IV estimates are slightly larger than those for reading. The coefficient on *RTW* is 0.0364 in column (6), with a standard error of 0.0108, and we can reject the null hypothesis that it is equal to zero at the 1% level of significance. This result suggests that RTW teachers had a positive impact on math achievement. We performed the Hausman test and can reject the null hypothesis that the OLS and IV estimates of RTW are equal at the 0.19% significance level.

6.3 Robustness

We perform six robustness checks. First, we apply inverse probability weights to account for students who are in the sample twice in a particular year, i.e., students who take two math or reading classes. Results for reading and math are presented in column (2) of Tables 5 and 6, respectively. The estimated coefficient on *RTW* for reading is 0.018, slightly smaller than the unweighted estimate but still statistically significant at the 5% level. For math, the estimate of *RTW* is 0.0357 relative to 0.0364 in the unweighted model and remains statistically different from zero at the 1% level.

Second, it is possible that the predicted probability of a RTW teacher working in a school-grade-year is non-linearly related to test scores. To control for any non-linearities, we include the quartic of the predicted values from equation (4) as a covariate in equation (5). The quartic allows for more flexibility than a linear, quadratic, or cubic term. Neither the reading nor the math estimates of the coefficient on *RTW* change, as shown in column (3) of Tables 5 and 6.

Third, we omit the last year the policy was in place and the first year after its expiration. Teachers and administrators might have anticipated the policy's end or thought that it would be renewed, which had happened several times before, and it is possible that they made different decisions about retirement and hiring as a result. To determine whether our estimates are being driven by these anticipatory effects, we re-estimate the probit model with the 2006-07 and 2007-08 school years only, and then predict the probability of having a RTW teacher for these years and for the 2010-11 and 2011-12 school years.[16] We repeat the IV estimation using these predicted values. Results are presented in column (4) of Tables 5 and 6. The coefficient on RTW for reading is 0.0203, which is slightly larger than the estimate with all sample years, and it is statistically different from zero at the 5% significance level. For math, the coefficient on RTW is 0.0332, slightly smaller than β_1 from the model with all years and statistically significant at the 5% level. This suggests that both the change in reading and math scores due to RTW are robust to anticipatory effects.

Fourth, we re-estimate equation (4) using a linear probability model instead of a probit model. Column (5) of Tables 5 and 6 show that the reading and math IV estimates decrease only slightly, indicating that the results are not being identified solely by the non-linearity of the predicted values from the probit specification.

Fifth, we do a placebo test and estimate reduced form models where the probability of a RTW teacher being in a grade and school is randomly assigned. Column (6) of Tables (5) and (6) show the estimation results. For reading, the effect of the random instrument is -0.003 (0.0022) and for math it is -0.0017 (0.0035). Since these coefficients are close to zero, we are confident that the estimated effects of RTW teachers are not just picking up randomness in test scores.

Finally, we test the robustness of our results to changes in the specification of equation (4). Table 7 shows the results for reading and Table 8 shows those for math. Column (1) in both tables shows the IV estimates using $\widehat{RTW}_{gst}$ from the probit with the core variables, as seen in column (6) of Tables 3 and 4. Column (2) uses the predicted probabilities from a model with the core

[16] Marginal effects from the probit with the last year of the policy omitted are presented in Table A2 of the appendix.

variables and different school-level proportions of students categorized as having different disabilities. In column (3), $\widehat{RTW}_{gst}$ is determined from a probit with the core and disability variables, as well as aggregate characteristics of the non-RTW teaching workforce, i.e., the average experience of non-RTW teachers in the previous year and the proportion of teachers eligible for retirement in the previous year. The estimated coefficient on RTW decreases slightly from column (1) to (2) and decreases even less between columns (1) and (3) for both math and reading. To sum, our results are generally robust to these six changes to our model specification.

7 Extensions

7.1 Heterogeneity by Student Ability

We are interested in whether there are differential effects of RTW teachers across students of different abilities. To examine this, we group students into quartiles based on their previous test score in reading or math, where the quartiles are defined within grade-years. Then, we estimate a version of equation (5) that includes dummy variables for each quartile instead of the previous test score, RTW, and interactions of the quartile indicators with RTW. In this case, RTW is instrumented with $\widehat{RTW}_{gst} * Post_t$ and the interaction terms are instrumented with $\widehat{RTW}_{gst} * Post_t * QuartileDummy$. Results are presented in Table 9.

In math, students in the bottom three quartiles of the previous test score who had a RTW teacher in their grade and school performed worse in the current year relative to students in the top quartile with a RTW teacher. The estimated coefficients on the interaction terms range from -0.12 to -0.18 and are statistically different from zero (column (2)). This suggests that RTW teachers were particularly important for top math students. In reading, however, the results are more ambiguous. The effect on students at the bottom of the distribution is positive, but statistically insignificant, while the effect on the second quartile is large and negative (-0.18) and the third quartile is close to zero (column (1)).

7.2 Heterogeneity by Grade

We also investigate whether RTW teachers have different effects for students in different grades by estimating equation (5) for grades 4-8 individually. Table 10 shows the results by subject. The effect of RTW teachers on each of 4^{th}-, 5^{th}-, and 6^{th}-grade reading scores is around 0.035 and is statistically different from zero at the 5% level, as shown in columns (1)-(3). RTW teachers do not have an effect on 7^{th} or 8^{th} grade reading scores, as shown in columns (4) and (5). For math, the results show that RTW teachers are most effective in grades 5, 6, and 8. The coefficient on *RTW* for 5^{th} graders is 0.06 and is statistically significant at the 5% level (column (7)). The effects for 6^{th} and 8^{th} graders are 0.059 and 0.041, respectively. Both of these are statistically different from zero. There is no effect on math scores for students in grades 4 and 7. Overall, the reading results show that RTW teachers are particularly effective in lower grade reading classes, whereas the math results show less of a pattern across grades.

7.3 Suspensions and Detentions

The positive effects of RTW teachers on student achievement could be explained by their ability to teach material better than other teachers, by their ability to manage student behavior in the classroom, some combination of these, or other reasons. We examine student disciplinary records to see whether student behavior changed after the RTW policy was discontinued. Specifically, we look at the effects of RTW teachers on out-of-school and in-school suspensions, and detentions. Schools are only required to report legally reportable offenses, which typically result in an out-of-school suspension, but many also report smaller incidences, which typically result in an in-school suspension or detention. The data include student-incident level information on out-of-school suspensions for the 2006-07 through the 2011-12 school years, and in-school suspensions and detentions for the 2007-08 through the 2011-12 school years. We create three binary variables equal to one if the student was ever suspended out of school, in school, or was given detention during the school year. If a student does not appear in this data, we assume that they were not involved in any disciplinary incidents during the year.

We estimate equation (5) using these binary discipline variables as the dependent variables and excluding the student's previous standardized test score. The results are presented in Table 11. Column (1) shows that the effect of RTW teachers on the probability of a student receiving an out-of-school suspension is -0.0092 (0.0041), which is statistically different from zero at the 5% level. In column (2), the effect on the probability of getting an in-school suspension is -0.0381 (0.0097), which is statistically significant at the 1% level. The likelihood of getting a detention decreases by 0.0205 (0.0058) and is statistically significant at the 1% level, as shown in column (3). These results suggest that RTW teachers have a positive impact on student behavior, which could help explain the positive effects on test scores.

8 Conclusion

The previous literature shows that teachers respond to the "push" and "pull" incentives embedded in DB pension plans. However, it is less clear how pension incentives impact the quality of the teaching workforce. Most of the research on pensions and teacher quality focuses on the teachers who respond to "pull" incentives and generally shows that less effective teachers are being pulled to stay longer than they otherwise would (Koedel, Podgursky, and Shi (2013), Fitzpatrick and Lovenheim (2014), Williams (2015)). One paper, Koedel, Podgursky, and Shi (2013), shows that teachers likely incentivized by the pension push are no different in terms of quality than those who were likely pulled to stay or had a regular retirement. We contribute to the literature in assessing the role of "push" factors by studying a RTW policy in North Carolina that effectively eliminated the pension push. RTW allowed retired teachers to return full-time and receive their pension benefits *and* their full-time salary concurrently after only a short break from work. The teachers who chose to return were likely those who would have kept teaching in absence of push incentives. Unlike, Koedel, Podgursky, and Shi (2013), we find that higher quality teachers seem to be affected by the pension push, though our results are relative to all other teachers rather than those likely impacted by the pension pull or those likely to have a regular retirement. We find that RTW teachers increased reading and math achievement by about 2 and 3.6 percent of a standard deviation, respectively, which suggests that DB pension plans incentivize higher-quality teachers to exit sooner than they otherwise would. Our results also suggest that RTW policies can bring effective teachers back into the teaching workforce and modestly improve average workforce quality.

These conclusions are tempered by the following caveats. First, they only apply to North Carolina. It is possible that a similar RTW policy implemented elsewhere could attract different kinds of teachers and lead to different impacts on student achievement. Second, our findings may also be interpreted as high-quality teachers strategically retiring earlier during the policy in order to return. Indeed, Mahler (2013) shows that that probability of retiring right at eligibility is higher during the policy than before. We expect that the people who understand their pensions and the policy's advantages are going to be the ones who decide to participate. In other words, we treat this as a feature of the policy.

Third, we do not conduct a cost-benefit analysis, so we do not know whether these increases in student achievement are big enough to offset the cost of hiring these teachers.

However, we do think about who would have been hired in absence of the policy, which we do not observe, to give some sense of how expensive these teachers are compared to alternative hires. If a novice teacher was hired instead of a RTW teacher, for instance, the school district would pay out a lower salary plus health benefits, as well as the RTW teacher's annuity and health benefits. If the district hired the RTW teacher, it would have to pay a higher salary plus her annuity and health benefits. Thus, the cost of hiring a RTW teacher rather than a novice teacher is the difference between the two teachers' salaries minus the amount that would have been paid for the novice's health benefits, meaning that the RTW teacher is relatively expensive for a given year. To some degree, this high cost is likely mitigated both by the fact that more experienced teachers are generally more effective than novices and by the gains in student achievement that we estimate in this paper. Additionally, the cost differential may decline over time as the novice teacher gains experience and is paid a higher salary. An alternative scenario might be that the school district hires another highly experienced, non-retired teacher instead of the RTW teacher. In this case, their salaries are likely similar and how much they each cost depends more on whether they are good teachers.

Finally, we do not directly address the RTW policy adopted by North Carolina from July 2019 through June 2021, though the conclusions do suggest that the newer policy might have had a modest positive impact on workforce quality. Like the one we analyze in this paper, the more recent policy allowed retired teachers to return to work full-time and collect both retirement benefits and earn a full-time salary. However, it required them to go back to high-need schools and limited their compensation. Instead of receiving the salary they retired with, they were paid on the first step of the salary schedule, unless they were certified in STEM subjects or special education, in which case they were paid on the sixth step. The policy expired in June 2021 but will possibly be extended until 2024 with several revisions. Studying this RTW policy seems like a promising avenue for future research.

References

Arias, E. (2011). United States Life Tables, 2007. *National Vital Statistics Reports, 59*(9).

Asch, B., Haider, S. & Zissimopoulos, J. (2005). Financial incentives and retirement: evidence from federal civil service workers. *Journal of Public Economics, 89,* 427-440.

Coile, C. & Gruber, J. (2000). Social Security and Retirement. NBER Working Paper No. 7830.

__________________. “Social Security Incentives for Retirement.” In *Themes in the Economics of Aging*, edited by David Wise, 311-354. University of Chicago Press, 2001.

Costrell, R. & McGee, J. (2010). Teacher pension incentives, retirement behavior, and potential for reform in Arkansas. *Education Finance and Policy*, *5*(4), 492-518.

Costrell, R. & Podgursky, M. (2009). Peaks, cliffs, and valleys: The peculiar incentives in teacher retirement systems and their consequences for school staffing. *Education Finance and Policy, 4*(2), 175–211.

Cronin, J., Kingsbury, G. G., McCall, M. S., & Bowe, B. (2005). The Impact of the No Child Left Behind Act on Student Achievement and Growth: 2005 Edition (Northwest Evaluation Association Technical Report). Lake Oswego, OR: Northwest Evaluation Association.

Figlio, D. N., & Ladd, H. (2008). School Accountability and Student Achievement. In H. Ladd & E. Fiske (Eds.), Handbook of Research in Education Finance and Policy (pp. 166-182). New York: Routledge.

Fitzpatrick, M. (2019). Pension Reform and Return to Work Policies. *Journal of Pension Economics and Finance*, *18*(4), 500-514.

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

Folwell, D. (2019). Teachers' and State Employees' Retirement Systems: Your Retirement Benefits. Raleigh: Department of State Treasurer.

Friedberg, L. & Turner, S. (2010). Labor market effects of pensions and implications for teachers. *Education Finance and Policy*, *5*(4), 463-491.

Harris, D. & Sass, T. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*, 798-812.

Ippolito, R. (1997). *Pension Plans and Employee Performance: Evidence, Analysis, and Policy*. University of Chicago Press.

Kim, D., Koedel, C., Kong, W., Ni, S., Podgursky, M. & Wu, W. (2021). Pensions and Late-Career Teacher Retention. *Education Finance and Policy*, *16*(1), 42-65.

Koedel, C. & Podgursky, M. (2016). Teacher Pensions. *Handbook of the Economics of Education, 5*, 281-303.

Koedel, C., Podgursky, M. & 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.

Mahler, P. (2013). Lifting the Salary Cap: The Effects of a Return-to-Work Policy on Teacher Retirement, Retention, and Quality. Working paper.

Ni, S. & 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. & Wang, X. (2021). Teacher Pension Enhancements and Staffing in an Urban School District. *Journal of Pension Economics and Finance,* 1-21.

Ni, S., Podgursky, M. & Wang, X. (forthcoming). Teacher Pension Plan Incentives, Retirement Decisions, and Workforce Quality. *Journal of Human Resources*.

North Carolina General Assembly Legislation S.L. 1998-212, S.L. 1998-217, S.L. 2000-67, S.L. 2001-424, S.L. 2002-126, SL. 2004-124, S.L. 2005-144, S.L. 2005-276, S.L. 2005-345, S.L. 2007-145, S.L. 2007-326.

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

Rivkin, S., Hanushek, E. & Kain, J. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, *73*(2), 417-458.

Rockoff, J. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*, *94*(2), 247-252.

Samwick, A. (1998). New evidence on pensions, social security, and the timing of retirement. *Journal of Public Economics, 70*, 207-236.

Stock, J. & Watson, M. *Introduction to Econometrics*. Addison Wesley: Boston, 2003.

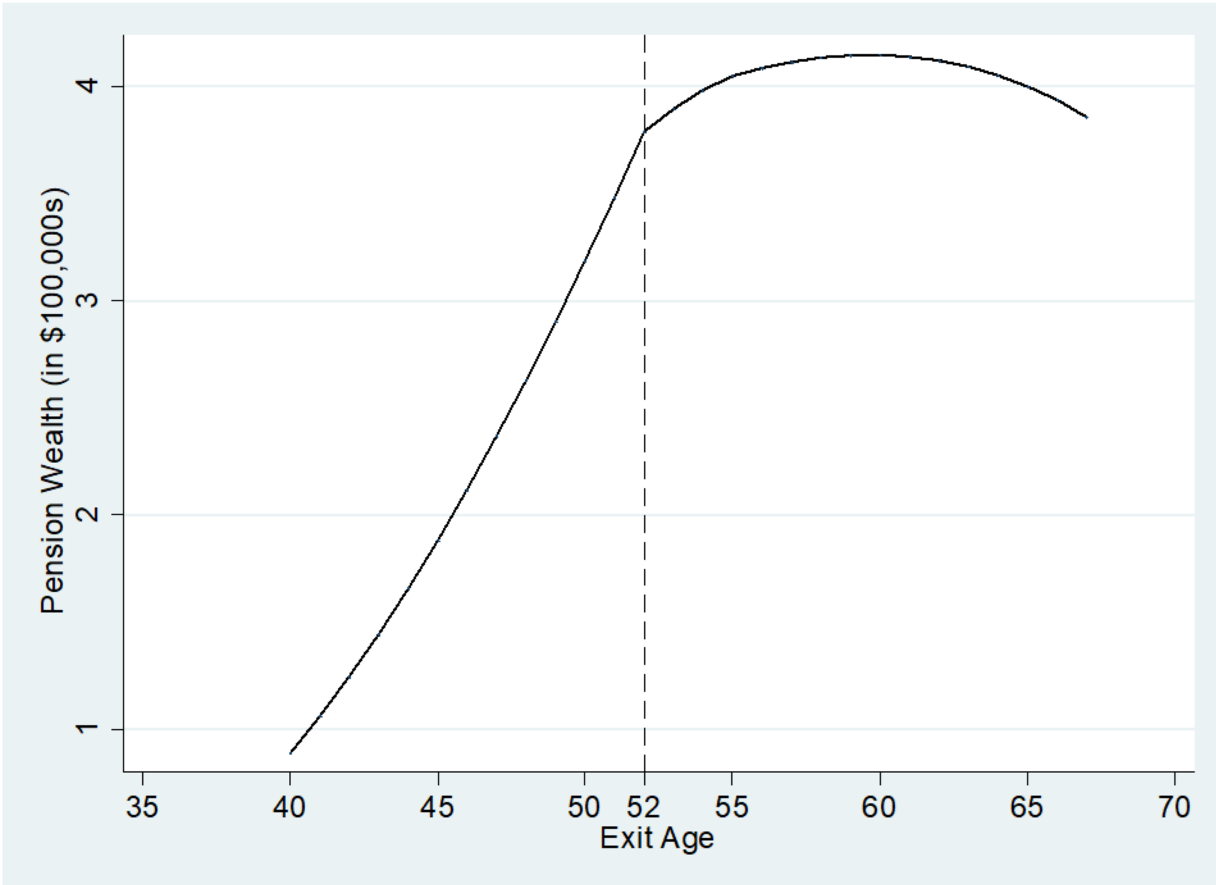
Stock, J. & Wise, D. (1990). Pensions, the Option Value of Work, and Retirement. *Econometrica*, *58*(5), 1151-1180.

Williams, K. (2015). Teacher Responses to Early Retirement Incentives and the Effect on Student Performance. Unpublished.

Wiswall, M. (2013). The Dynamics of Teacher Quality. *Journal of Public Economics, 100*, 61-78.

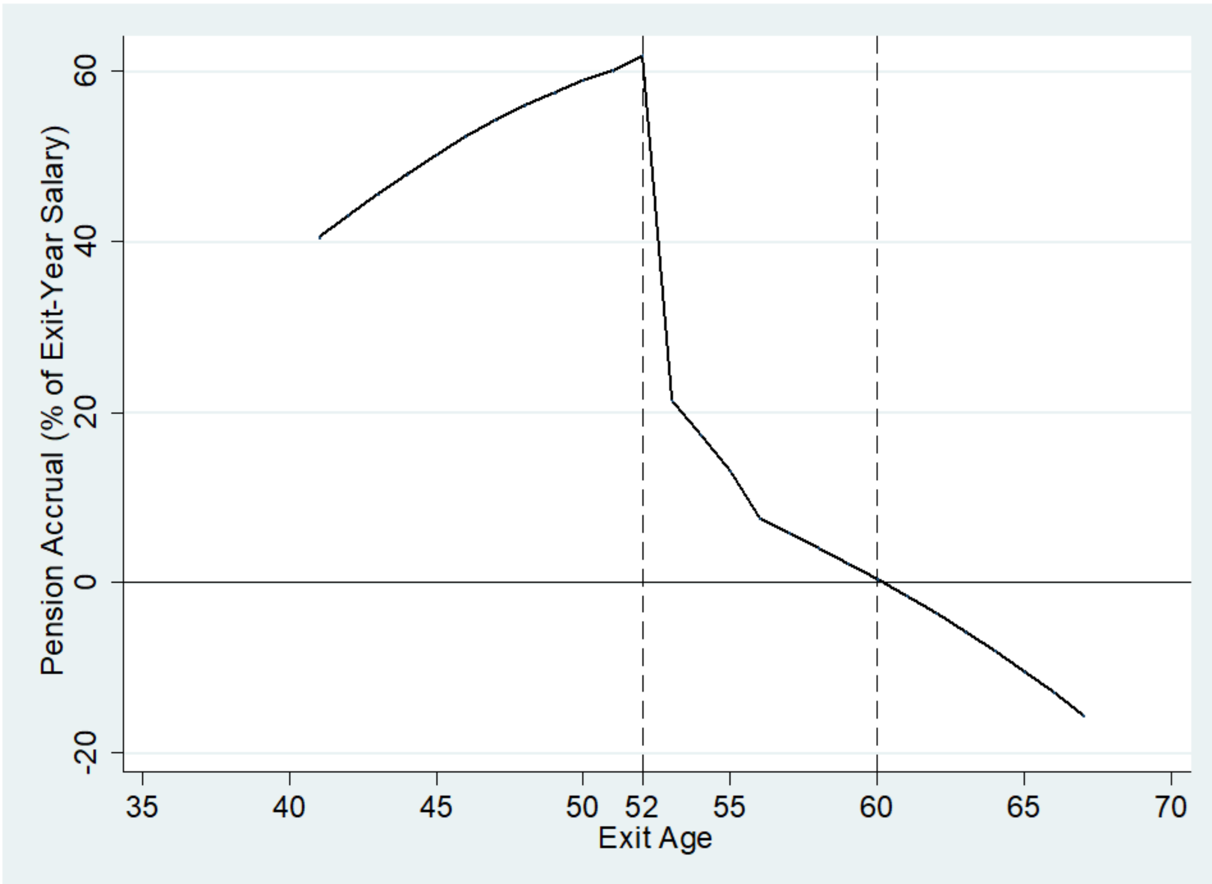
Figures

Figure 1: Pension Wealth (in $100,000s) by Exit Age for Teachers in North Carolina



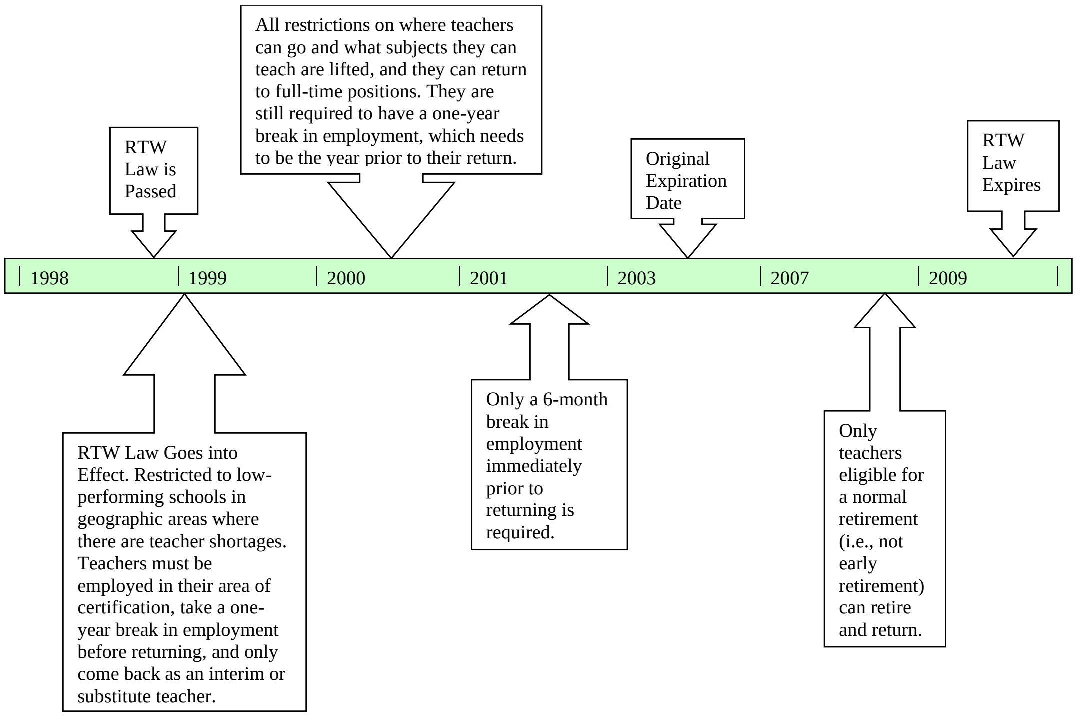
Notes: This graph shows the pension wealth for a non-Hispanic white female with a bachelor's degree who began teaching at age 22 and earned the median salary of teachers with her level of experience in her last year of teaching (i.e., the 2006-07 school year). The y-axis is pension wealth in $100,000s, and the x-axis is the age at which the teacher exits the teaching workforce. The teacher is eligible for retirement after 30 years of experience, as shown by the vertical dashed line at age 52.

Figure 2: Pension Accrual (as % of Exit Year Salary) by Exit Age for Teachers in North Carolina



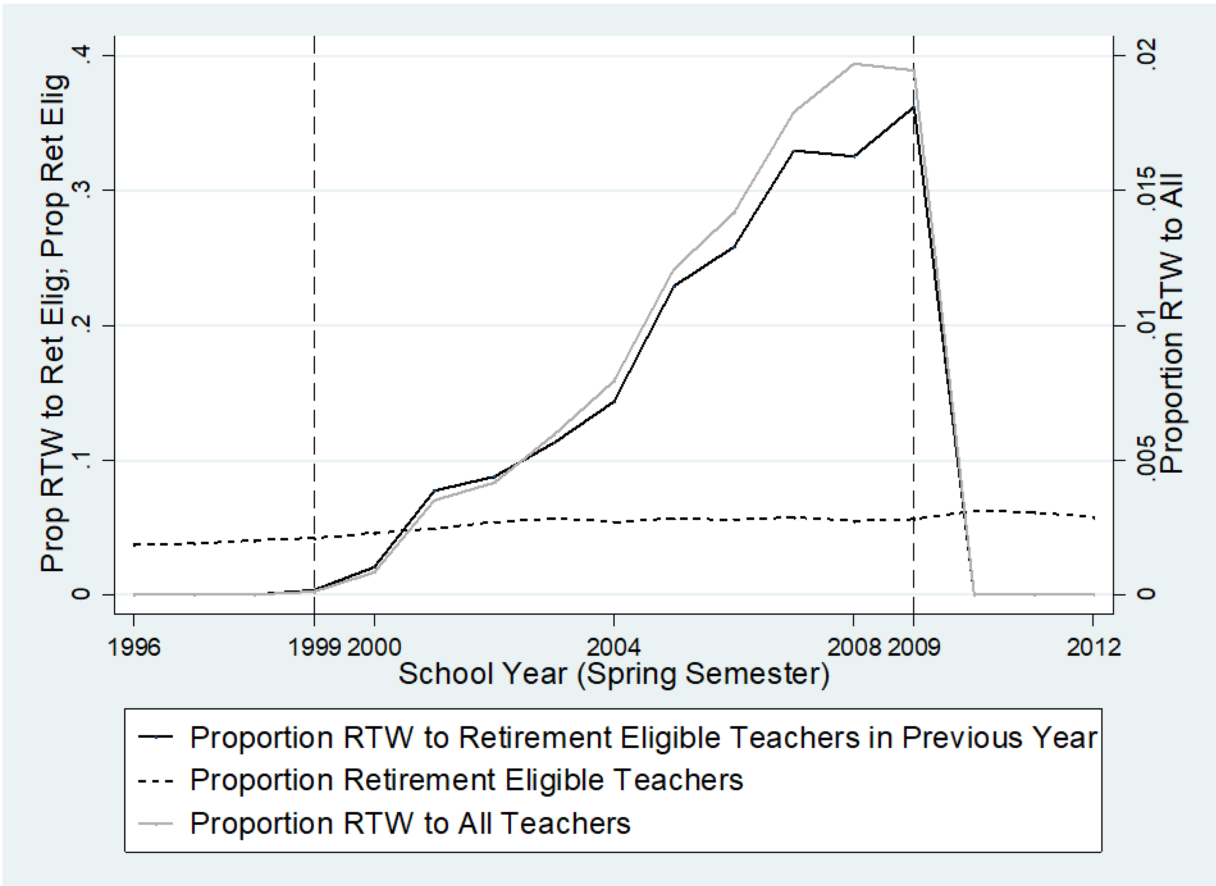
Notes: This graph shows the accrual profile for a non-Hispanic white female with a bachelor's degree who began teaching at age 22 and earned the median salary of teachers with her level of experience in her last year of teaching (i.e., the 2006-07 school year). The y-axis is the teacher's pension accrual as a percentage of her exit-year salary, and the x-axis is the age at which the teacher exits the teaching workforce. The vertical dashed lines at ages 52 and 60 show where the teacher is eligible for full retirement benefits and where her accrual becomes negative, respectively.

Figure 3: Abbreviated RTW Policy Timeline



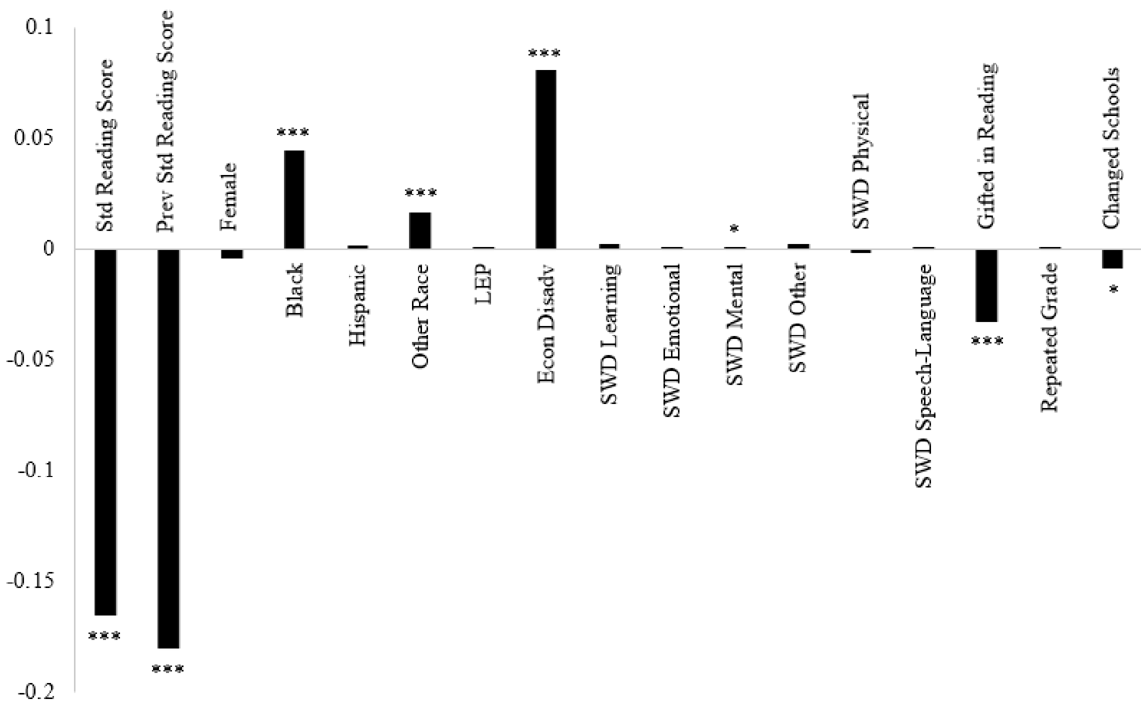
Notes: Information from North Carolina General Assembly Legislation S.L. 1998-212, S.L. 1998-217, S.L. 2000 67, S.L. 2001-424, S.L. 2002-126, SL. 2004-124, S.L. 2005-144, S.L. 2005-276, S.L. 2005-345, S.L. 2007-145, S.L. 2007-326.

Figure 4: Take-Up of the RTW Policy Over Time as a Fraction of All Teachers and Retirement Eligible Teachers



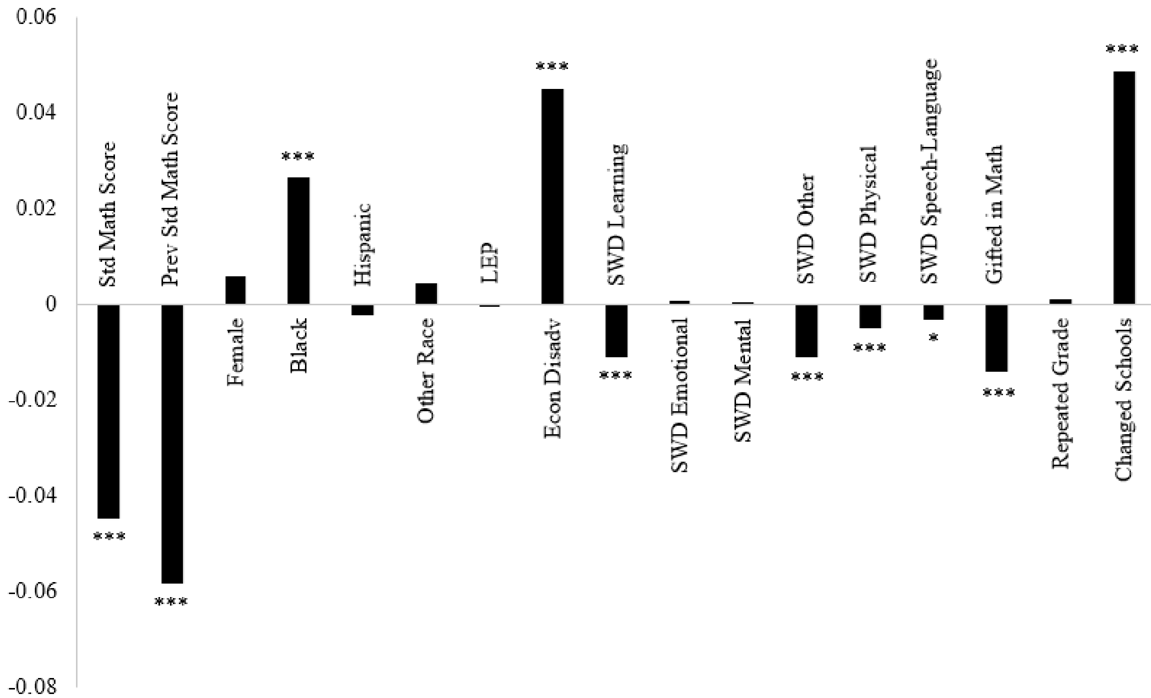
Notes: This graph shows how many teachers returned to work during the policy. The left y-axis shows the proportion of retirement eligible teachers as well as the proportion of RTW to retirement eligible teachers. The right y-axis shows the proportion of RTW to all teachers. The x-axis is the spring semester of each school year from 1996-2012, and the vertical dashed lines indicate the first and last years of the policy.

Figure 5: Differences in the Average Characteristics of Students in Reading Classes Who Had and Did Not Have RTW Teachers



Notes: These are student-level differences in means for students in grades 4-8 during the 2008-09 school year for students who took reading classes. The bars show the average for students who had a RTW teacher minus the average for those who did not have a RTW teacher. LEP is limited English proficient; Econ Disadv is economically disadvantaged; SWD is a student with a disability; Gifted in Reading means the student is academically or intellectually gifted in reading; Repeated Grade means the student repeated the previous grade; and Changed Schools means the student switched schools from the prior year. The stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Figure 6: Differences in the Average Characteristics of Students in Math Classes Who Had and Did Not Have RTW Teachers



Notes: These are student-level differences in means for students in grades 4-8 during the 2008-09 school year for students who took math classes. The bars show the average for students who had a RTW teacher minus the average for those who did not have a RTW teacher. LEP is limited English proficient; Econ Disadv is economically disadvantaged; SWD is a student with a disability; Gifted in Math means the student is academically or intellectually gifted in math; Repeated Grade means the student repeated the previous grade; and Changed Schools means the student switched schools from the prior year. The stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Tables

Table 1: Marginal Effects from Probit Regressions of RTW on School & Grade Characteristics

	Core (1)	+ SWD (2)	+ Teacher Characteristics (3)
Proportion Female Students in School-Grade	-0.080	-0.085	-0.083
	(0.125)	(0.126)	(0.126)
Proportion Black Students in School-Grade	0.162	0.174	0.176
	(0.112)	(0.113)	(0.114)
Proportion Hispanic Students in School-Grade	-0.075	-0.074	-0.065
	(0.194)	(0.195)	(0.194)
Proportion Asian Students in School-Grade	-0.432	-0.452	-0.451
	(0.473)	(0.467)	(0.469)
Proportion Native American Students in School-Grade	0.600	0.591	0.573
	(0.417)	(0.409)	(0.415)
Proportion Economically Disadvantaged Students in School	0.096	0.068	0.059
	(0.119)	(0.122)	(0.123)
Proportion Limited English Proficient Students in School	-0.998*	-0.873	-0.854
	(0.552)	(0.553)	(0.557)
Enrollment	0.0001	0.0002	0.0002
	(0.00015)	(0.00015)	(0.00015)
Proportion Students in School with Emotional Disability		1.784***	1.757***
		(0.576)	(0.585)
Proportion Students in School with Learning Disability		0.362	0.435
		(0.615)	(0.610)
Proportion Students in School with Mental Disability		-1.261	-1.323
		(0.938)	(0.946)
Proportion Students in School with Physical Disability		-0.301	-0.262
		(0.720)	(0.731)
Proportion Students in School with Speech/Language Disability		0.070	0.086
		(0.831)	(0.840)
Average Experience of Non-RTW Teachers in the School in the Previous Year			0.006
			(0.009)
Proportion of Retirement Eligible Teachers in the School in the Previous Year			0.287
			(0.320)
Observations	4,082	4,082	4,066

Notes: These are marginal effects of the probit regression given by equation (4). The sample used for estimation is at the school-by-grade-by-year level for the 2006-07 through the 2008-09 school years. Column (1) includes a set of core variables. Column (2) adds student with disability (SWD) variables. Column (3) adds aggregate characteristics of non-RTW teachers. Each column includes school, grade, and year fixed effects, and standard errors are clustered at the school-by-grade level and shown in parentheses. Stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Table 2: Decomposition of the Variation in the First Stage into Grade, School, and School-Grade Components

	Partial SS	DF	MS	F-Statistic	P-Value
Panel A: Reading ANOVA					
Model	36,071.84	4,710	7.66	515.18	0.00
Grade	3.87 E-15	4	9.67 E-16	0.00	1.00
School	14,128.45	1,913	7.39	496.81	0.00
School-Grade	617.38	2,793	0.22	14.87	0.00
Residual	37,981.40	2,554,953	0.01		
Total	74,053.24	2,559,663	0.03		
Panel B: Math ANOVA					
Model	35,215.97	4,710	7.48	500.86	0.00
Grade	1.58 E-14	4	3.94 E-15	0.00	1.00
School	13,840.70	1,913	7.24	484.67	0.00
School-Grade	571.09	2,793	0.20	13.70	0.00
Residual	36,617.62	2,452,960	0.01		
Total	71,833.59	2,457,670	0.03		

Notes: Analysis of variance for the first stage models in both reading (*Panel A*) and math (*Panel B*). The variation in the models is decomposed into grade, school, and school-by-grade components. No other variables are included.

Table 3: Ordinary Least Squares and Instrumental Variable Estimates of the Effects of RTW Teachers on Standardized Reading Test Scores

	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
RTW	0.0023	0.0050	0.0064*	0.0267	0.0162**	0.0198***
	(0.0055)	(0.0035)	(0.0034)	(0.0131)	(0.0067)	(0.0068)
Previous Std Test Score		X	X		X	X
Student Characteristics			X			X
Predicted RTW				X	X	X
Observations	2,514,985	2,514,985	2,506,129	2,514,985	2,514,985	2,506,129
R-squared	0.1178	0.6797	0.6957	0.1178	0.6797	0.6957
Hausman p-value: 0.0198						

Notes: OLS and IV estimation results. The sample used for estimation is at the student-by-class-by-year level for the 2006-07 through the 2011-12 school years. The dependent variable is a student's standardized reading test score. "RTW" is a binary indicator equal to 1 for school-grade-years where RTW teachers worked and 0 otherwise. "RTW" is instrumented with $\widehat{RTW}_{gst} * Post_t$ in columns 4-6. All specifications include school, grade, and year fixed effects. Student characteristics include female, Black, Hispanic, other race, LEP status, economic disadvantage, student with disability (SWD) indicators, gifted in reading, and indicators for whether a student repeated the previous grade or changed schools from the prior year. Omitted categories are White and Other Disability. "Predicted RTW" is the predicted probability of a school-grade-year having a RTW teacher from the probit regression in equation (4). The p-value of the Hausman specification test between columns (3) and (6) is reported in the last row. Standard errors are clustered at the school-by-grade level and are shown in parentheses. Stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Table 4: Ordinary Least Squares and Instrumental Variable Estimates of the Effects of RTW Teachers on Standardized Math Test Scores

	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
RTW	-0.0022	0.0068	0.0075	0.0331**	0.0341***	0.0364***
	(0.007)	(0.0056)	(0.0055)	(0.0151)	(0.0110)	(0.0108)
Previous Std Test Score		X	X		X	X
Student Characteristics			X			X
Predicted RTW				X	X	X
Observations	2,413,045	2,413,045	2,404,535	2,413,045	2,413,045	2,404,535
R-squared	0.138	0.7067	0.7208	0.138	0.7066	0.7208
Hausman p-value: 0.0019						

Notes: OLS and IV estimation results. The sample used for estimation is at the student-by-class-by-year level for the 2006-07 through the 2011-12 school years. The dependent variable is a student's standardized math test score. "RTW" is a binary indicator equal to 1 for school-grade-years where RTW teachers worked and 0 otherwise. "RTW" is instrumented with $\widehat{RTW}_{gst} * Post_t$ in columns 4-6. All specifications include school, grade, and year fixed effects. Student characteristics include female, Black, Hispanic, other race, LEP status, economic disadvantage, student with disability (SWD) indicators, gifted in math, and indicators for whether a student repeated the previous grade or changed schools from the prior year. Omitted categories are White and Other Disability. "Predicted RTW" is the predicted probability of a school-grade-year having a RTW teacher from the probit regression in equation (4). The p-value of the Hausman specification test between columns (3) and (6) is reported in the last row. Standard errors are clustered at the school-by-grade level and are shown in parentheses. Stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Table 5: Instrumental Variable Estimates of the Effects of RTW Teachers on Standardized Reading Test Scores

	IV					Reduced Form
	Base	Weighted	Quartic in Predicted RTW	Transition Years Omitted	Linear Probability Model	Randomized Instrument
	(1)	(2)	(3)	(4)	(5)	(6)
RTW	0.0198***	0.0180***	0.0196***	0.0203**	0.0167**	
	(0.0068)	(0.0067)	(0.0068)	(0.0080)	(0.0065)	
Random Instrument						-0.0030
						(0.0022)
Observations	2,506,129	2,506,129	2,506,129	1,648,413	2,506,129	2,550,790
R-squared	0.6957	0.6923	0.6957	0.6912	0.6957	0.6957

Notes: IV estimation results (for all but column (6), which shows the estimated effect of the instrument directly on test scores). The sample used for estimation is at the student-by-class-by-year level for the 2006-07 through the 2011-12 school years. The dependent variable is a student's standardized reading test score. "RTW" is a binary indicator equal to 1 for school-grade-years where RTW teachers worked and 0 otherwise. "RTW" is instrumented with $\widehat{RTW}_{gst} * Post_t$ in all specifications. All specifications include previous standardized test scores, student characteristics, predicted RTW, and school, grade, and year fixed effects. Student characteristics include female, Black, Hispanic, other race, LEP status, economic disadvantage, student with disability (SWD) indicators, gifted in reading, and indicators for whether a student repeated the previous grade or changed schools from the prior year. Omitted categories are White and Other Disability. "Predicted RTW" is the predicted probability of a school-grade-year having a RTW teacher from the probit regression in equation (4). Column (1) repeats the IV estimation results from column (6) of Table 2, the base model. Inverse probability weights are applied in column (2) to account for students who take two reading classes in a year. Column (3) includes a quartic in the predicted values from equation (4). In column (4), the transition years are omitted, meaning the 2008-09 and 2009-10 school years are not included in the IV and are not used to predict $\widehat{RTW}_{gst}$. Column (5) shows IV estimates when a linear probability model is used to calculate "Predicted RTW" rather than a probit model. Column (6) shows reduced form estimates after the probability of having a RTW teacher in a grade and school is randomly assigned. Standard errors are clustered at the school-by-grade level and are shown in parentheses. Stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Table 6: Instrumental Variable Estimates of the Effects of RTW Teachers on Standardized Math Test Scores

	IV					Reduced Form
	Base	Weighted	Quartic in Predicted RTW	Transition Years Omitted	Linear Probability Model	Randomized Instrument
	(1)	(2)	(3)	(4)	(5)	(6)
RTW	0.0364***	0.0357***	0.0362***	0.0332**	0.0321***	
	(0.0108)	(0.0110)	(0.0108)	(0.0130)	(0.0104)	
Random Instrument						-0.0017
						(0.0035)
Observations	2,404,535	2,404,535	2,404,535	1,578,603	2,404,535	2,449,144
R-squared	0.7208	0.7194	0.7208	0.7224	0.7208	0.7210

Notes: IV estimation results (for all but column (6), which shows the estimated effect of the instrument directly on test scores). The sample used for estimation is at the student-by-class-by-year level for the 2006-07 through the 2011-12 school years. The dependent variable is a student's standardized math test score. "RTW" is a binary indicator equal to 1 for school-grade-years where RTW teachers worked and 0 otherwise. "RTW" is instrumented with $\widehat{RTW}_{gst} * Post_t$ in all specifications. All specifications include previous standardized test scores, student characteristics, predicted RTW, and school, grade, and year fixed effects. Student characteristics include female, Black, Hispanic, other race, LEP status, economic disadvantage, student with disability (SWD) indicators, gifted in math, and indicators for whether a student repeated the previous grade or changed schools from the prior year. Omitted categories are White and Other Disability. "Predicted RTW" is the predicted probability of a school-grade-year having a RTW teacher from the probit regression in equation (4). Column (1) repeats the IV estimation results from column (6) of Table 2, the base model. Inverse probability weights are applied in column (2) to account for students who take two math classes in a year. Column (3) includes a quartic in the predicted values from equation (4). In column (4), the transition years are omitted, meaning the 2008-09 and 2009-10 school years are not included in the IV and are not used to predict $\widehat{RTW}_{gst}$. Column (5) shows IV estimates when a linear probability model is used to calculate "Predicted RTW" rather than a probit model. Column (6) shows reduced form estimates after the probability of having a RTW teacher in a grade and school is randomly assigned. Standard errors are clustered at the school-by-grade level and are shown in parentheses. Stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Table 7: Instrumental Variable Estimates of the Effects of RTW Teachers on Standardized Reading Test Scores Using Different Probit Specifications

	Core Probit	Core Probit + SWD	Core Probit + SWD + Teacher Characteristics
	(1)	(2)	(3)
RTW	0.0198***	0.0187***	0.0189***
	(0.0068)	(0.0067)	(0.0066)
Observations	2,506,129	2,506,129	2,501,514
R-squared	0.6957	0.6957	0.6957

Notes: IV estimation results. The sample used for estimation is at the student-by-class-by-year level for the 2006-07 through the 2011-12 school years. The dependent variable is a student's standardized reading test score. "RTW" is a binary indicator equal to 1 for school-grade-years where RTW teachers worked and 0 otherwise. "RTW" is instrumented with $\widehat{RTW_{gst}} * Post_t$ in all specifications. All specifications include previous standardized test scores, student characteristics, predicted RTW, and school, grade, and year fixed effects. Student characteristics include female, Black, Hispanic, other race, LEP status, economic disadvantage, student with disability (SWD) indicators, gifted in reading, and indicators for whether a student repeated the previous grade or changed schools from the prior year. Omitted categories are White and Other Disability. "Predicted RTW" is the predicted probability of a school-grade-year having a RTW teacher from the probit regression in equation (4). In column (1), $\widehat{RTW_{gst}}$ is from the probit regression using only core variables. In column (2), $\widehat{RTW_{gst}}$ is from the probit with additional student with disability (SWD) variables. In column (3), $\widehat{RTW_{gst}}$ is from the probit specified with SWD variables and characteristics of the non-RTW teaching workforce. Standard errors are clustered at the school-by-grade level and are shown in parentheses. Stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Table 8: Instrumental Variable Estimates of the Effects of RTW Teachers on Standardized Math Test Scores Using Different Probit Specifications

	Core Probit	Core Probit + SWD	Core Probit + SWD + Teacher Characteristics
	(1)	(2)	(3)
RTW	0.0364***	0.0341***	0.0350***
	(0.0108)	(0.0106)	(0.0106)
Observations	2,404,535	2,404,535	2,399,908
R-squared	0.7208	0.7208	0.7208

Notes: IV estimation results. The sample used for estimation is at the student-by-class-by-year level for the 2006-07 through the 2011-12 school years. The dependent variable is a student's standardized math test score. "RTW" is a binary indicator equal to 1 for school-grade-years where RTW teachers worked and 0 otherwise. "RTW" is instrumented with $\widehat{RTW_{gst}} * Post_t$ in all specifications. All specifications include previous standardized test scores, student characteristics, predicted RTW, and school, grade, and year fixed effects. Student characteristics include female, Black, Hispanic, other race, LEP status, economic disadvantage, student with disability (SWD) indicators, gifted in math, and indicators for whether a student repeated the previous grade or changed schools from the prior year. Omitted categories are White and Other Disability. "Predicted RTW" is the predicted probability of a school-grade-year having a RTW teacher from the probit regression in equation (4). In column (1), $\widehat{RTW_{gst}}$ is from the probit regression using only core variables. In column (2), $\widehat{RTW_{gst}}$ is from the probit with additional student with disability (SWD) variables. In column (3), $\widehat{RTW_{gst}}$ is from the probit specified with SWD variables and characteristics of the non-RTW teaching workforce. Standard errors are clustered at the school-by-grade level and are shown in parentheses. Stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Table 9: Heterogenous Effects of RTW Teachers on Standardized Reading and Math Test Scores Across Students with Different Abilities

	Reading (1)	Math (2)
RTW*Quartile1	0.0705	-0.1206*
	(0.0598)	(0.0726)
RTW*Quartile2	-0.1820***	-0.1837**
	(0.0540)	(0.0767)
RTW*Quartile3	0.0071	-0.1658**
	(0.0527)	(0.0711)
RTW	0.0474	0.1696***
	(0.0376)	(0.0523)
Quartile1	-1.6901***	-1.7156***
	(0.0048)	(0.0058)
Quartile2	-0.9670***	-1.0605***
	(0.0048)	(0.0055)
Quartile3	-0.4761***	-0.5367***
	(0.0037)	(0.0050)
Observations	2,506,129	2,404,535
R-Squared	0.6561	0.6804

Notes: IV estimates of equation (5) with the following modifications. The previous test score is substituted for dummy variables indicating the quartile of the previous score within grade-years. Interactions of the quartile indicators and *RTW* are also included. The instruments for these interaction terms are $\widehat{RTW}_{gst} * Post_t * QuartileDummy$. The sample used for estimation is at the student-by-class-by-year level for the 2006-07 through the 2011-12 school years. The dependent variable is a student's standardized math test score. "RTW" is a binary indicator equal to 1 for school-grade-years where RTW teachers worked and 0 otherwise. "RTW" is instrumented with $\widehat{RTW}_{gst} * Post_t$ in all specifications. All specifications include student characteristics, predicted RTW, and school, grade, and year fixed effects. Student characteristics include female, Black, Hispanic, other race, LEP status, economic disadvantage, student with disability (SWD) indicators, gifted in math, and indicators for whether a student repeated the previous grade or changed schools from the prior year. Omitted categories are White, Other Disability, Quartile4, and RTW*Quartile4. "Predicted RTW" is the predicted probability of a school-grade-year having a RTW teacher from the probit regression in equation (4). Standard errors are clustered at the school-by-grade level and are shown in parentheses. Stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Table 10: Heterogenous Effects of RTW Teachers on Standardized Reading and Math Test Scores by Grade

	Reading					Math				
	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
RTW	0.034**	0.038**	0.035**	0.016	0.001	0.026	0.060**	0.059**	0.021	0.041*
	(0.017)	(0.016)	(0.015)	(0.016)	(0.014)	(0.024)	(0.024)	(0.029)	(0.023)	(0.021)
Observations	555,048	561,962	477,097	457,178	454,844	533,266	543,211	453,638	433,577	440,843
R-Squared	0.6902	0.6903	0.6988	0.7083	0.7074	0.7122	0.7316	0.7305	0.7461	0.7296

Notes: IV estimation results by grade. The sample used for estimation is at the student-by-class-by-year level for the 2006-07 through the 2011-12 school years. The dependent variable is a student's standardized reading or math test score. "RTW" is a binary indicator equal to 1 for school-grade-years where RTW teachers worked and 0 otherwise. "RTW" is instrumented with $\widehat{RTW}_{gst} * Post_t$ in all specifications. All specifications include previous standardized test scores, student characteristics, predicted RTW, and school and year fixed effects. Student characteristics include female, Black, Hispanic, other race, LEP status, economic disadvantage, student with disability (SWD) indicators, gifted in reading or math, and indicators for whether a student repeated the previous grade or changed schools from the prior year. Omitted categories are White and Other Disability. "Predicted RTW" is the predicted probability of a school-grade-year having a RTW teacher from the probit regression in equation (4). Standard errors are clustered at the school-by-grade level and are shown in parentheses. Stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Table 11: Marginal Effects of RTW Teachers on Suspensions and Detentions

	Out-of-School Suspension	In-School Suspension	Detention
	(1)	(2)	(3)
RTW	-0.0092**	-0.0381***	-0.0205***
	(0.0041)	(0.0097)	(0.0058)
Observations	2,268,593	1,917,085	1,917,085
R-squared	0.1135	0.1418	0.0971

Notes: IV estimation results. The sample is at the student-year level. Column (1) includes the 2006-07 through the 2011-12 school years, while columns (2) and (3) omit 2006-07 for data availability reasons. The dependent variables are indicators for whether the student was suspended (out-of-school or in-school) or received a detention during the school year. “RTW” is a binary indicator equal to 1 for school-grade-years where RTW teachers worked and 0 otherwise. “RTW” is instrumented with $\widehat{RTW}_{gst} * Post_t$ in all specifications. All specifications include student characteristics, predicted RTW, and school and year fixed effects. Student characteristics include female, Black, Hispanic, other race, LEP status, economic disadvantage, student with disability (SWD) indicators, gifted in reading or math, and indicators for whether a student repeated the previous grade or changed schools from the prior year. Omitted categories are White and Other Disability. “Predicted RTW” is the predicted probability of a school-grade-year having a RTW teacher from the probit regression in equation (4). Standard errors are clustered at the school-by-grade level and are shown in parentheses. Stars indicate statistical significance (* 10%, ** 5%, *** 1%).

Appendix

Table A1: Detailed RTW Policy Timeline

Law passed	Oct 30, 1998	Jun 30, 2000	Sept 26, 2001	Sept 30, 2002	Jul 20, 2004	Aug 11, 2005	July 31, 2007
Law effective	Jan 1, 1999	Jun 30, 2000	Jul 1, 2001	Sept 30, 2002	Jun 30, 2004	Aug 1, 2005	Oct 1, 2007
Expiration	Jun 30, 2003			Jun 30, 2004	Jun 30, 2005	Jun 30, 2007	Oct 1, 2009
Specifics of law:							
Restrictions on who can return with respect to their retirement date	None						No restrictions if retire prior to Oct 1, 2007; only those eligible for normal retirement if retire after Oct 1, 2007
Mandatory break in employment before returning to work	1 year (other than as a substitute teacher)	1 year immediately preceding reemployment (other than as substitute teacher)	6 months immediately preceding reemployment (other than as substitute teacher or part-time tutor)			6 months immediately preceding reemployment	
Restrictions on returning school	Must be low-performing	None					
Restrictions on returning employment	**Not permanent (only sub or interim)**	None					
Restrictions on returning teacher certification	Employed in area of certification; school in area where there is a shortage of teachers with beneficiary's certification	None					
% of returning salary that LEAs must pay to retirement system	0 %				11.7 %		

Notes: Source is Mahler (2013).

Table A2: Marginal Effects from Probit Regressions of RTW on School & Grade Characteristics Omitting the Last Year of the Policy

	Core (1)	+ SWD (2)	+ Teacher Characteristics (3)
Proportion Female Students in School-Grade	0.098	0.077	0.094
	(0.225)	(0.230)	(0.229)
Proportion Black Students in School-Grade	0.252	0.255	0.262
	(0.203)	(0.204)	(0.202)
Proportion Hispanic Students in School-Grade	-0.069	-0.040	-0.009
	(0.324)	(0.331)	(0.329)
Proportion Asian Students in School-Grade	-1.371	-1.349	-1.301
	(0.894)	(0.883)	(0.872)
Proportion Native American Students in School-Grade	0.792	0.844	0.830
	(0.660)	(0.657)	(0.661)
Proportion Economically Disadvantaged Students in School	-0.042	-0.066	-0.105
	(0.213)	(0.211)	(0.213)
Proportion Limited English Proficient Students in School	0.224	-0.046	-0.177
	(1.020)	(1.011)	(1.023)
Enrollment	-0.00001	-0.00006	-0.00011
	(0.00022)	(0.00022)	(0.00023)
Proportion Students in School with Emotional Disability		2.246	2.204
		(1.525)	(1.535)
Proportion Students in School with Learning Disability		1.630	1.520
		(1.154)	(1.159)
Proportion Students in School with Mental Disability		-1.539	-1.681
		(1.684)	(1.674)
Proportion Students in School with Physical Disability		-2.837**	-2.898***
		(1.128)	(1.119)
Proportion Students in School with Speech/Language Disability		-2.115	-2.589
		(1.976)	(2.021)
Average Experience of Non-RTW Teachers in the School in the Previous Year			0.004
			(0.016)
Proportion of Retirement Eligible Teachers in the School in the Previous Year			0.710
			(0.539)
Observations	2,176	2,176	2,169

Notes: These are marginal effects of the probit regression given by equation (4). The sample used for estimation is at the school-by-grade-by-year level for the 2006-07 and 2007-08 school years. Column (1) includes a set of core variables. Column (2) adds student with disability (SWD) variables. Column (3) adds aggregate characteristics of non-RTW teachers. Each column includes school, grade, and year fixed effects, and standard errors are clustered at the school-by-grade level and shown in parentheses. Stars indicate statistical significance (* 10%, ** 5%, *** 1%).