It's Not (Just) About the Money:
Pay and the Value of Working Conditions in Teaching[*]

Carolyn Tsao[†]

October 24, 2024

Updated frequently - [click here for latest version]

Abstract

This paper tests and quantifies whether teachers earn rents when taking into account both pay and non-wage amenities. Using quasi-experimental designs, novel administrative data, and a survey experiment, I estimate the gap in pay and the gap in the value of working conditions between teaching and teachers' next-best jobs, and sum the estimates to obtain the difference in overall compensation between jobs. Using a fuzzy regression kink design that leverages variation in who becomes certified to teach in Kentucky, I find that teaching pays a premium of $18-21,000/yr (33-40% of the teaching salary) above teachers' next-best options. Event studies around teacher exits yield similar results. Fielding a stated preference experiment survey to Kentucky teachers, I find that teaching offers relatively poor working conditions and estimate that teachers are willing to pay 29-35% of their salary to switch to their next-best job due to its better working conditions. Taken together, my findings suggest that teaching offers negligible rents; instead, teaching pays a large premium that mostly functions as a compensating differential, consistent with reports of teacher shortages today.

1   Introduction

Do public school teachers earn rents? The answer has significant fiscal implications as teacher salaries represent over half of public school spending, totaling over \$500 billion in the US in 2020-21. Even spending just a small fraction of teacher salaries on rents would cost billions of dollars. However, there are both reasons to believe, and to doubt, that teachers earn rents. On one hand, teaching is associated with amenities like pensions and job stability and is often supported by strong unions. On the other hand, widespread reports of "teacher shortages" and evidence that teachers earn less than other college-educated workers suggest that teachers may not be capturing rents at all [The New York Times, 2023, Allegretto and Mishel, 2018]. Consistent with this uncertainty, prior studies find mixed evidence of rent extraction in teaching (e.g., Hoxby [1996], Lovenheim [2009], Cowen and Strunk [2015], Allegretto and Mishel [2018]).

Many of these studies, however, are observational and leave out a crucial aspect of compensation: the value of working conditions. Given that workers place substantial value on non-wage aspects of jobs (e.g., Mas and Pallais [2017], Maestas et al. [2023]), it is necessary to account for both pay and the value of non-wage job attributes when assessing rents. Doing so is especially important when studying occupations with poor working conditions where workers may be paid a compensating differential, which could otherwise be mistaken as a rent. Teaching may in fact be one such occupation, as over 50% of teacher strikes in the past decade have demanded better working conditions in addition to higher pay, citing increased workplace stress and student behavioral issues [Lyon et al., 2024, The Atlantic, 2013].

This paper tests and quantifies whether teachers earn rents taking both pay and the value of working conditions into account. The answer comes down to comparing the *gap in pay* to the *gap in the value of working conditions* between teaching and teachers' next-best jobs. If teaching pays a large premium, the evidence could be consistent with rents. However, if teaching also offers less valuable working conditions, the evidence could instead be consistent with compensating differentials, with the pay premium functioning as compensation for poor working conditions in teaching.

Estimating the gaps requires addressing two key challenges: selection into jobs and data availability. Workers select into occupations for idiosyncratic reasons, making it difficult to identify teachers' next-best jobs using observational data. Even if workers were randomly assigned to teaching, estimating the pay gap would require panel data on workers' wages and occupations outside of teaching, which is difficult to come by. The data needed to estimate the gap in working conditions is even more difficult to come by, as there is no systematic data on the incidence of general (i.e., not school-specific) job attributes in teaching and no existing

estimates of teachers' willingness-to-pay for general job attributes. Indeed, the first dataset to systematically document differences in working conditions across a representative sample of U.S. workers only came into existence in 2015 [Maestas et al., 2023].

In this paper, I address all of the challenges mentioned above. In doing so, I provide the first comprehensive set of estimates of how much teaching differs from teachers' next-best jobs in terms of both pay and the value of working conditions in a U.S. context: specifically, in the state-wide labor market of Kentucky.

First, I identify and estimate the pay gap between teaching and teachers' next-best jobs. I use two complementary quasi-experimental designs—a fuzzy regression kink (RK) design and an event study design—to leverage as-good-as-random variation in who enters teaching as well as variation in who exits teaching. Each has its own advantage: the RK leverages as-good-as-random variation in the certification process and estimates the pay gap for inexperienced teachers, while the event study estimates the pay gap for more experienced teachers. I apply both designs to a unique novel panel dataset constructed by linking five administrative datasets, ultimately covering the near-universe of workers in the state of Kentucky.

In the main empirical design, the fuzzy RK, I circumvent selection in who is observed working as a teacher by leveraging quasi-experimental variation in the process that determines who obtains teacher certification in Kentucky. As is the case in many states, the first step towards teacher certification in Kentucky is to pass a standardized entry exam, which can be retaken. Over 40% of test-takers score near the cutoff, a product of the fact that the score is only used as a pass/fail indicator rather than as a continuous signal of quality. I show that for individuals who score near the exam cutoff on their first attempt, the exam creates two sources of variation in who becomes a teacher that are plausibly exogenous to test-takers' future earnings. First, the cutoff induces as-good-as-random variation in who passes or fails on their first attempt. Second, the cutoff combined with *the distance of one's first-attempt score from the cutoff* induces as-good-as-random variation in how much individuals are encouraged to retake the test. I thus use the interaction between *whether* and *how far* one falls to the right or left of the passing cutoff on their first attempt as an instrument for whether one becomes a teacher. I apply a statistical result from Dong [2018] to demonstrate that a *fuzzy regression kink design* (RK) identifies the treatment effect of becoming a teacher on future earnings.

My main finding from the quasi-experimental analyses is that in Kentucky, the teaching job offers a large pay premium compared to prospective teachers' next-best job options in the short run. From the RK, I find that for test-takers who score near the cutoff on their first attempt, their next-best in-state options four years later pay an estimated $18,000-22,000 less

per year than teaching does.[1] Those who pass the exam but do not pursue teacher certification are either not employed in the state four years later or are employed in industries such as education, health care, retail, or accommodation and food services. Similarly, event studies examining how annual earnings change around teacher exits show that exiting teachers tend to leave to the same industries and earn on average $20,000-25,000 less per year four years after exit. The estimates are large, amounting to 33%-40% of a Kentucky teacher's salary.

Second, I field a survey to a representative sample of teachers from Kentucky to elicit information about, and estimate the value of, the working conditions in teaching. To discipline my survey design and to ensure that my estimates can be benchmarked to willingness to pay estimates in the literature, the survey instrument I use is an exact replica of the American Working Conditions Survey (AWCS) conducted by Maestas et al. [2023], which was the first survey to systematically document differences in working conditions across a representative sample of U.S. workers. The first part of the survey asks workers about nine key job attributes[2], while the second part is a stated-preference experiment in which workers are presented pairs of jobs with randomized job attributes and asked which jobs they prefer. The stated-preference experiments are what allow me to estimate teachers' willingness-to-pay for each job attribute.

My first main finding from the survey is that teaching is significantly less likely to offer highly-valued working conditions compared to the jobs of other college graduates. Compared to the job of the average college graduate in the AWCS, teaching is significantly less likely to offer eight of the nine working conditions in the survey, including: more than 15 days of paid time off (5% in teaching vs. 69% in other college graduates' jobs), low physical demands, i.e. mostly sitting (1% vs. 58%), scheduling flexibility (12% vs. 68%), a relaxed pace of work (16% vs. 36%), and the option to telecommute (4% vs. 55%).[3]

My second main finding from the survey is that the average teacher in Kentucky is willing to pay 29-35% of their salary for a job that offers the desirable working conditions that teaching does not have. To obtain this estimate, I adapt the method used by Maestas et al. [2023] to estimate workers' willingness to pay for "the best job" over "the worst job" (i.e., for a job that offers all the most desirable conditions over a job that offers only the least desirable conditions). Using the choice data from the stated preference experiments, I estimate the average teacher's willingness to pay for each of the nine working conditions in the survey.

---

[1]The result is robust to varying bandwidths, but due to small sample size, the conservative 95% confidence intervals as recommended by Lee et al. [2022] include zero at smaller bandwidths. Nonetheless, even with the most conservative standard errors we are able to rule out large pay differences in the opposite direction.

[2]The attributes include: schedule flexibility, telecommuting opportunities, physical demands, pace of work, autonomy, paid time off, working with others, job training opportunities, and impact on society.

[3]In fact, the only positively-valued amenity that teaching is significantly more likely to offer is the opportunity to make a positive impact on one's community or society (64% vs. 39%).

Based on the finding from the pay gap estimation that teachers' "next-best jobs" are low-paying jobs in the education/healthcare sectors, I take the subset of below-median paying jobs in the education/health sector in the AWCS as a proxy for teachers' average "next-best" job. Then, following the best-to-worst job comparison in Maestas et al. [2023], I transform the estimated preference parameters to obtain teachers' willingness to pay for the attributes of the average "next-best" job over the attributes that the teaching job provides.

Taken together, my findings suggest that Kentucky teachers earn negligible rents. Instead, I find that teaching offers large pay premium that is almost entirely compensation for undesirable working conditions. Extending my results from Kentucky to other states suggests that over $145 billion is being spent annually to compensate teachers for poor working conditions. As such, identifying cost-effective ways to close the gap in working conditions between teaching and other jobs may be a more pressing policy issue than whether teachers earn rents. Education policy is taking steps that align with this conclusion, with the Biden administration calling on Congress in 2021 to invest $9 billion in improving both teacher pay and teaching conditions.

This paper contributes to three strands of literature. The first strand investigates whether public sector workers earn rents, and particularly how much public sector unions impact salaries through collective bargaining (e.g. Hoxby [1996], Lovenheim [2009]). Relative to this literature, I demonstrate a new approach to establishing whether public sector workers earn rents—by pinning down the gap in pay and the gap in working conditions—that does not require data on unions or bargaining, which can be hard to find. The second strand characterizes who becomes a teacher, wherein the key challenge is finding exogenous variation in job choice (e.g. Hoxby and Leigh [2004] use changes in unionization laws; Nagler et al. [2020] use business cycles). I use two new sources of variation—quasi-experimental variation in teacher certification and experiment-induced variation—to document new facts about teachers' next-best jobs and their preferences over working conditions. Finally, the third strand investigates the pay gap between teaching and teachers' next-best jobs by using event study designs to examine whether exiting teachers leave to higher- or lower-paying jobs [Goldhaber et al., 2022, Champion et al., 2011, Scafidi et al., 2006, Stinebrickner, 2002]. I add to the literature a new set of estimates of the pay gap using variation in who exits teaching *and* new variation in who becomes a teacher.

The rest of the paper proceeds as follows. Section 2 describes how the pay gap is estimated, from data and methods to results. Similarly, Section 3 describes how the gap in the value of working conditions in estimated. Section 4 discusses the implications of our estimates, as well as how the estimates can be reconciled with previous findings that may seem at odds with our results. Section 5 concludes.

2 Estimating the Pay Gap

2.1 Identification Challenge

We are interested in estimating the pay gap between teaching and teachers' next-best job options. The identification challenge is selection into teaching: the decision for one to become a teacher or to become qualified to teach, i.e. by earning a teacher certification, is endogenous to earnings potential.

To circumvent this challenge, we leverage two sources of variation in who is a teacher. First, we leverage exogenous variation in whether one gets certified to teach, induced by the cutoff on one of the required exams for entry into teacher preparation programs in Kentucky. This approach identifies the treatment effect of becoming a teacher or getting certified on short-term earnings, or equivalently, the pay gap between teachers' next-best options and teaching for inexperienced teachers. Second, we leverage job switches induced by exits from teaching to other jobs. This approach allows us to estimate the initial gap in pay that arises after a teacher move—in other words, the pay gap between teachers' next-best options and teaching for experienced, exiting teachers.

The following sections provide background on the key features of the teacher certification process that we leverage in our identification strategy, as well as the data that we use, before detailing the empirical strategies and results.

2.2 Key Features of the Teacher Certification Process

In most U.S. states, the traditional[4] path to becoming eligible to teach in public schools is a multi-year process colloquially known as the "teacher pipeline." The process typically consists of three stages: (1) enroll in an educator preparation program (EPP), which may have an entry exam requirement, (2) graduate from the EPP and thus earning a state-approved teaching certificate, which may have an exit exam requirement, and (3) apply for, be offered, and accept a position in a public school as a teacher. EPPs are commonly comprised of up to two years of coursework, which college students may opt to combine with the last two years of their four-year

---

[4]Over the past decade, a combination of political pressures during the Obama administration and general discontent in the teaching profession has led to increased demand and supply of individuals obtaining certification through "alternative routes," which allows individuals to bypass the multi-year traditional pipeline to become "emergency certified," etc., under the premise of filling slots in schools that need to be filled. However, the vast majority of public school teachers today hold a traditional teaching certificate earned through the traditional pipeline.

college degree.[5]

The majority of state-approved EPPs use Praxis exams as EPP entry and/or exit exams.[6] "The Praxis" is a common name used to describe all the teaching-related standardized tests offered by the testing institution ETS, which offers many other tests including the GRE and the SAT. The Praxis tests include a wide range of exams that test skills ranging from core understanding of math, reading, and writing to specific knowledge in fields of study (e.g. American history, geography) or levels of education (e.g. elementary school modules). Not all of the Praxis exams test pedagogical knowledge—some, like the Praxis CASE math exam, test the test-takers' middle-school level proficiency in mathematics.

In all states that use the Praxis, the Praxis is only used as a means for test individuals for admission to, or graduation from, teaching preparation programs.[7] However, states vary in which Praxis exams they require at what stages of the pipeline and in what scores constitute passing scores. State departments of education and ETS are transparent about these differences online.[8] In Kentucky, our setting of study, the teacher pipeline consists of the same three stages as the typical U.S. state.

We focus on one particular set of Praxis exams: the Praxis CASE ("Core Academic Skills for Educators"), the Praxis' core exams in math, reading, and writing. The Praxis CASE exams are commonly used as an entry requirement to EPPs; in Kentucky for example, the state mandates that individuals must pass the Praxis CASE exams to be admitted to an EPP. As such, most test-takers who attempt one of the three exams opt to sit all three tests at once. The pricing scheme also incentivizes taking all three at once: each exam can be taken on its own for $90, or the whole suite can be taken for $150. Retaking is allowed and technically unlimited, although test-takers must wait 28 days between attempts and face the same pricing schedule. Based on national Praxis statistics and the pricing schemes of test preparation programs, the math component is the most frequently retaken component.[9] [10]

The test design makes it difficult for the test-taker to predict or target a specific score. The

---

[5]Some states also require that students engage in a form of teacher training after the EPP before they receive their full certificate, but this is relatively rare cite examples.

[6]As of now, twenty-six states require certain passing scores on the Praxis at some point in the teacher pipeline. Others, e.g. Texas, also require exams but write and administer their own state exam.

[7]This feature makes the Praxis unlike other exams offered by ETS like the GRE or the SAT, which can be used to apply to wide ranges of college majors.

[8]See websites. As of April 2023, twenty states/provinces use the Praxis CASE exam in their teacher pipeline, the majority of which utilize a cutoff of 150 on the math exam, 156 on the reading exam, and 162 on the writing component.

[9]Magoosh, for example, offers two programs: a 6-month preparation program for all three exams at $99, and a 6-month math-only program for $79.

[10]A sizeable minority accept students on the condition that they pass the CASE during the program. However, regardless of conditional admittance or not, students must pass all three tests in order to graduate.

exams are timed and administered on computers in monitored exam rooms with no Internet access. Scores are computed and reported immediately to the test-taker at the end of the exam. Each exam question carries its own weight corresponding to its difficulty level. The weights are not revealed to the test-taker but are instead used to convert the raw scores into standardized scores, which range from 100 to 200. The timed nature of the exam introduces time pressure—test preparation programs advise test-takers on strategies to best manage their time and try to maximize their score rather than aim to pass.

The use of the Praxis CASE in Kentucky has varied over the past few decades. Since the mid-2000s, the Educator Professional Standard Board (EPSB), which governs the certification rules for teachers in Kentucky, has required that applicants pass a version of the Praxis CASE to enter state teacher certification programs. However, the state's usage of the Praxis CASE has become more standardized over time. Prior to 2012, most EPPs in Kentucky required the earlier version of the Praxis CASE, the "Praxis I," for entry, but set their own cutoff scores. In 2012, the state mandated a common set of cutoff scores for all EPPs. Finally, in 2014, the Praxis I was discontinued by ETS, resulting in the Praxis CASE becoming the only set of required exams for entering programs with common cutoffs across the state. Test-takers could not "mix and match" test scores from the Praxis I and Praxis CASE—to be eligible for an EPP, they had to take the complete suite of Praxis CASE exams. As of April 2023, ETS has not made any more changes to the Praxis CASE since its introduction in 2013.

In recent years, the necessity of the Praxis in the teacher pipeline has been brought into question by Kentucky governance. In a discussion on the state of the teacher pipeline in an interim legislative session in June 2022, legislators raised concerns that the Praxis was serving as more of a barrier than an informative screening device into teaching. In response, the EPSB lowered the cutoffs on the Praxis CASE in January 2023.

2.3   Matched Employer-Employee Data

Our main administrative data is comprised of five statewide datasets from Kentucky linked using SSNs. We start with data on Praxis test-takers and their progress along the teacher pipeline comes from the Educator Professional Standards Board (EPSB). For each person, these data include the date and scores on all Praxis attempts and annual indicators on the type (e.g. full, emergency, temporary) and validity (i.e. valid or expired) of their teacher certification status.

To obtain information on postsecondary enrollment and completion, we merge these data with person-year level records from the Council on Postsecondary Education (CPE) and the Kentucky Higher Education Assistance Authority (KHEAA) to obtain person-year records for

whether the person is enrolled in an in-state postsecondary institution, holds a postsecondary degree, or holds a Pell grant.

To obtain information on childbirth, we merge these data with Vital Statistics data, giving person-year records for whether a person appears as a parent (father or mother) on a birth certificate for a newborn child in a given year.

To obtain information on general employment, we again merge these data with the Kentucky UI database on all workers between the ages of 18 to 60, giving us quarterly earnings, 6-digit industry NAICS codes associated with each source of quarterly income, and quarterly indicators for whether one is working in a public school district. We create a crosswalk between firm names under the education NAICS code and the official list of private schools on the Kentucky Department of Education (KDE) website to identify private schools where possible.[11] The Kentucky UI wage records have the standard UI data limitation in that it does not separately identify individuals who are unemployed from individuals who might be working outside of the state; as a result, we refer to individuals in either category as "non-employed."

Finally, to obtain details on public school employment, we merge these data with the Kentucky Longitudinal Data System maintained by KYSTATS to bring in person-year records on all staff in the public school system, including their district and school of employment, occupation within the school system (e.g. teacher or principal), and links to classrooms and students' standardized tests. These data allow us to observe when individuals move between schools or districts or change occupations within public schools.

Our final sample includes the union of all individuals who ever attempted the Praxis in Kentucky between 2008 and 2022 and all individuals who ever appear as employed in the Kentucky UI data system between 2009 until 2022. Throughout, we obtain age, gender, race, and ethnicity from whichever database the person first appears in, with near full coverage.

Because our linked administrative data does not contain information on occupations outside of the public school system, we supplement our analyses of our administrative data with an analysis of the matched sample in the Current Population Survey (CPS), which contains panel data with two observations one year apart for a nationally representative sample of individuals. From 2009 to 2023, nearly 5,000 individuals in the matched CPS report being a teacher in the first round but not being a teacher in the second. We are able to observe the second occupation for nearly all of these individuals.

---

[11]Details in the appendix.

2.4 Descriptive statistics on Praxis test-takers

The main two analysis samples we study are Praxis test-takers from our administrative data and survey respondents from our own survey.

Table 1 shows sample means describing all individuals who attempted the Praxis CASE for the first time between 2013 and 2017 in Kentucky.[12] Panel A shows key demographic characteristics. First-time test-takers are on average in their third year of college[14] and are mostly women and white, in similar proportions to the demographic composition of teachers in the state. One-fifth hold a Pell grant—lower than the state average—and two-thirds are working in the year prior to taking the exam.

Table 1: Summary statistics on main Praxis sample

	Mean	S.D.	N
Panel A: Sample characteristics at first attempt			
Age	24.35	7.57	10,127
Female	0.76	0.43	10,127
White	0.92	0.28	10,127
First-time enrollee in post-secondary institution	0.77	0.42	10,127
Holds a Pell grant	0.20	0.40	10,127
Earnings 1yr prior (2018 USD)	11,384.91	12,917.88	6,744
Panel B: Praxis performance and retaking			
Math score (passing cutoff: 150)	158.03	23.50	10,127
Pass math on first attempt	0.64	0.48	10,127
Among those who fail on first attempt...			3,645
% who retake at least once	0.80		3,645
% who pass on first retake	0.33		3,645
Total number of attempts at the Praxis	1.63	1.39	10,127
Panel C: Pipeline progression			
Attempt Praxis for the first time	1.00		10,127
Ever pass entire Praxis CASE	0.80		10,127
Enroll in an EPP within 3 years of first attempt	0.72		10,127
Get certified within 4 years of first attempt	0.53		10,127
Become teacher within 4 years of first attempt	0.46		10,127

Note: Sample includes Praxis first-attempters between 2013-2017, inclusive.

Panel B summarizes statistics related to Praxis performance and retaking behavior.[15] The

---

[12]We choose the starting date of 2013 because the Praxis CASE was revamped in 2013 but has not changed since, allowing us to cleanly identify whether an individual is attempting the test for the first time.[13] We choose the end date of 2017 to allow our analysis to look at the short-term employment and earnings outcomes for these individuals four years out.

[14]Although the mean age is 24, the median age is 21, consistent with our understanding that individuals are most likely to take the Praxis CASE in the third year of college.

[15]Should be able to update to include 2018 when the 2022 UI data is available.

majority (80%) of test-takers eventually pass the entire Praxis CASE.[16] However, these passes are not all obtained on the first attempt: for the least-passed component, math, just under two-thirds of the first-time test takers pass on the first attempt. The gap between the initial and eventual passing rate is explained by retakes: 80% who do not pass on the first try attempt it again, and around 80% again who do not pass on the second try attempt it again. The pass rate dwindles as attempts increase, from 35% on the first attempt to 30% on the second attempt and 25% on the third. The majority (93%) of first-time attempters retake the exam no more than three times.

Panel C shows that there is significant "leakage" of individuals from the pipeline over time. Only 53% of all first-time attempters become certified to teach within four years of taking the Praxis. Most of the leakage occurs at the Praxis stage and at the certification completion stage; 87% of individuals who get certified find work as a teacher shortly thereafter.

2.5 Empirical Strategy I: Regression Kink Design

In this section we describe the fuzzy RK, our main empirical strategy for estimating the pay gap between teaching and teachers' next-best options.

2.5.1 Identification

Our identification argument proceeds as follows. We restrict our attention to first-attempt scores only. Under the assumption that test-takers cannot precisely manipulate their test scores, those who score near the cutoff on their first attempt are comparable in every way except that (1) those on the right can continue directly on to certification, while (2) those on the left can only continue on if they retake and pass the exam. Furthermore, under the assumption that there is an "encouragement-to-retake effect" of scoring near the cutoff that is increasing one's initial score, those who score just below the cutoff will be more likely to retake the exam than those who score epsilon further below the cutoff.

In sum, for those who score near the cutoff on their first attempt, the cost of certification is exogenously varied by *whether* and *how far* they are from the cutoff. Under these assumptions, the interaction between whether one passes on their first attempt and how far they score from the cutoff on their first attempt may be a valid instrument for whether one gets certified.

We demonstrate that both assumptions appear to hold and that the instrument is valid—that is, relevant and excluded—in Section 2.5.4.

---

[16]During our sample period, the cutoff in Kentucky on the math section is 150, below the mean math score of 156. For comparison, most states that require the Praxis CASE in some form implemented the same math cutoff of 150, and the mean math score in Kentucky was similar to the nationwide mean.

2.5.2 Why an RK and not an RD?

From the institutional details of our setting, it is not immediately obvious why it is appropriate to use an RK and not appropriate to use an RD in our setting. The underlying principle of the fuzzy RD design (Hahn, Todd, and Van der Klaauw 2001) is that as long as there is randomization in the running variable, then even with imperfect compliance, the running variable can be used as an instrument for treatment. In theory, the same principle can be applied to our setting. Each attempt at the exam is a sharp RD, and while retaking introduces fuzziness and possibly a kink, there may be a discontinuity at the passing cutoff that can be used to instrument for treatment status. This is in fact precisely the approach taken by Jepsen et al. [2016], who use the first score on the GED as the running variable in a fuzzy RD to evaluate the long term effect of earning a GED on earnings.

However, our setting deviates from the classic fuzzy RD setting in that we find a kink and virtually *no discontinuity* in the likelihood of certification at the cutoff. Applying a statistical result from Dong [2018], we formally demonstrate that the kink combined with the lack of discontinuity allows us to identify the treatment effect of certification on earnings. The proofs (Appendix B) show that (1) if there is both a discontinuity and a kink, a fuzzy RD but not a fuzzy RK will identify a weighted average treatment effect, and (2) if there is a kink but no discontinuity, a fuzzy RK but not a fuzzy RD will identify a weighted average treatment effect. Guided by these results, we use a fuzzy RK in our main analysis.

In the case of the RK, the key identifying assumptions needed in addition to those needed for the traditional RD is that among those who fail on their first attempt, the likelihood of retaking is increasing in one's first-attempt score (the "encouragement" effect) and in one's intrinsic desire to get certified (the "intrinsic motivation" effect), such that the likelihood of treatment near the first-attempt cutoff is the same on both sides.

As Appendix C shows, the possibility for retaking creates an additional source of identifying variation: a kink in the likelihood of treatment at the cutoff. Those who do not pass on their first attempt are increasingly likely to be treated as a function of their initial score, while those who pass on their first attempt are equally likely to be treated.

The discontinuity can only be zero if and only if retaking behavior is determined by both the "encouragement" and "intrinsic motivation" effects described above. We find supporting evidence that the "encouragement" effect is at play in our setting: the likelihood of retaking increases in the first-attempt score, even after controlling for a host of baseline test-taker characteristics including gender, race, earnings, age, and whether one holds a Pell grant. Finally, because there is no discontinuity in the likelihood of certification at the cutoff despite not ev-

eryone below the cutoff retaking the exam, it must be that the "intrinsic motivation" factor is at play as well.

2.5.3 Estimating equations

Leveraging the lack of discontinuity and the kink in the likelihood of getting certified at the cutoff, we use a fuzzy RK design to estimate the effect of certification on earnings via 2SLS:

$$\text{Structural eqn.} \quad Y_{i,t+4} = \quad \beta_0 + \beta_1 C_{i,t+4} + \beta_2 S_{it} + \epsilon_{it} \tag{1}$$

$$\text{First stage.} \quad C_{i,t+4} = \quad \alpha_0 + \alpha_1 \underbrace{(D_{it} \times S_{1i})}_{Z_{it}} + \alpha_2 S_{it} + \eta_{it},$$

where $D_{it} = \mathbb{I}(S_{1i} \geq 0)$. We interpret β_1/α_1 as a weighted average treatment effect of becoming certified on the outcome $Y_{i,t+4}$, earnings four years after the first test attempt.

LATE interpretation. The RD and RK estimands can be interpreted as the weighted average treatment effect across the distribution of individual types (DiNardo and Lee). Each estimand uses different weights, suggesting different interpretations.

To shed light on these differences, we turn to the alternative interpretation of the 2SLS estimator as the "local average treatment effect" (LATE). The LATE is the average treatment effect for "compliers," those who let the instrument dictate their receipt of the treatment.

First, consider the RD where the instrument is passing math on the first attempt and the treatment is ever passing the complete Praxis. Compliers are individuals for whom their performance on the math test on their first attempt perfectly predicts whether they pass all three exams: (1) if they pass math on their first attempt, go on to pass all three exams, but (2) if they fail math on their first attempt, never pass all three exams.

Second, consider the RK where the instrument is passing math on the first attempt interacted with one's score, and the treatment is getting certified. The concept of a complier does not naturally carry over into a setting where individuals can take multiple attempts at treatment. What does it mean for one to "comply" with their original passing or failing status if they can retake the exam? In Appendix D we borrow the classic concepts of compliers/always-takers/never-takers/defiers and re-cast the definitions to take retaking behavior into account. Combined with our identification proofs in B, we find that the RK estimator is a weighted average treatment effect for what we call "re-taking compliers": individuals whose choice to retake is responsive to their first-attempt score.

(a) Conditional on passing read/write on first attempt.
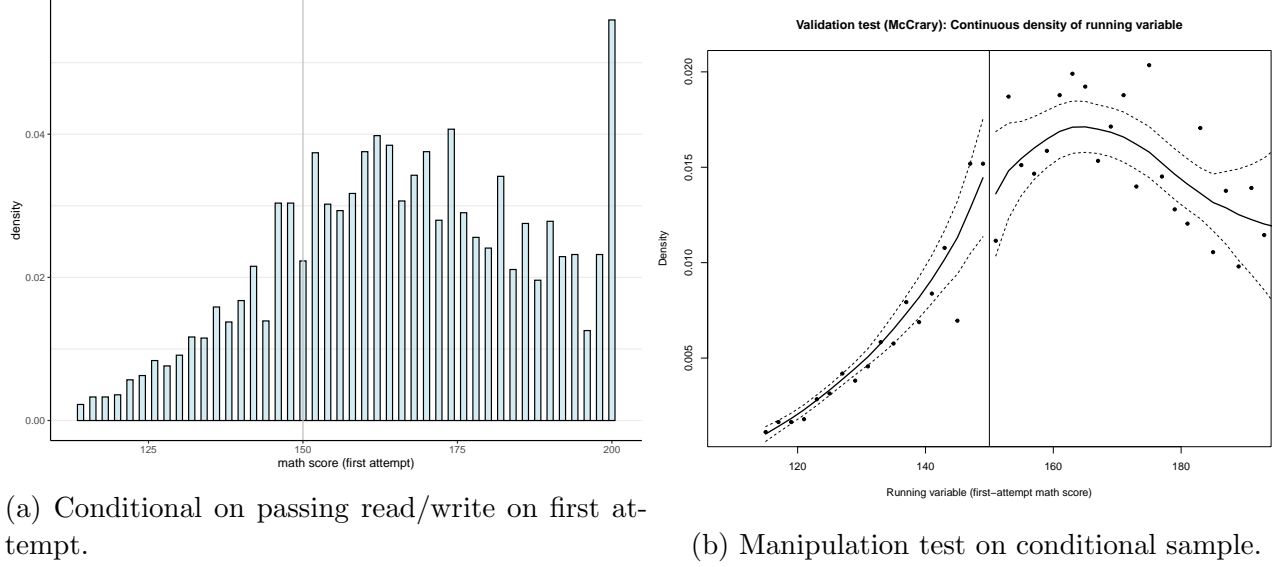
(b) Manipulation test on conditional sample.

Figure 1: Histogram of first-attempt math scores and McCrary (2008) test for manipulation, plotting fitted polynomials to the right and left of the test score cutoff.

2.5.4 Evaluating the quasi-experiment

Our empirical strategy is to focus on individuals who score near the passing threshold on the Praxis CASE and treat whether they pass or fail as a near experiment. This strategy requires that scores are as good as randomly assigned near the passing threshold. In this section we provide tests of these assumptions.

Smoothness in the running variable. A key assumption for valid inference in both the RD and RK design is that the density of the running variable—in our case, the first-attempt math score—is smooth through the cutoff. Were there a kink or a jump in the discontinuity, we might suspect endogenous sorting of individuals around the passing cutoff, which would invalidate the design. Figure 1 shows the distribution of first-attempt math scores around the cutoff of 150 for the main analysis sample, along with fitted polynomials to the right and left of the cutoff.[17] There appears to be no jump or kink in the distribution at the cutoff.[18]

Balance. Another implication of no sorting around the threshold is that individuals on either side of the cutoff are "balanced" on pre-determined characteristics prior to taking the Praxis for the first time. We examine this in three ways: (1) by visually examining the distributions

---

[17]Each bin represents a single score; gaps arise between the bars because only even number scores are possible on all components of the Praxis CASE.

[18]Because the test scores can only take on integer values, the formal McCrary test [McCrary, 2008] for manipulation around the cutoff reports discontinuities at every point.

(a) Covariates index: Pr(employed in 4yrs), predicted.
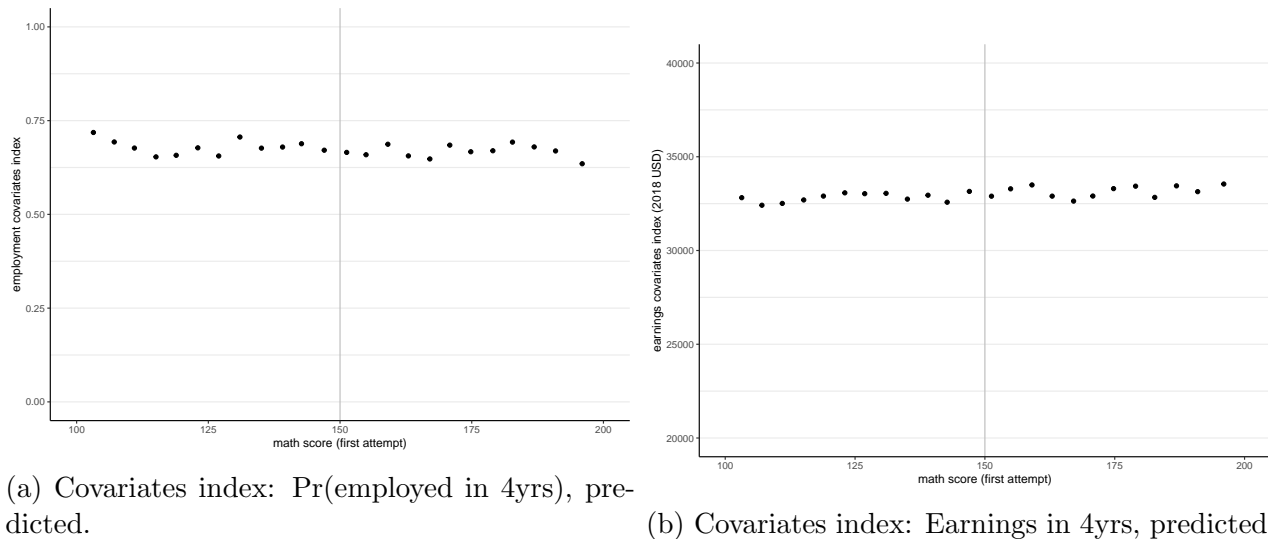
(b) Covariates index: Earnings in 4yrs, predicted.

Figure 2: Tests of selection around the cutoff on baseline covariates.

of each of a large set of baseline characteristics over first-attempt scores, including gender, race (Black, White, Hispanic, and others), age, whether one holds a Pell grant, whether one is enrolled in a Master's or Bachelor's program, whether one is already working, the industry one is employed in if so (education, retail, healthcare, and others), and the year and month of the exam; (2) by combining the baseline characteristics into a single "covariates index" —i.e. predicting the outcome of interest (earnings, employment, or industry) using baseline characteristics only, as in Card et al. [2015]—and testing whether the index is predictive of the treatment (Figure 2), and (3) by conducting paired t-tests comparing differences in baseline characteristics of individuals who score within 10-15 points to the right and left of the cutoff. All three tests suggest that there is balance on baseline characteristics.

2.5.5 Evaluating instrument validity

We argue that the instrument is both relevant and excluded.

Relevance is satisfied by the pattern repeated in Figure 3: the likelihood that an individual becomes certified (or a teacher) is non-decreasing in their first-attempt score. Interestingly, the increasing portion of the function is driven by the retaking behavior among the test-takers. Figure 4 shows that conditional on not passing on one's first attempt, the probability of retaking the exam is increasing in one's first-attempt score. We interpret this as an "encouragement" effect of scoring closer to the cutoff: the closer one scores initially, the more encouraged they are to retake the test in the hopes of passing on a future attempt.

For the exclusion restriction to be satisfied, it must be the case that the first attempt score

cannot affect earnings through any channel other than through its effect on the likelihood one gets certified. Perhaps the most compelling story that would violate the exclusion restriction would be if those who failed initially and chose to retake became, through the retaking process, additionally motivated to be a teacher or gained additional human capital in the process of studying, both of which could increase their future earnings potential. This story however seems unlikely to hold true given that test-takers need not wait long to retake the test (the modal wait time is 28 days) and given that the entry Praxis exam tests middle-school math and reading skills. We therefore conclude that it is reasonable to assume that the exclusion restriction holds.

(a) Ever pass math.

(b) Ever pass whole Praxis CASE.

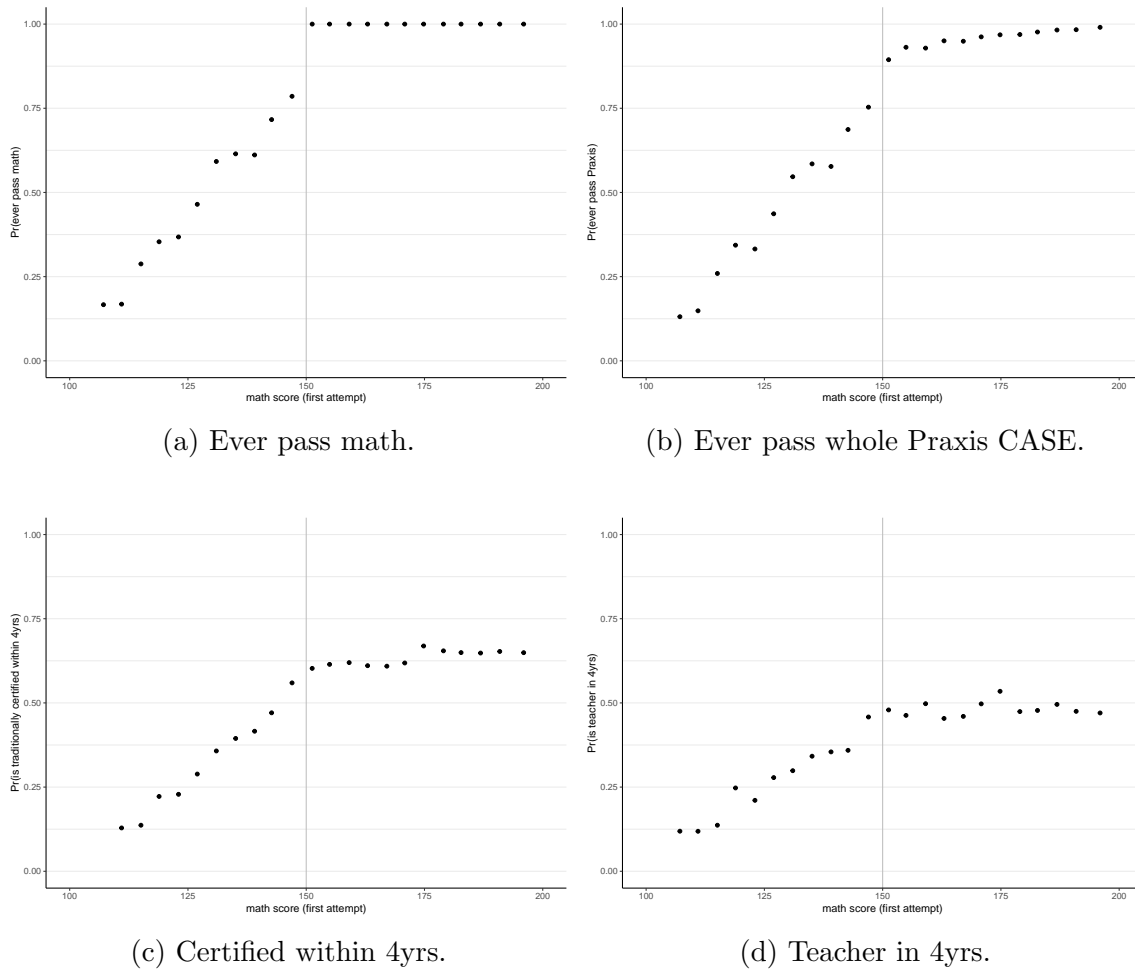(c) Certified within 4yrs.

(d) Teacher in 4yrs.

Figure 3: "First stage": Binscatters of the likelihood of reaching various stages along the teacher pipeline, plotted over bins of first-attempt Praxis CASE math scores.
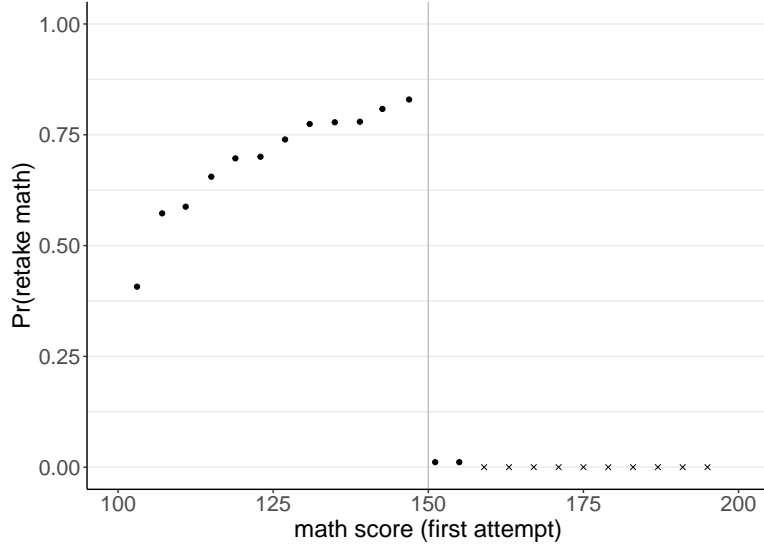
Figure 4: Binscatter of the share of individuals who retake the exam at least once, plotted over first-attempt math scores. X's represent cells with too few observations to show.

2.6   Regression Kink Estimates of the Pay Gap

2.6.1   First-stage and reduced form plots

Figure 3 shows the first-stage relationships between the first-attempt test scores and the likelihood of reaching any point of the teacher pipeline: ever passing the Praxis CASE, enrolling in an EPP, getting certified, and becoming a teacher. Our main treatment of interest is getting certified, but for completion we show the progression of first-stage relationships that we are able to observe.

Each plot shows binned averages over first-attempt scores of the observed likelihood that individuals reach a given point in the teacher pipeline. There is a strong discontinuity at the cutoff in the likelihood that one ever passes math. However, the discontinuity shrinks to virtually zero as we progress along the pipeline. In its place, there emerges a kink at the cutoff in the likelihood one reaches any point of the pipeline. Due to our small sample sizes, the corresponding F-stats are small (Figure G.1).

We also show graphical evidence of the "reduced-form" relationship between the first-attempt test scores and our outcomes of interest—earnings and attachment to the labor force four years after attempting the Praxis for the first time—in Figure 5. There appears to be no discontinuity or kink at the test score cutoff in the likelihood that individuals are employed, suggesting no effect of engaging in the teacher pipeline on being in the workforce. However, there is a weak kink at the cutoff in short-run earnings.

(a) Pr(employed in 4yrs in KY).

(b) Earnings in 4yrs (2018 USD).
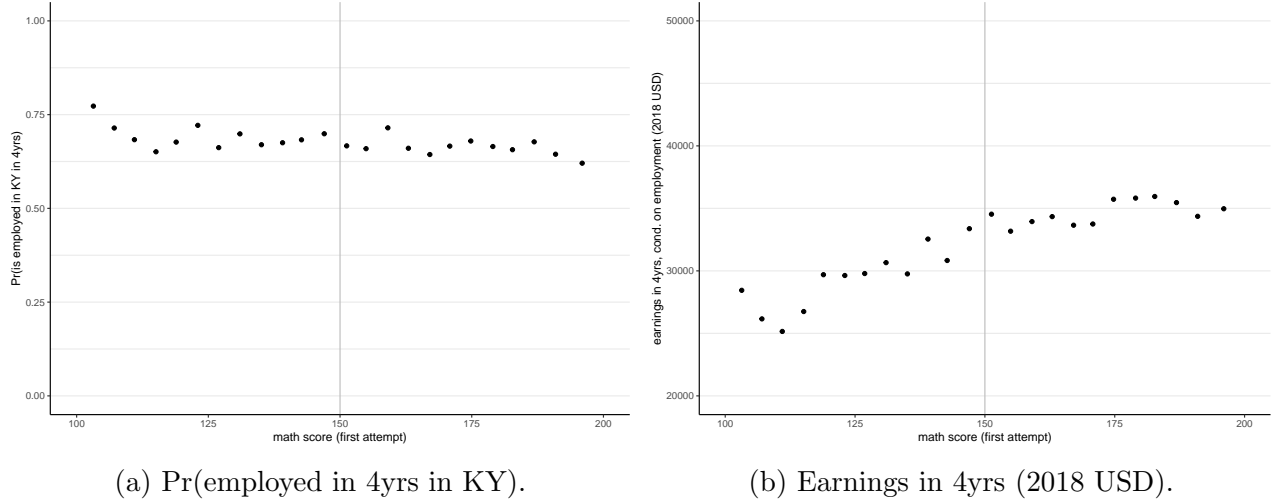
Figure 5: "Reduced form": Binscatters of likelihood of being employed and earnings 4 years after one's first attempt on the Praxis CASE math, plotted over bins of first-attempt math scores.

2.6.2 Treatment effect estimates

We now present our RD and main RK estimates. We maintain a hands-above-the-table approach when it comes to bandwidth selection: we estimate the weighted average treatment effects for a wide range of bandwidths, ranging from an under-powered bandwidth of two up to a large bandwidth of thirty. We show the estimates at the optimal bandwidth recommended by Calonico et al. [2020] in Table 2 and for the full range of bandwidths in Figure 6.

Table 2: IV Estimates of the Effects of Passing the Praxis, Becoming Certified, and Becoming a Teacher on Employment and Earnings

	Outcome: Pr(Employed in 4yrs)		Outcome: Earnings in 4yrs, cond. on emp.	
	First Stage (1)	Struct. Model (2)	First Stage (3)	Struct. Model (4)
Treatment A: Passing the entire Praxis				
RD estimate (conventional std. error)	0.12 (0.02)	-0.25 (0.23)		
VtF confidence interval		[-.7, .2]		
Treatment B: Becoming Certified				
RK estimate (conventional std. error)			-0.010 (0.004)	21,685.2 (11,380.9)
VtF confidence interval				[-621.4, 43,991.8]
Treatment C: Becoming a Teacher				
RK estimate (conventional std. error)			-0.012 (0.004)	18,329.5 (8232.3)
VtF confidence interval				[2194.2, 34,464.7]
Bandwidth	14	14	14	14
N	4,352	4,352	2,958	2,958

Note: All estimates are reported for the full sample of first-time Praxis test-takers between 2013-2017, using a bandwidth of 14 (i.e. comparing individuals with scores 136, 138, 140, 142, 144, 146, 148 to those with scores 150, 152, 154, 156, 158, 160, 162, 164). Conventional standard errors are reported in round brackets, and VtF confidence intervals are reported in McCrary et al. (2024).

Figures 6 shows the RK treatment effect estimates for becoming certified across a range of bandwidths. The RD treatment effect estimates for passing the Praxis are also shown in Figures G.2 and G.3. We compute the MSE-optimal bandwidth and indicate the resulting value with a dotted line [Calonico et al., 2020]. The estimates are shown Table 2 for a fixed bandwidth of fourteen, at which the estimates begin to stabilize. From here onward we focus on discussing the stabilized estimates.

The RD estimates in Figure G.2 suggest that there is no effect—or if anything, a slight negative effect—of passing the Praxis on whether one appears as employed in the state UI system in the short-run. Given this null effect, in our analysis of treatment effects on earnings, we restrict our sample to those who are employed (i.e. have non-zero earnings). Again, under the RD in Figure G.3, there appears to be no effect of passing the Praxis on earnings.

Our RK estimated effects suggest that getting certified has no effect on the likelihood of

short-run employment (Figure G.4), either. However, the main RK estimates shown in Figure 6 suggest there is a large, positive effect of getting certified on earnings, around $17,000-$25,000.

(a) First stage: Becoming certified.

(b) First stage: Becoming a teacher.

(c) Effect of becoming certified on earnings.

(d) Effect of becoming a teacher on earnings.
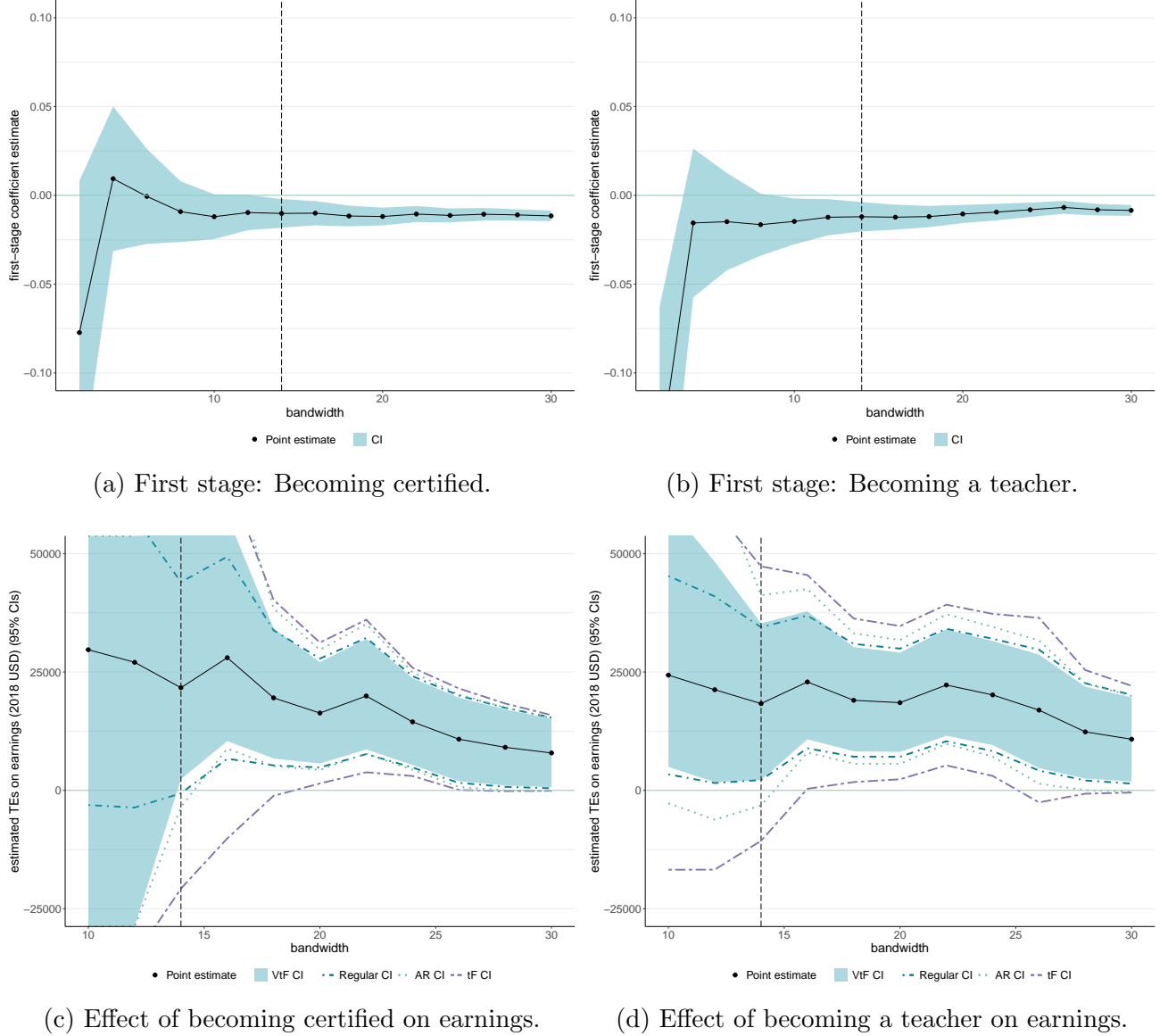
Figure 6: RK estimates of the effect on earnings in 4yrs of becoming certified or a teacher, over bandwidths ranging from 2-30.

2.6.3   Interpretation

The differences between our RD and RK estimates are intuitive given the context. We expect the skills developed along the pipeline to have the most significant pay-off for those who complete their training. Indeed, given that the majority of those who become certified also immediately become teachers, a key driver of the earnings premium of getting certified appears to be finding

employment as a teacher.

Our RK results indicate that for "retaking compliers" who let their first-attempt score dictate whether they retake the exam, there is a large positive premium of getting certified. Implicit in this interpretation is a form of monotonicity in the retaking behavior: we assume "retaking never-takers" do not exist—that is, people who would only retake the exam if they passed on their first attempt and would otherwise not retake the exam. The RK estimate also does not apply for "retaking always-takers," that is, people who would retake the exam no matter what, even if they had passed on their first attempt. Between "retaking compliers" and "retaking always-takers," one might expect the always-takers to have the larger treatment effect, in which case our RK estimates provide a lower bound on the positive treatment effect of certification on short-run earnings.

2.6.4 Robustness checks

Due to the size of Kentucky and the proximity of many of the major counties to the border of neighboring states, the state education department has a unique agreement with neighboring states allowing teachers to carry their certification across state borders. As a result, we omit one college institution from our sample—Northern Kentucky University—which is particularly known to send its graduates across Kentucky's border to work in Cincinnati, Ohio. The results are not sensitive to this omission.

In addition, because of our small sample sizes and resulting weak first-stage in the RK, we present the classic robust 2SLS standard errors as well as Anderson-Rubin confidence intervals (CIs), tF CIs resulting from implementing the Lee et al. (2021) tF procedure, and the VtF CIs from implementing the Lee et al. (2024) VtF procedure. As expected, the tF confidence intervals are by far the widest, encompassing zero for bandwidths less than fourteen. However, our preferred estimates using the VtF intervals closely correspond with the regular CI estimates and are narrower than the AR CIs.

The point estimates also do not change dramatically when we investigate the effect on longer-run earnings five or six years after taking the test, although the sample size becomes considerably smaller given how recent our testing data is.

2.7   Empirical Strategy II: Event Study

The second strategy we use to examine the pay gap between teaching and teachers' next-best options is an event study design. Specifically, we examine how pay changes for teachers who we observe exiting teaching and changing jobs in two data sources: our Kentucky administrative

data, and in the Current Population Survey (CPS) matched sample. The advantage of the event study over the RK is that the event study permits us to "get away from the (test score) cutoff," in that we do not expect that leaving teachers are necessarily those who would have scored near the passing cutoff on the Praxis. The disadvantage of this design however is that the resulting estimated pay changes cannot be interpreted as causal, as moves may be endogenous to pay. Nonetheless, the exercise provides complementary evidence to the RK estimates using a different source of variation.
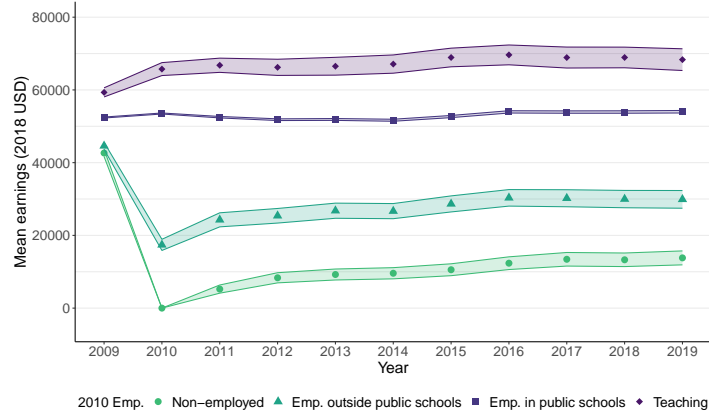
(a) Annual earnings.
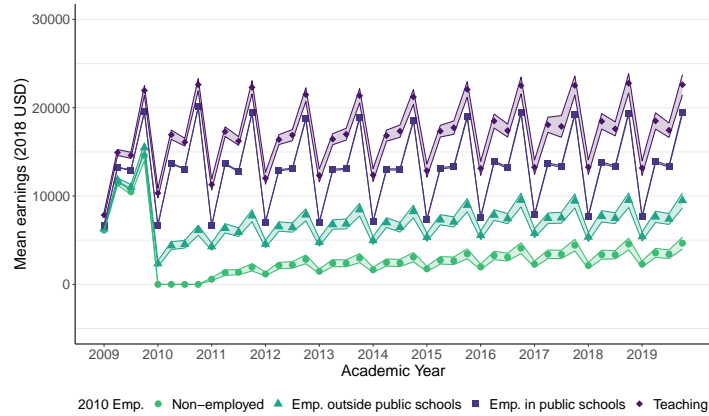
(b) Quarterly earnings.

Figure 7: Patterns in earnings for teachers in 2009 by type of transition made in 2010.

2.7.1   Event study specification

Using the administrative data, we identify the sample of teachers who move out of teaching at some point between 2009 and 2018, whose move cannot be attributed to retirement, and who are employed as a teacher for at least three years prior to the move. We stack their administrative data such that their dates of exit align at $t = 0$, and estimate the following specification via

OLS:

$$Y_{it} = \beta_0 + \sum_{\tau \in \{-2,0,1,2\}} \alpha_\tau \cdot \mathbb{I}(t = \tau) + X_{i,t=-1}\gamma + \varepsilon_{it}, \tag{2}$$

where the coefficients α_τ capture the adjusted earnings difference between year $t = -1$ and year τ, and $X_{i,t=-1}$ are controls describing the teachers' characteristics one year prior to exit.

We also estimate a two-period version of the estimating equation on the sample of individuals who are observed leaving teaching in the matched Current Population Survey (CPS). Studying wage changes in the CPS presents two main advantages. First, doing so provides a useful validation exercise, to check whether the large pay gap we observe in the administrative data is a function of the data rather than reality. Second, the CPS contains data on occupations and hours across industries, which the administrative data do not, allowing us to learn about the types of occupations and part-time/full-time work that teachers leave teaching for.

2.8 Event Study Estimates of the Pay Gap

Figure 7 illustrates the raw data underlying the event study design: specifically, the earnings trajectories for the same sample of individuals who leave teaching in 2010, as well as the earnings trajectories for individuals who stay in teaching for the entire period of 2009 till 2019. The only group that continually earns more than they would have as teachers are those who change to other public school employment; these individuals mostly move to administrative positions which offer higher pay scales than teaching. In contrast, those who move to employment outside of teaching earn on average $21,000 less in their first year away from teaching. Their earnings recover slightly in the following years as they transition back into teaching. Those who leave to non-employment mostly do not return to teaching or other employment and thus consistently earn the least.[19]

The raw means, while illustrative, may reflect pay differences biased by different initial levels of pay across districts or different ages among leaving teachers. We therefore turn to Figure 8, which shows the estimated coefficients from the event study specification in which we control for pre-determined observables at the time of exit including gender, race, age and years of experience teaching at the time of move, and district fixed effects.

Our regression-adjusted estimates suggest that leaving teachers experience a pay decrease of between $18,000 and $25,000 a year upon leaving teaching. We also find suggestive evidence that the earnings drop is not purely driven by selection in worker transitions; instead, the gap persists in cases where the separations are plausibly exogenous to individual-level unobservables.

---

[19]We code non-employment as zero earnings to maintain the same sample throughout.
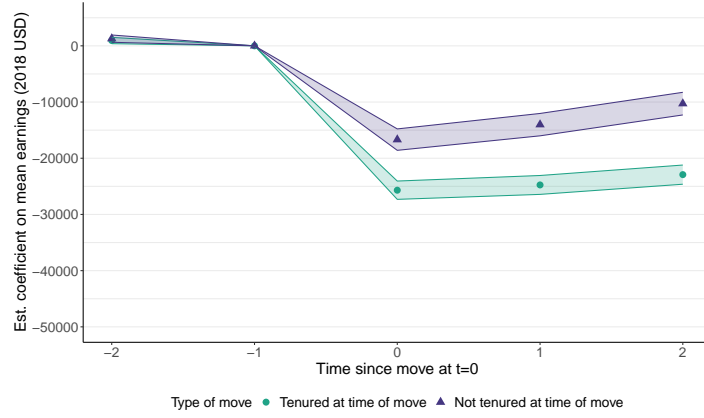
(a) Est. earnings differences relative to $t = -1$, by movers' tenure status
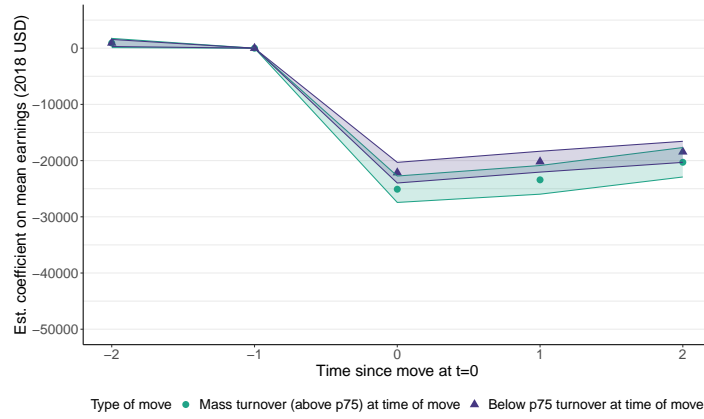
(b) Est. earnings differences relative to $t = -1$, by degree of turnover at school

Figure 8: Event studies for the sample of teachers who exit the public school system and maintain at least 3 quarters of wages in the following years. Outcome is total earnings over the year. Observation counts are the same for every point shown. Non-employment is coded as 0.

Our data are rich enough such that we (1) can distinguish voluntary exits from involuntary ones, and (2) can identify when exits coincide with a more systematic turnover event at a school. First, to investigate whether voluntary separations might be driving the gap, we examine the earnings trajectories separately for teachers who were tenured versus not tenured at exit.[20] We find that voluntary moves made by tenured teachers do coincide with larger earnings drops than plausibly involuntary ones made by untenured teachers. However, even for untenured teachers, the earnings drop is still on the order of $15,000 to $25,000. Secondly, inspired by the mass layoffs literature, we study what happens to earnings for teachers whose move coincides with a "mass turnover" event at their school, which are less likely to be driven by individual-level decisions. We define a "mass turnover" as a school turnover rate that falls above the 75th percentile in a given year. We find similarly large earnings drops for teachers who leave from "mass turnover" schools.[21]

We draw similar conclusions from our analysis of the matched CPS. From the matched CPS, we find that teachers leave to jobs that pay a median of $13,000 less (or an average of $5,000, due to some extreme positive outliers) per year than teaching. The most popular sectors (occupations) one year after exit include education (e.g. teacher assistants, administrators, childcare), healthcare (e.g. nurses, social workers, counselors), administrative (e.g. secretaries, administrative assistants, office clerks), and retail (e.g. sales worker supervisors, cashiers, waiters). The CPS also shows that nearly half of leaving teachers exit to full-time jobs while one-fifth exits to part-time work.

Taken together, we interpret the event study findings as corroborating our RK estimates of a large estimated pay gap. Our event study estimates are very similar in magnitude to our RK estimates, and in light of the occupations analysis from the CPS, are plausible given that part-time or even full-time work as a cashier or secretary may pay much less than teaching does per year. Furthermore, we find suggestive evidence that the gap can persists even ten years after the initial exit.

2.9   Putting the pay gap estimates in context

Using two empirical strategies that leverage different sources of variation in who is a teacher in Kentucky, we find consistent evidence that the pay gap between teaching and teachers' next-best jobs ranges between $18,000 to $21,000 a year, in terms of 2018 US dollars. How do we interpret this range of estimates? Can we think of it as the pay gap for the average teacher?

---

[20]Teachers enjoy a degree of job security after earning "tenure" (working for four years consecutively in the same district), which effectively protects them from being fired, making their moves largely voluntary.

[21]Additionally, we look at moves that occur at the same time as a principal change. Our findings are similar.

And how does the pay gap compare to what teachers actually earn in Kentucky?

We interpret our range of estimates as providing an *upper bound* on the pay gap for the average teacher. The reason is because our RK treatment effects are weighted towards marginal test-passers, amounting to nearly 40% of our sample of test-takers in the optimal bandwidth around the cutoff. Our strategy thus gives no weight to individuals who would have passed the test regardless, including a sizable share of individuals who obtain the highest possible score of 200 on their first attempt. Taking performance on a standardized test as a proxy for human capital, lower-scoring students may also be expected to have lower earnings potential; indeed, we find in the reduced form analysis that the short-term future income is weakly increasing in initial test scores. The test-takers driving our results are therefore those with relatively low earnings potential compared to the average teacher in our sample. Similarly, while we do not observe reasons for teacher exit, one possible reason especially pre-tenure is poor performance. While it remains an open question whether teacher value-added or other measures of teacher effectiveness correlate with general ability, it is possible that the event study estimates are being driven by lower general ability individuals. As such, both designs deliver the pay gap between teaching and the next-best jobs of individuals with relatively low earnings potential, leading us to conclude that the estimated pay gaps serve as an upper bound.

Finally, we use a back-of-the-envelope calculation to put our estimates in context. Table 3 shows the results of dividing our estimates of the pay gap by benchmark values of Kentucky teachers' annual earnings. For the RK estimation, where the main outcome is earnings two years after college, we use as a benchmark the average teaching salary across districts for those with two years of experience and the most common set of credentials[22]: approximately $45,000 in 2018 USD.[23] Likewise, for the event study analysis where the main outcome is earnings one year after leaving teaching, we use as a benchmark the average teaching salary one year prior to exit: around $55,000 in 2018 USD.[24] Table 3, Column (2) shows our preferred calculation[25], which suggests that an upper bound estimate of the pay gap is equivalent to between 33-40% of a Kentucky teacher's annual earnings, depending on the benchmarks used.

---

[22]These credentials can be thought of as courses that can be obtained within the first two years of teaching.

[23]There is essentially no variation in these average salaries between 2018 and 2022 after adjusting for inflation. In addition, the $45,000 salary serves as a lower bound on one's actual annual earnings: taking up additional responsibilities around the school can generate additional pay based on another fixed salary scale, and taking a second job outside of teaching is also common.

[24]Based on the salary schedules, this implies that the average leaving teacher has between 5-10 years of experience.

[25]Our preferred estimate is the effect of actually working as a teacher on earnings, which at the optimal bandwidth falls around $18,000, and is virtually stable over the surrounding bandwidths as seen in Figure 6. For transparency, we also show our calculations our estimate of the effect of becoming certified to teach on earnings, which at the optimal bandwidth falls at around $21,000. However, as Figure 6 shows, this point estimate shrinks and appears to stabilize at lower values as the bandwidth and first-stage F-stat grows.

Table 3: Benchmarking the pay gap estimates

| | | Pay Gap / Salary | |
	Salary (1)	Gap = $18,000 (2)	Gap = $21,000 (3)
Most relevant benchmark for RK sample:			
2yrs experience, Rank I	$45,000	40%	47%
Most relevant benchmark for event study sample:			
Avg earnings before exit	$55,000	33%	38%

Note: The salary schedules used come from the 2019 salary table for teachers, available online from the Kentucky Department of Education.

3 Estimating the Working Conditions Gap

3.1 Identification Challenge

Our goal is to estimate the difference in Kentucky teachers' valuations of the non-wage amenities in teaching and the non-wage amenities in their next-best job options. There are two key identification challenges to doing so. The first is identify and estimate teachers' willingness to pay for amenities in general. The second is identifying the non-wage amenities offered by teachers' next-best jobs. Neither set of quantities can be inferred directly from observational data on job choices, as workers select into jobs for unobserved reasons, including their valuations over amenities. Even longitudinal data that tracks job transitions, with which one could feasibly control for worker- or firm-specific attributes, rarely contains detailed information on the amenities each job provides.

To overcome the first challenge, we administer a set of stated-preference experiments to a representative sample of teachers in Kentucky. In the experiment, we present teachers with pairs of jobs with randomly varied job attributes to choose between, thus allowing us to observe the trade-offs teachers face as well as the choices they make. Using the resulting choice data, we estimate a simple discrete choice model to back out teachers' willingness to pay for each attribute included in the experiment.

To overcome the second challenge, we identify the most popular industries that teachers take their next-best jobs in and then leverage the survey data from Maestas et al. [2023] to identify the amenities offered by the average job in these industries. We triangulate the most popular industries using multiple data sources: the industries of the next-best jobs as found using the RK design and administrative data; the industries of the next-best jobs after exit as found using the event study design in the administrative data and the CPS; and the industries

teachers report in the survey itself, in which we ask directly teachers what their next-best job option would be. We then take the industry-level averages of the amenities reported in the American Working Conditions Survey (AWCS), which provides data on a representative sample of American workers across industries.

The following sections detail our survey design, estimation strategy, and results.

3.2 Data on Working Conditions: Survey of Kentucky Teachers

3.2.1 Survey implementation

To gain an understanding of how working conditions vary between teaching and other jobs, we field an online survey on a representative (once weighted) sample of teachers in the state of Kentucky.

The purpose of our survey is two-fold: to elicit information about the working conditions that teachers experience in Kentucky K-12 schools, and to conduct a stated-preference experiment that generates data with which we can estimate teachers' willingness-to-pay for said working conditions. The survey was coded using the online survey tool Qualtrics, which allows users to send unique links to the survey out to respondents via email. To target Kentucky teachers, we obtained a complete list of school staff and their email addresses through MCH Strategic Data and emailed unique Qualtrics survey links directly to staffs' school email addresses. We conducted two rounds of surveys in May 2024, resulting in 587 surveys of teachers and 48 surveys of other school staff, with a total of 2,745 choices from the choice experiment.[26] The email response rate was 6%, exactly the average email survey response rate according to Qualtrics.[27]

Our approach when designing the survey was to replicate the American Working Conditions Survey (AWCS) as closely as possible. The AWCS was the first survey designed to elicit "detailed information about a broad range of working conditions in the American workplace," and was first fielded to a nationally representative sample of U.S. workers in 2015 by Maestas et al. [2023] through the RAND American Life Panel (ALP). The original survey contained two parts: a short initial questionnaire asking respondents about the job attributes at their current job, and a stated-preference module that presents respondents with ten pairs of hypothetical

---

[26]Each respondent was asked to respond to a minimum of three job choices, with the option to evaluate one or two additional choices. The majority chose to evaluate the maximum of five job choices, leading the total number of choices exceeding three times the number of respondents.

[27]According to Delighted by Qualtrics summary statistics.

jobs and asks respondents which job they prefer in each case.

Our survey asks exactly the same questions as the AWCS does, and we follow the experimental design described in Maestas et al. [2023] exactly in the stated preference module as well. We conduct our own survey rather than use the data from the AWCS because our population of interest is specifically Kentucky teachers. To ensure an apples-to-apples comparison of the results across our research designs, it is important that our quasi-experiment-based results and our survey results are both derived from data on Kentucky teachers. In the AWCS however, occupation is not collected for the majority of respondents, and while a small sample of 46 (out of the total 1,748 respondents) report "teacher" as their occupation,[28] these individuals reside in various places across the country.

There are two key advantages to modeling our survey off of the AWCS: credibility and comparability. First, for a survey on working conditions to be time-efficient but also comprehensive and credible, it is crucial that the survey asks about a small but core set of job attributes that workers deem important. Knowing this, when designing the AWCS, Maestas et al. [2023] conducted a thorough literature review across fields as well as survey pilots to identify job attributes that respondents said were important to them and that varied across jobs. By asking teachers in our survey about the same job attributes used by Maestas et al. [2023] in the AWCS, we can be confident that we are assessing general job attributes of the teaching job that workers would consider salient when choosing between teaching and other jobs. Second, given our survey aims to provide the first set of estimates of teachers' willingness-to-pay for general job attributes, it would be difficult to assess the accuracy of the survey results without a set of benchmark estimates. However, the AWCS provides two natural sets of benchmark estimates: specifically, Maestas et al. [2023] provide willingness-to-pay estimates for women and college-educated workers separately, two groups that are demographically similar to teachers. By using the same survey questions as the AWCS, we therefore obtain natural benchmark estimates with which we can compare our results.

The only place where we diverge from the AWCS is at the end of our survey, in which respondents are asked to bring to mind an alternative, non-teaching job that they would consider doing—in the case of former teachers, they are asked about their current employment—to gather information about the "next-best" occupation and industry they would consider working in.

Table 4 shows summary statistics on the demographic composition of the survey respondents and compares them to the population of public school teachers in KY, which we observe in the public school administrative data. The survey sample closely resembles the population

---

[28]This is a conservative estimate; some of these 46 respondents appear to have only typed in part of their occupation, e.g. "teach" or "teac".

Table 4: Summary statistics on teachers in own survey and KY administrative data

	Survey	Administrative data
	(1)	(2)
White	0.94	0.95
	(0.24)	(0.21)
Woman	0.77	0.78
	(0.42)	(0.41)
Age: 26-29	0.08	0.11
	(0.27)	(0.31)
Age: 30-34	0.10	0.15
	(0.30)	(0.35)
Age: 35-39	0.13	0.16
	(0.33)	(0.37)
Age: 40-44	0.19	0.15
	(0.39)	(0.36)
Age: 45-49	0.15	0.15
	(0.36)	(0.36)
Age: 50 plus	0.30	0.23
	(0.46)	(0.42)
Experience: 3-5 years	0.09	0.14
	(0.29)	(0.35)
Experience: 5-10 years	0.15	0.19
	(0.36)	(0.40)
Experience: 10+ years	0.69	0.57
	(0.46)	(0.50)
Teaches at high school	0.35	0.26
	(0.48)	(0.44)
Teaches at elementary school	0.62	0.48
	(0.48)	(0.50)
Teaches at Title I school	0.71	0.67
	(0.46)	(0.47)
N	598	39385

Note: This table presents means (standard deviations in parentheses) describing teachers who responded to the survey in KY and the full KY teacher population in 2019.

in terms of race (95% white), gender (78% women), and likelihood of working at a Title I school (67-71%). However, older teachers and teachers with more teaching experience are over-represented in the survey. Due to these differences, we construct survey weights for the survey respondents based on the administrative population data and reweight the respondents in all following analyses.

3.3 Working Conditions in Teaching

We begin by examining the incidence of the nine core non-wage job attributes selected by Maestas et al. [2023] as reported by Kentucky teachers.[29] We then compare teachers' reports to those of non-teachers—in particular, college-educated workers and women—in the AWCS. Table 5 presents summary statistics comparing the share of teachers in our survey who report having each job attribute to the share of college-educated workers who did the same in the AWCS. All means are weighted using the corresponding survey weights.

Table 5 shows that teaching is significantly less likely than the average job held by a college-educated worker to offer valued working conditions, even those that are commonplace in other jobs. Among college-educated workers, the majority report being able to set their own schedule (67%), being able to telecommute (54%), and having more than 15 days of paid time off (68%). In contrast, among teachers, only a small share report having some scheduling flexibility (12%), being able to telecommute (4%), or having more than 15 days of paid time off (5%). Teachers are also significantly more likely to have more physically demanding work (2% report mostly sitting, compared to 58% of college-educated workers).

These differences may not be so surprising given the traditional structure of the school day and the school year: classroom teachers are required to be physically present at school five days a week, and while school does not run over the summer, summers are not technically paid time off. However, the teaching job also differs on other less obvious dimensions. Perhaps most starkly, teachers are significantly more likely than all workers—college-educated workers, women workers, and the average American worker—to experience a stressful "fast-pace" at work (87% vs. 64% for college-educated, 67% for women, and 70% for all workers). The teaching job is also less likely to offer training opportunities to develop skills that will transfer to other jobs (53% vs. 74%), less likely to give workers autonomy at work (81% vs. 91%), and more likely to require working alone (41% vs. 33%).

In fact, the only positively-valued amenity that teaching is significantly more likely to offer is the opportunity to make a positive impact on one's community or society (64% vs. 39% for

---

[29]We acknowledge that these nine attributes do not cover every non-monetary aspect of a job, but agree with Maestas et al. [2023] assessment that they do cover a broad spectrum.

jobs of college-educated workers).

Table 5: Incidence of working conditions in teaching and the AWCS

	KY Teachers (1)	AWCS		
		All workers (2)	Women (3)	College educ. (4)
Contractual hours per week	38.8	39.6	37.3	40.2
	(4.5)	(9.9)	(9.6)	(10.2)
Total hours per week	49.9			
	(9.8)			
Hourly wage (in 2015 USD)	28.4	30.3	26.0	38.0
	(11.9)	(36.6)	(28.0)	(30.1)
% with each working condition				
Sets own schedule	10.3	56.5	58.8	67.5
Telecommute	4.1	36.4	37.0	54.9
Moderate physical activity	87.6	38.4	38.1	36.7
Mostly sitting	2.5	42.9	49.4	57.9
Relaxed pace	12.8	29.8	32.6	35.5
Choose how do work	81.1	86.4	86.4	91.8
1-14 paid time off (PTO) days	94.3	26.0	25.1	21.3
15+ PTO days	4.1	59.7	61.2	68.7
Team-based, evaluated on own	55.9	49.1	51.3	52.1
Work by self	41.1	32.4	35.1	33.5
Training opportunities	54.0	70.0	65.4	74.0
Frequent opp. positive impact	61.3	34.5	39.5	39.0
Observations	554	1738	968	908

The data on the small sample of teachers in the AWCS corroborate our survey findings. We identify teachers in the AWCS by merging occupation information from the RAND American Life Panel. Table H.3 shows summary statistics on the differences between the AWCS teachers and other workers.[30] Compared to the average college-educated worker, AWCS teachers are significantly less likely to have scheduling flexibility, the option to telecommute, a relaxed pace at work, or more than 15 days of paid time off. Furthermore, compared to the average working woman, teachers (who are mostly women) earn similar pay but are significantly less likely to have scheduling flexibility or the option to telecommute. As in our data, the only working conditions that teachers in the AWCS are significantly more likely to report having frequent opportunities to make a positive impact with their work. While the advantage of our survey is

---

[30]Statistical tests of these differences are reported in the Appendix in Tables H.4 and H.5.

that we are able to reach a larger sample of teachers who are from the same population covered by our administrative data, it is reassuring that we reach the same qualitative conclusions as we would with the small sample of teachers in the AWCS: that is, that teaching offers lower amenity value than the jobs of observationally similar workers.

Table 6: Comparison of Working Conditions in Teaching to Average Jobs in Other Industries in 2015 (AWCS)

| | Teacher (1) | Industry in AWCS | | | | | | |
		Ed/Health (2)	Prof (3)	Leisure (4)	Govt (5)	Finance (6)	Trade (7)	Info (8)
Next-best industry %	-	36.6	12.5	11.5	6.8	4.1	3.9	3.6
Mean hours per week	39.0	38.3	39.7	31.8	40.0	40.0	40.6	41.9
Mean hourly wage (in 2015)	25.9	30.8	34.1	16.3	32.5	37.0	33.6	29.6
% with each working condition								
Sets own schedule	14.7	56.3	64.8	66.5	55.8	76.0	51.7	70.6
Telecommute	13.4	32.6	52.8	21.9	38.9	58.4	21.1	61.9
Moderate physical demands	88.2	50.1	32.3	54.4	38.4	14.9	39.0	21.7
Mostly sitting	5.8	36.3	50.6	13.2	54.1	81.5	33.2	77.4
Relaxed pace	13.9	33.5	33.2	13.8	30.7	27.1	23.4	25.3
Choose how do work	93.5	90.4	88.0	86.0	93.4	86.7	75.9	96.3
1-14 PTO days	61.2	26.8	30.4	21.9	19.7	17.9	38.8	12.3
15+ PTO days	28.6	61.1	59.0	28.0	75.2	74.5	49.9	68.1
Team-based, evaluate own	68.1	56.3	43.6	58.1	50.7	43.0	45.8	55.8
Work by self	26.0	27.5	38.8	25.0	37.4	37.6	31.0	31.7
Training opportunities	63.3	67.3	73.9	57.9	78.8	70.5	66.8	70.8
Frequent opp. positive impact	61.8	51.9	25.6	31.8	49.0	35.9	26.3	19.2
N	46	385	241	59	152	146	233	52

Finally, we examine the characteristics of the industries that Kentucky teachers reported they would work in for their next-best jobs.[31] The first row of Table 6 gives the distribution of responses from our survey. Over a third report that their next-best job would be in either the education or healthcare sector. The next most popular industries included professional and business services (12.5%), leisure and hospitality (11.5%), government (6.8%), and finance, trade, and information. Our survey responses are in line with the actual job choices that we observe in the RK and event study designs, where we found teachers' next-best jobs are in education, healthcare, administrative and support services (which classifies as under professional

---

[31]We are inquiring about occupations in future iterations of the survey.

and business services), and retail (which classifies as leisure and hospitality or trade).

The columns of Table 6 then compare the job attributes in teaching to the attributes of the average job in each next-best industry. The average job in all other industries is substantially more likely to offer many of the desirable job attributes that teaching does not have, including: scheduling flexibility, telecommuting options, relaxed pace of work, and more than 15 days paid time off. The average job is equally likely to provide autonomy at work, less likely to require moderate physical demands, more likely to involve working alone, and less likely to provide frequent opportunities to make a positive impact.

It is important to note that the average job in all industries, aside from leisure and hospitality, pays more on average than teaching does. One may therefore be concerned that the non-wage attributes described in Table 6 are not representative of the attributes in teachers' true next-best options, given that our pay gap estimates suggest that teachers' next-best options pay substantially less than teaching does per year. To investigate whether this is true, we also examine the attributes of the average job by industry that pays less than the mean wage in the industry, as shown in Table H.6. We find that the distribution of attributes by industry is similar, as are the qualitative comparisons between the average industry jobs and teaching.

3.4 Estimating Teachers' Willingness to Pay for Working Conditions

To estimate teachers' willingness to pay for each of the nine job attributes listed in Table 5, we conducted three stated-preference experiments with each survey respondent. To maximize the size of our data but minimize survey fatigue, which can be especially salient with online surveys, we also gave respondents the option to respond to at most two additional stated-preference experiments, such that the maximum number a respondent could have completed was five. No additional incentives were provided; survey respondents were simply told that completing the additional experiments would "be a huge help to the study." While the majority actually did choose to respond to at least one more stated-preference experiment, our main results use the first three experiments to avoid any selection bias that may be introduced by those who chose to answer more questions.

We detail our experimental design and estimation strategy below, both of which follow Maestas et al. [2023] as closely as possible. Following their example, we also include "trick" questions to address the concern with that survey respondents may only skim the questions and thus provide inaccurate responses. To identify inattentive respondents, we include two attention check questions at the start of the first and second third of the survey, one of which occurs in

between the first and second stated-preference experiment. The first question uses the "reverse wording" technique: it asks a question that was already asked in the survey, but changes the direction of the multiple choice scale. Those who provide inconsistent responses to the true question and the reverse wording question are labelled as "inattentive." The second question is similar to one that was used in the AWCS: the text contains a plausibly genuine question about job preferences but also specifies in the introduction of the text that the respondent should respond in a specific way, regardless of their true answer to the question. Those who do not respond in the specified manner are labelled as "inattentive." Our results are not sensitive to the inclusion or exclusion of the inattentive respondents.

3.4.1 Experimental design

In the first part of the survey, respondents are first asked about the pay, hours, and working conditions in their current/former teaching job, including all nine core job attributes: schedule flexibility, telecommuting opportunities, physical demands, pace of work, autonomy, paid time off, working with others, job training opportunities, and impact on society.[32] To obtain normalized measures of pay, we allow respondents to report their teaching earnings at the hourly, weekly, biweekly, twice-monthly, monthly, or annual level, and ask them to report their hours worked per week and weeks worked per year. We then use these responses to calculate each respondent's hourly wage. These responses give us the summary statistics in Table 5, but also provide the basis for the construction of the hypothetical job profiles presented in the stated-preference experiments.

In the second part of the survey, respondents are presented with a minimum of three stated-preference experiments in which they are asked to select between two jobs which partially vary in terms of job attributes, hours, and wages, exactly following the technical and visual design described in Maestas et al. [2023]. We provide a screenshot of our presentation of a hypothetical job pair in Figure G.5, which can be compared with Appendix Figure 3 in Maestas et al. [2023].

We construct the hypothetical job profiles in two stages. First, we define the respondent's "baseline job" using the pay, hours, and job attributes of the respondent's current/former teaching job that they reported on in the first part of the survey. The baseline job then functions as an initial set of values that we can randomly vary subsets of to construct hypothetical job profiles in the second stage. As described in Maestas et al. [2023], using the baseline job is advantageous because the resulting hypothetical job profiles bear resemblance to a job that the respondent is familiar with. However, to address concerns that respondents may prefer job profiles that they are familiar with, which could bias the willingness-to-pay estimates, we also

---

[32]See Table XX for a list of the job attributes and their potential values.

define a "common" baseline job—essentially the average teaching job we expected based on the survey pilot—and use it in one of the three experiments presented. The values of the common baseline job are shown in Appendix XX. We find that the inclusion of the common baseline jobs does not affect our willingness-to-pay estimates.

Next, we create hypothetical job choices Job A and Job B by randomly selecting (without replacement) two of the nine non-wage job attributes to vary across the two hypothetical jobs. We also vary the offered wage and hours per week between the two jobs, such that in total, each job is characterized by eleven characteristics. Hours are randomly varied to be one of 5-hour intervals between 15 and 60 hours per week. Wages are generated to be similar on average to the hourly wage that the respondent reported earning. Specifically, following Maestas et al. [2023], given the respondent's actual hourly wage w_i obtained in the first part of the survey, the hypothetical wages for Job A and Job A are given by $\theta_A w_i$, where $\theta_A \sim N(1, 0.1^2)$ and $\theta_B w_i$, where $\theta_B \sim N(1, 0.1^2)$ respectively. To ensure that the wage values do not stray too far from the respondent's reported wage, we truncate θ_A and θ_B at 0.75 and 1.25. To facilitate ease of interpretation for the respondent, we display the earnings in the hypothetical job choice in terms of the units in which the respondent reported their earnings (hourly, weekly, biweekly, monthly, or annually), and round to the nearest \$.50 if hourly, \$10 if weekly, biweekly, or monthly, and \$100 if annually. To convert hourly wages to annual earnings, we assumed the job constituted 52 weeks of work. If the weekly work hours exceeded 40, we called the excess time overtime pay and assigned it a wage of 1.5 times the randomly assigned hourly wage. If presenting hourly wages, we also presented the implied weekly earnings below, including overtime pay.

Finally, especially given that we only asked three experiments per person compared to ten per person in Maestas et al. [2023], we took steps to maximize precision by limiting situations where one job dominated the other on all dimensions. We used the willingness-to-pay estimates from the college-educated workers in Maestas et al. [2023] as a benchmark with which to rank job attribute values, but not the attributes themselves. When one of the randomly generated jobs dominated the other on all aspects, we redrew the scaling parameters θ_A and θ_B and recomputed the wage offered. If one job still dominated, we redrew the attribute values. These steps reduced the likelihood that one job would dominate the other in all aspects.

As a result, each pair consisted of a Job A and a Job B that were identical in every way except for two non-wage attributes, the hourly wage, and the hours worked per week. We displayed a table comparing the attributes of Job A (in the first column) and Job B (in the second column) side-by-side, and asked respondents to choose whether they "Strongly Prefer Job A," "Prefer Job A," "Prefer Job B," or "Strongly Prefer Job B." Our presentation of the job pairs as shown in Figure G.5 is virtually identical to the presentation Maestas et al. [2023]

use, shown in their Appendix Figure 3. The above process was repeated a minimum of three times per respondent to produce three experiments; after the third, respondents are given the option to evaluate one to two additional job pairs, if willing.

3.4.2 Estimation

Because we randomize the attributes across the hypothetical job choices, we are able to use the choice data from the stated preference experiments to estimate Kentucky teachers' willingness-to-pay for each job attribute. Following Maestas et al. [2023], we assume that the respondents have the following indirect utility function:

$$V_{ij} = \alpha + A'_{ij}\beta_i + \delta_i \ln w_{ij} + \varepsilon_{ij}, \tag{3}$$

where A_{ij} is the vector of job attributes offered by job j and w_{ij} is the corresponding wage. Then under the assumption that ε_{ij} is i.i.d. EVI, a standard assumption in discrete choice models and in the willingness-to-pay literature, the likelihood that worker i chooses teaching over job k is given by the expression

$$\Pr(V_{it} > V_{ik}) = \frac{\exp[(A'_{it} - A'_{ik})\beta_i + \delta_i(\ln w_{it} - \ln w_{ik})]}{1 + \exp[(A'_{it} - A'_{ik})\beta_i + \delta_i(\ln w_{it} - \ln w_{ik})]}. \tag{4}$$

In theory the β_i and δ_i subscripts allow for heterogeneity in preferences over amenities and pay. However, unlike Maestas et al. [2023], we are focusing on a specific, less variable group of individuals: Kentucky teachers. We therefore assume that $\beta_i = \beta$ and $\delta_i = \delta$ and use a standard logit model to estimate β and δ. We aggregate the four possible responses into a single dummy variable that indicates whether the workers prefer Job A or not, and use survey weights in estimation. We also conduct as many of the robustness checks that Maestas et al. [2023] explore that our sample size allows, including: (i) excluding "inattentive" respondents, i.e. those who incorrectly answered both trick questions, (ii) estimating a probit model instead of a logit model to evaluate sensitivity in the error term distributional assumption, (iii) excluding experiments that used the "common" baseline job instead of the respondent's own baseline job, (iv) estimating the model without survey weights. Despite our smaller sample, we find that our willingness-to-pay estimates are largely robust to all of these variations.

Finally, we transform the estimated parameters from the discrete choice model into a measure of the teacher's willingness-to-pay for each desirable job attribute r. If teacher i is willing to pay WTP_i^r for an attribute r, then they should be indifferent between a job that pays

37

w_i but does not offer r and a job that pays $w_i - \text{WTP}_i^r$ but offers r:

$$\delta \ln w_i = \beta^r + \delta \ln[w_i - \text{WTP}_i^r], \tag{5}$$

where β^r denotes the average teacher's marginal utility from attribute r and δ is the average teacher's marginal utility of the log wage. Rearranging gives our object of interest:

$$\text{WTP}_i^r = w_i \left[1 - \exp\left(\frac{-\beta^r}{\delta} \right) \right]. \tag{6}$$

To report our results, we follow Maestas et al. [2023] in presenting our estimates in terms of $1 - \exp\left(\frac{-\beta^r}{\delta} \right)$, such that gaining attribute r in one's job is equivalent to a $100 \left[1 - \exp\left(\frac{-\beta^r}{\delta} \right) \right] \%$ wage increase.

We also use our estimates to construct summary measures that capture the average Kentucky teacher's full valuation of the desirable amenities that teaching does not offer. More specifically, we summarize (1) the average teacher's total valuation of all the desirable amenities teaching does not offer, and (2) the total amenity value the average teacher would get from switching from teaching to the average teacher's next-best job. We define the job attributes of the average teacher's next-best job in two steps. First, we take the most popular industries that teachers' next-best jobs were found in based on the RK analysis, the event study analysis, and as reported in the survey. Then, using the AWCS as our reference dataset, we group the responses in the AWCS by industry and summarize the incidence of job attributes within each industry. We take the average set of attributes reported in each industry as proxies for the attributes in the average teacher's next-best job.

Our measures are based off of the summary measure used by Maestas et al. [2023] to capture respondents' full valuation in terms of amenities of the "best" job (i.e. one with all the desirable attributes) relative to the "worst" job (i.e. one with none of the desirable attributes). We begin by restating this measure using slightly different notation which will allow us to refer to the most preferred as well as the second most preferred value of each attribute. Let S denote the set of all attributes. Let $\beta^{s,1}$ denote the willingness to pay for the highest attribute value of s, and in cases where there are more than two values for attribute s, let $\beta^{s,2}$ denote the willingness to pay for the second highest attribute value of s.[33] For example, for workers in the AWCS and taking s to be the physical demands attribute, $\beta^{s,1}$ is the valuation of moderate physical demands (the most preferred value) while $\beta^{s,2}$ is the valuation for mostly sitting.

Using our notation, we can write the Maestas et al. [2023] definition of respondent i's

---

[33]If the attribute has only two values, $\beta^{s,2}$ is 0.

willingness to pay for the "best" job relative to the "worst" job as

$$\text{WTP}_i^{\text{FULL}} = w_i \left[1 - \exp\left(\frac{-\sum_{s=1}^{S} \beta^{s,1}}{\delta} \right) \right], \tag{7}$$

meaning we add up the coefficients for the most preferred value of each attribute. Note that in this baseline measure, we are careful not to "double count" attribute values, which would inflate our summary measures. If an attribute has more than two values, we only use the coefficient associated with the attribute value with the highest willingness-to-pay estimate.

For our first measure, we adapt (7) to capture respondents' valuation over only the desirable amenities teaching typically does not offer. We leverage our new notation to account for differences between a job that offers all the desirable amenities versus that a job that offers *some* desirable amenities rather than none at all. Our first measure is given by

$$\text{WTP}_i^{\text{BEST}} = w_i \left[1 - \exp\left(\frac{-\sum_{s=1}^{S} (\beta^{s,1} - \beta^{s,2} \cdot \mathbb{I}_{\text{teach}}^{s,2})(1 - \mathbb{I}_{\text{teach}}^{s,1})}{\delta} \right) \right], \tag{8}$$

where $\mathbb{I}_{\text{teach}}^{s,j}$ is an indicator variable for whether the average Kentucky teaching job offers the jth highest attribute value of s or not, meaning we add up the coefficients only for desirable attributes that teaching does *not* offer. Overall, the measure can be thought of as the total fraction of one's salary that the average teacher would be willing to pay to switch from teaching to the "best" job, i.e. one that offers all the most desirable attributes. The expression also makes clear why it is necessary to include multiple attribute values, not only the highest valued ones, when valuing a job relative to teaching. Take for example the paid time off attribute, which has three values—no time, 10 days, and 20 days—that workers value in increasing order. If teaching offers 10 days paid time off (meaning $\mathbb{I}_{\text{teach}}^{s,2} = 1$), then when evaluating how much teachers would value having 10 additional days paid time off, we would want to include teachers' valuations of 20 days paid time off *netting out* their valuation of the 10 days paid time off that they already get as teachers, which is given by $(\beta^{s,1} - \beta^{s,2} \cdot \mathbb{I}_{\text{teach}}^{s,2})$.

For our second measure, we again adapt (7) but this time to capture differences between teaching and teachers' next-best jobs. We define teacher i's willingness to pay for the average teachers' next-best job over teaching as

$$\text{WTP}_i^{\text{NXBEST}} = w_i \left[1 - \exp\left(\frac{-\sum_{s=1}^{S} (\beta^{s,1} - \beta^{s,2} \cdot \mathbb{I}_{\text{teach}}^{s,2})(\mathbb{I}_{\text{nxbest}}^{s,1} - \mathbb{I}_{\text{teach}}^{s,1})}{\delta} \right) \right], \tag{9}$$

where $\mathbb{I}_{\text{nxbest}}^{s,1}$ is an indicator variable for whether the average next-best job has the highest value

of attribute s or not, meaning we add up the coefficients for the most preferred values of the attributes that the next-best job offers and net out the values of the attributes that teaching already offers.[34] The measure can be thought of the total fraction of one's salary that the average teacher would be willing to pay (in terms of amenities alone) to switch from teaching to their next-best job—or, in other words, the compensating differential between teaching and teachers' next-best jobs.

For all three summary measures, we compute standard errors using the delta method and cluster by respondent.

Our willingness-to-pay parameters are identified so long as the respondents' choices were made under the belief that the jobs in each pair were truly the same with the exception of the randomly varied characteristics highlighted in the table. As such, we implicitly assume that respondents followed the survey instruction to assume that the pair of jobs shown are identical in every way, including in ways that are not explicitly specified in the table. We also assume that there were no systematic differences in how respondents rated the subjective amenities like heavy versus moderate physical activity, which we think is reasonable given that our survey respondents are somewhat homogeneous (e.g. mostly women) and engage in similar work.

3.5 Main Estimates of Teachers' Willingness to Pay for Working Conditions

Our main estimates of Kentucky teachers' willingness-to-pay for each job attribute are presented in Table 7, alongside columns showing the willingness to pay estimates from Maestas et al. [2023] for all workers and for college-educated workers. As described in the previous section, the estimates have been transformed such that each number can be interpreted as the percentage wage increase (or decrease) that a teacher would need to compensate for removing (or adding) a given job attribute. For example, the first estimate in Column (1) indicates that the average Kentucky teacher is willing to pay 8 percent of their salary to be able to set their own schedule, relative to a job in which they cannot set their own schedule. Although our sample size is significantly smaller than that of the AWCS, because our population of study is specific and relatively homogeneous, our estimates are precise.

We highlight three key findings. First, we find that Kentucky teachers are willing to pay large amounts for better working conditions. The most valued job attribute is paid time off:

---

[34]In other words, the sum in the expression puts positive weight on β^r if r is offered by the next-best job but not by teaching, no weight on β^r if r is offered in both or neither of the jobs, and negative weight on β^r if r is offered by teaching but not by the next-best job.

switching from a job with no paid time off to one with 20 days of paid time off is equivalent to a 26.8 percent wage increase. A job requiring moderate physical activity as opposed to heavy physical activity is valued at 15.9 percent more of one's wage. Working in a team but being independently evaluated is equivalent to a 9.9 percent wage increase over working in a team but being evaluated as a team. Switching from a fast-paced to a relaxed-pace job is equivalent to a 9.3 percent wage increase.

Second, we find that teachers' preferences are very similar to those of the average American worker and those of the average college-educated worker. The only job attribute that teachers value statistically significantly more than other workers is having a relaxed pace at work: the average college-educated worker in the US values switching from a fast pace to a relaxed pace as a 5.2 percent wage increase, while the average Kentucky teacher values the same switch nearly twice as much, as a 9.3 percent wage increase. One explanation for this difference is prevalence of teacher "burnout:" surveys show that stress and burnout is common among especially starting teachers and contributes to high turnover rates in teaching, especially among starting teachers CITE. Workers who are burned out at the time of taking the survey may value more highly a relaxed pace of work.

Some of the insignificant differences in preferences are noteworthy as well. Perhaps most notably, teachers and other workers share the same willingness to pay for having frequent opportunities to impact the community or society with one's work. This finding may be surprising given that, as shown in Table 5, teaching is significantly more likely than other jobs to offer such opportunities. These estimates suggest that teachers are not strongly selected on community-impact-related altruism. We also find that, although teachers are significantly more likely to work alone than the average American worker, teachers marginally prefer working in teams and being evaluated alone over working alone.

Third, we find that because the average teaching job does not offer most of the nine core job attributes we study, teachers are willing to pay a substantial amount of their salary to switch to a job with more desirable amenities. To establish a benchmark, we begin by estimating the "full" valuation of the amenities using the measure from Maestas et al. [2023]. We estimate that for teachers in our sample, switching from the least desirable job to the most desirable job, in terms of amenities, is equivalent to a 62.2 percent wage increase. Reassuringly, our estimate is very similar to the estimates in Maestas et al. [2023] of the value of switching from the "worst" job to the "best" job for the average American worker (a 55 percent wage increase) or for the average college-educated worker (a 60 percent wage increase).

We then assess the wage impact of two more plausible job changes for Kentucky teachers: switching from teaching to the "best" job, and switching from teaching to teachers' average

Table 7: Willingness-to-pay estimates for job attributes

	Teachers (1)	AWCS All workers (2)	College-educ. (3)
Sets own schedule	0.080	0.089	0.104
vs. Schedule set by manager	(0.018)	(0.007)	(0.008)
Telecommute	0.063	0.042	0.069
vs. No telecommuting	(0.015)	(0.007)	(0.008)
Moderate physical activity	0.159	0.145	0.168
vs. Heavy physical activity	(0.021)	(0.010)	(0.013)
Sitting	0.119	0.116	0.137
vs. Heavy physical activity	(0.026)	(0.010)	(0.013)
Relaxed	0.093	0.043	0.052
vs. Fast pace	(0.017)	(0.007)	(0.007)
Choose how do work	0.050	0.040	0.059
vs. Tasks well defined	(0.017)	(0.007)	(0.007)
10 days PTO	0.196	0.164	0.158
vs. No days PTO	(0.024)	(0.009)	(0.010)
20 days PTO	0.268	0.230	0.226
vs. No days PTO	(0.025)	(0.010)	(0.011)
Team-based, evaluate own	0.099	0.065	0.074
vs. Team-based, evaluate team	(0.021)	(0.010)	(0.011)
Work by self	0.048	0.086	0.073
vs. Team-based, evaluate team	(0.022)	(0.010)	(0.012)
Training opportunities	0.041	0.054	0.064
vs. Already have skills	(0.016)	(0.007)	(0.008)
Frequent opp. to serve	0.043	0.036	0.039
vs. Occasional opp. to serve	(0.015)	(0.007)	(0.008)
Best job	0.622	0.550	0.600
vs. Worst job	(0.033)	(0.016)	(0.016)
Best job	0.347		
vs. Teaching	(0.037)		
Next-best job	0.286		
vs. Teaching	(0.037)		
N	1,662	17,380	9,080

next-best job. Recall that we obtain proxies for the attributes offered by teachers' next-best jobs using a combination of our previous findings from the administrative data and the survey itself. For the first type of job change, we estimate that switching from teaching to a job that offers all the most desirable attributes, in terms of amenities, is equivalent to a 36 percent wage increase. For the second type of job change, we estimate that switching from teaching to the average teachers' next-best job, in terms of amenities, is equivalent to a 29 percent wage increase.

4 Discussion

Using independent empirical designs and datasets, we find that in Kentucky, teaching offers a large pay premium relative to teachers' next-best options, which we upper bound at being between 33-40% of the average teachers' annual earnings. We also find that, through a survey of Kentucky teachers and a set of stated-preference experiments, that Kentucky teachers are willing to pay 35% of their salary for a job with desirable attributes that teaching does not offer, and 29% of their salary for their next-best job.

Combining our most conservative estimates using the following expression,

$$\text{Rent} \approx \underbrace{(\text{Pay}_T - \text{Pay}_{NB})}_{\textbf{Pay gap}} + \underbrace{(\text{WorkConditions}_T - \text{WorkConditions}_{NB})}_{\textbf{Working conditions gap}}.$$

we find that teachers in Kentucky earn a rent of around 4% of their salary (minimum -2%, maximum 11%) and a compensating differential of around 29% of their salary (minimum 29%, maximum 35%). With our range of estimates, we cannot reject that teachers may earn no rent at all.

Our results therefore suggest although teaching pays more than teachers' next-best jobs, most of the pay premium is functioning as a compensating differential rather than a rent. In other words, the overall compensation in teaching is actually relatively similar to teachers' next-best options once both pay and working conditions are taken into account.

Assessing magnitudes. Our estimates of the gap in the value of working conditions are large. As such, one may be concerned that we are overestimating the gap because our survey does not cover all the non-wage aspects of teaching: in particular, the survey (as of October 2024) does not include desirable amenities common to teaching like "summers off" or shorter work days.

We present two justifications for why our estimate of the working conditions gap is plausible

and specifically not an overestimate. First, we compare the pay and working conditions between public and private schools. Private school teaching jobs are the most similar to public school teaching jobs in terms of the structure of the work day and the skills involved. However, private schools are also able to select students to enroll, leading to smaller class sizes and fewer behavioral issues [The Atlantic, 2013]. Incidentally, in 2020-21, the average private school teacher in the US also earned around \$19,000 *less* than the average public school teacher in 2018 USD [National Center for Education Statistics, 2024]—a remarkably similar number to the gap in the value of working conditions that we estimate. While only suggestive, the private school/public school pay gap is consistent in terms of both direction and magnitude with our finding that public school teachers may demand a large compensating differential to make up for poor working conditions driven by student behavior and stress.

Second, although our survey (as of October 2024) does not include desirable teaching amenities like "summers off," it also does not include undesirable amenities common to teaching like student behavioral issues or exposure to violence in schools. It is therefore not obvious that our current estimate of the value of teaching conditions is biased upward.[35]

Our estimates of the pay gap are also large. Similarly however, we argue that the magnitude is plausible, especially under the interpretation of the point estimates as an upper bound (as discussed in Section 2.9). The only remaining concern is that the estimates may be biased due to our earnings data being from a state UI system, as it is well known that UI systems do not cover moves between states. For example, if the lower-paying opportunities for exiting teachers are local but the higher-paying opportunities for exiting teachers involve moving across state borders, then our estimates of the pay gap could be biased upward. However, our estimates are in line with previous studies that have used event study designs to examine whether teachers leave teaching to higher or lower pay in multi-state data, including the NLSY and the LEHD. These papers find similarly sized declines in pay following exits from teaching.

Qualitative evidence.   Our conclusion is internally consistent with Kentucky teachers' views expressed in the open-ended portion of the survey (Figure 9): when asked if there was anything more the teachers wanted to share about their experience choosing between teaching and other jobs, the most commonly raised issue was [student] behavior rather than pay. The conclusion is also consistent with reports of teacher shortages and increasingly challenging working conditions in teaching, especially since COVID-19.

---

[35]An updated version of the survey that includes additional teaching conditions will be included in this paper in early 2025.

Figure 9: Word cloud of most common words used in response to the final (open-ended) question of the survey, aside from generic words including "school," "kids," "students." Larger font size indicates greater frequency of mention. The question prompt was: "Is there anything more that you would like us to know about your experience choosing between being a teacher versus pursuing a different job or career path?"

Comparison to prior work. Our conclusions may initially seem at odds with prior work that documents rent extraction in teaching. However, once we take our context and approach into account, our findings do not contradict but rather add additional context to prior work. Many of the papers on rent extraction in the public sector focus on the role of unions in driving up such rents. However, while teacher unions are known to be politically strong and active when it comes to collective bargaining, teacher unions are not as prevalent in the U.S. South: a 2011 report from Kentucky showed that less than 30% of teachers at the time were covered by a union that actively in collective bargaining (CITE LRC). Our findings may therefore be more generalizable to other states with similar union involvement than to the average U.S. state.

We are also not aware of other studies that incorporate the value of non-wage amenities when evaluating whether workers extract rents. In light of growing evidence that workers may have high valuations over certain amenities, it is important to take non-wage attributes of jobs into account when assessing whether the overall compensation jobs offer, which prior measures of rent extraction may not. For example, the most common metric with which previous studies have evaluated rent extraction in teaching is total funding inputs for staff expenses. However, most of the job attributes that we consider in our survey do not necessarily entail an explicit cost (e.g. working in a team versus working alone), and thus the presence or lack of such attributes would not be included in such metrics.

Finally, this paper provides new perspective on a few well-documented facts about teachers. The first fact is that teachers have become increasingly negatively selected on academic performance (e.g. Bacolod [2007], Corcoran et al. [2004]). Prior work provides several explanations for the downward trend, including wage compression [Hoxby and Leigh, 2004] and increasing labor market opportunities for women. My findings suggest a third explanation: teachers may

be becoming more negatively selected because the working conditions in teaching are declining in quality. In the face of such conditions, the workers who select into teaching are those with relatively low-paying next-best options who are willing to accept the pay premium teaching offers as a compensating differential. While the theory that job attributes may affect selection into teaching has received considerable attention (CITE HANUSHEK RIVKIN), this paper is one of the first to provide empirical evidence that supports this theory.

The second fact is that schools hold significant monopsony power over teachers (e.g. Ransom and Sims [2010]). Numerous studies document that within the public school system, school districts hold monopsony power over teachers: for example, Ransom and Sims [2010] estimate that school districts in Missouri have the power to reduce wages by 27% compared to a frictionless world in which workers are informed and mobile. The authors posit that school districts' market power can be attributed to teachers' locational preferences [Boyd et al., 2013] and the nature of teacher pensions—specifically, the fact that switching districts after one has begun accruing in the schools' pension plan can induce a significant financial loss. However, less attention has been given to the idea that schools may hold market power in the broader labor market as well due to the fact that teaching offers uncommon yet desirable amenities, like pensions, tenure, and "summers off." Based on my findings, one may therefore posit that schools have the power to reduce wages *and* sustain relatively poor working conditions due to their market power across districts and in the broader labor market, derived from workers' preferences, institutional features, and unique amenity provision.

The third fact is that teachers in the US earn less than their similarly-educated counterparts. Numerous high-profile policy reports document the observational pay gap between teachers and other college graduates using the Current Population Survey (CPS) (e.g. Allegretto and Mishel [2018]) and have used these estimates to argue that teachers are undercompensated relative to their alternative job options. However, the results of other studies bring into question whether the observational pay gap is an unbiased estimate of the true pay gap between teaching and teachers' next-best options. Using event study designs to examine how pay changes for exiting teachers in the NLSY [Stinebrickner, 2002], GA [Scafidi et al., 2006], and WA [Goldhaber et al., 2022], these studies consistently find that exiting teachers tend to leave to lower-paying jobs. My findings show that these two seemingly opposed facts are not at odds with one another. Using quasi-experimental variation in who becomes a teacher to address the selection issues that arise when using observational data like the CPS, I find that teachers are negatively selected on income, consistent with the observational pay gap. At the same time, I find that consistent with the event study results, teaching is a high-paying option relative to teachers' next-best options *because* teaching offers a compensating differential for

relatively poor working conditions.

One limitation of our study is that our data and samples come primarily from one state, which limits the generalizability of our findings. Given that the matched CPS evidence resulted in a smaller, but still negative, pay gap, there is likely geographic heterogeneity in the differences in pay and working conditions between teaching and teachers' next-best options. That said, the small number of existing studies that examine whether workers who exit teaching leave to higher-or lower-paying jobs consistently find that exiting teachers leave to lower-paying jobs, suggesting that Kentucky is not unique in the large pay premium we find for teachers.

Another limitation of our study is that because our survey targets general, non-teaching-specific amenities, we leave untouched a "middle ground" of working conditions and amenities that are common in teaching and exist but are uncommon in other jobs. The open-ended responses to our survey suggest that these amenities/disamenities include: workplace safety risks (poor/violent student behavior), external communication/stakeholder interaction (parent involvement), public respect, job stability (tenure), and extended unpaid leave ("summers off"). Because these job attributes are a mix of desirable and undesirable traits, it is unclear how including workers' valuations of these attributes would affect our estimates. We aim to quantify these valuations in future iterations of the survey.

5   Conclusion

This paper revisits the long-standing belief that teachers, as public sector workers who are largely unionized, are able to extract rents—that is, are able to earn greater overall compensation as a teacher than they would in their alternative job. Motivated by the observation that this belief appears to be at odds with recent reports of teacher shortages and challenging working conditions in teaching, I investigate how the *overall* compensation in teaching today compares to that of teachers' next-best jobs in the state of Kentucky.

I evaluate the difference in overall compensation between teaching and teachers' next-best jobs in three steps: first by assessing the pay gap, then by assessing the gap in amenity values, and finally by combining the two gaps. I use two quasi-experimental strategies to estimate the pay gap. I develop an empirical strategy to identify the treatment effects of becoming a teacher in a dynamic RD setting, leveraging the fact that test-takers can retake the exam if they do not pass on their first attempt. I also use an event study design to estimate the effect of leaving teaching on earnings. I estimate both models using novel linked data that allow me to track the short-term earnings, education, and employment outcomes for individuals who express an interest in pursuing teaching by taking the teaching entry exam in Kentucky. The data and

strategies allow me to provide the first quasi-experimental estimates of how much teachers' next-best options pay in a U.S. setting.

To estimate the gap in amenity values, I field a survey to current and former teachers in Kentucky to elicit information about their working conditions that includes a stated-preference experiment. My survey instrument is identical on key dimensions to the instrument used in the AWCS by Maestas et al. [2023], allowing me to compare my estimates to those of the representative sample of U.S. workers in the AWCS. The survey is the first to systematically document the differences in working conditions between teaching and teachers' next-best options.

From the RK, I find that teaching offers a pay premium: that is, interested teachers have substantially lower-paying next-best options than teaching in the short-run. From the survey, I find that teaching offers poorer working conditions than teachers' next-best jobs. Remarkably, the estimates of the pay gap between teaching and teachers' next-best options from the RK and the estimates of teachers' willingness-to-pay for a job with better working conditions from the survey are very similar, at around 33-40% of a Kentucky teacher's salary. Taken together, the similarity in estimates across the two independent designs and datasets suggests that the overall compensation offered by teaching and teachers' next-best jobs is similar. The large pay gap that exists is thus primarily a compensating differential rather than a rent.

The idea that teaching offers a large pay premium and that disamenities may be accounting for a sizeable share of the labor cost of teaching suggests that there could be large benefits in evaluating the costs and benefits of improving workplace amenities for teachers. Which working conditions do teachers value most highly that teaching does not offer but can be improved at low cost? And what is the disamenity value associated with current policy proposals that aim to improve student outcomes but worsen teachers' working conditions, like standardized testing or banned disciplinary policies? Our findings suggest that investigating these questions may be a critical step in addressing the hiring challenges that schools are facing today.

References

S. Allegretto and L. Mishel. The teacher pay penalty has hit a new high: Trends in the teacher wage and compensation gaps through 2017. *Economic Policy Institute*, 2018.

J. D. Angrist and V. Lavy. Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly journal of economics*, 114(2):533–575, 1999.

M. P. Bacolod. Do alternative opportunities matter? the role of female labor markets in the decline of teacher quality. *The Review of Economics and Statistics*, 89(4):737–751, 2007.

D. Boyd, H. Lankford, S. Loeb, and J. Wyckoff. Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics*, 31(1):83–117, 2013.

S. Calonico, M. D. Cattaneo, and M. H. Farrell. Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2): 192–210, 2020.

D. Card, D. S. Lee, Z. Pei, and A. Weber. Inference on causal effects in a generalized regression kink design. *Econometrica*, 83(6):2453–2483, 2015.

S. Champion, A. Mastri, and K. Shaw. The teachers who leave: Pulled by opportunity or pushed by accountability?, 2011.

J. J. Conlon and D. Patel. What jobs come to mind? stereotypes about fields of study. Technical report, Working Paper, 2022.

S. P. Corcoran, W. N. Evans, and R. M. Schwab. Changing labor-market opportunities for women and the quality of teachers, 1957–2000. *American Economic Review*, 94(2):230–235, 2004.

J. M. Cowen and K. O. Strunk. The impact of teachers' unions on educational outcomes: What we know and what we need to learn. *Economics of Education Review*, 48:208–223, 2015.

Y. Dong. Jump or kink? regression probability jump and kink design for treatment effect evaluation. *Unpublished manuscript*, 2018.

D. Goldhaber, J. Krieg, S. Liddle, and R. Theobald. Out of the gate, but not necessarily teaching: A descriptive portrait of early-career earnings for those who are credentialed to teach. *Education Finance and Policy*, pages 1–40, 2022.

C. M. Hoxby. How teachers' unions affect education production. *The Quarterly Journal of Economics*, 111(3):671–718, 1996.

C. M. Hoxby and A. Leigh. Pulled away or pushed out? explaining the decline of teacher aptitude in the united states. *American Economic Review*, 94(2):236–240, 2004.

C. Jepsen, P. Mueser, and K. Troske. Labor market returns to the ged using regression discontinuity analysis. *Journal of Political Economy*, 124(3):621–649, 2016.

A. C. Johnston. Preferences, selection, and the structure of teacher pay. *Available at SSRN 3532779*, 2020.

D. S. Lee, J. McCrary, M. J. Moreira, and J. Porter. Valid t-ratio inference for iv. *American Economic Review*, 112(10):3260–3290, 2022.

M. F. Lovenheim. The effect of teachers' unions on education production: Evidence from union election certifications in three midwestern states. *Journal of Labor Economics*, 27(4):525–587, 2009.

M. A. Lyon, M. A. Kraft, and M. P. Steinberg. The causes and consequences of us teacher strikes. Technical report, National Bureau of Economic Research, 2024.

N. Maestas, K. J. Mullen, D. Powell, T. Von Wachter, and J. B. Wenger. The value of working conditions in the united states and implications for the structure of wages. *American Economic Review*, 113(7):2007–2047, 2023.

A. Mas and A. Pallais. Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–3759, 2017.

J. McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2):698–714, 2008.

M. Nagler, M. Piopiunik, and M. R. West. Weak markets, strong teachers: Recession at career start and teacher effectiveness. *Journal of Labor Economics*, 38(2):453–500, 2020.

National Center for Education Statistics. Public school expenditures. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences, 2024. Retrieved Oct 16 2024, from https://nces.ed.gov/programs/coe/indicator/cmb.

M. R. Ransom and D. P. Sims. Estimating the firm's labor supply curve in a "new monopsony" framework: Schoolteachers in missouri. *Journal of Labor Economics*, 28(2):331–355, 2010.

B. Scafidi, D. L. Sjoquist, and T. R. Stinebrickner. Do teachers really leave for higher paying jobs in alternative occupations? *Advances in Economic Analysis & Policy*, 6(1), 2006.

T. R. Stinebrickner. An analysis of occupational change and departure from the labor force: Evidence of the reasons that teachers leave. *Journal of Human Resources*, pages 192–216, 2002.

The Atlantic. Why are private school teachers paid less than public school teachers? 2013. URL https://www.theatlantic.com/education/archive/2013/10/why-are-private-school-teachers-paid-less-than-public-school-teachers/280829/. Accessed: [Date of Access].

The New York Times. As more teachers quit, schools scramble to fill vacancies. March 2023. URL https://www.nytimes.com/2023/03/13/education/teachers-quitting-burnout.html.

Appendices

A   Conceptual Framework

In this section, we provide a conceptual framework that defines the relationship between rents and compensating differentials and motivates our empirical approach to evaluating the extent to which teachers earn rents.

Consider the aggregate demand and supply curves for teaching labor depicted in Figure 10. All the subfigures share the same values marked on the y-axis: the going teacher salary, w_T^*, and the going average pay at teachers' non-teaching, next-best job, w_{NT}^*. There is a teaching premium, i.e. $w_T^* > w_{NT}^*$. The demand curve is shown for a fixed budget for teacher salaries: the lower the teaching salary, the more teachers employed. We assume that the budget constraint is binding, which is reasonable given that smaller class sizes are both beneficial to students and preferred by teachers [Angrist and Lavy, 1999, Johnston, 2020].

(a) Pay gap = Comp. diff. only

(b) Pay gap = Rent only

(c) Pay gap = Comp. diff. + rent
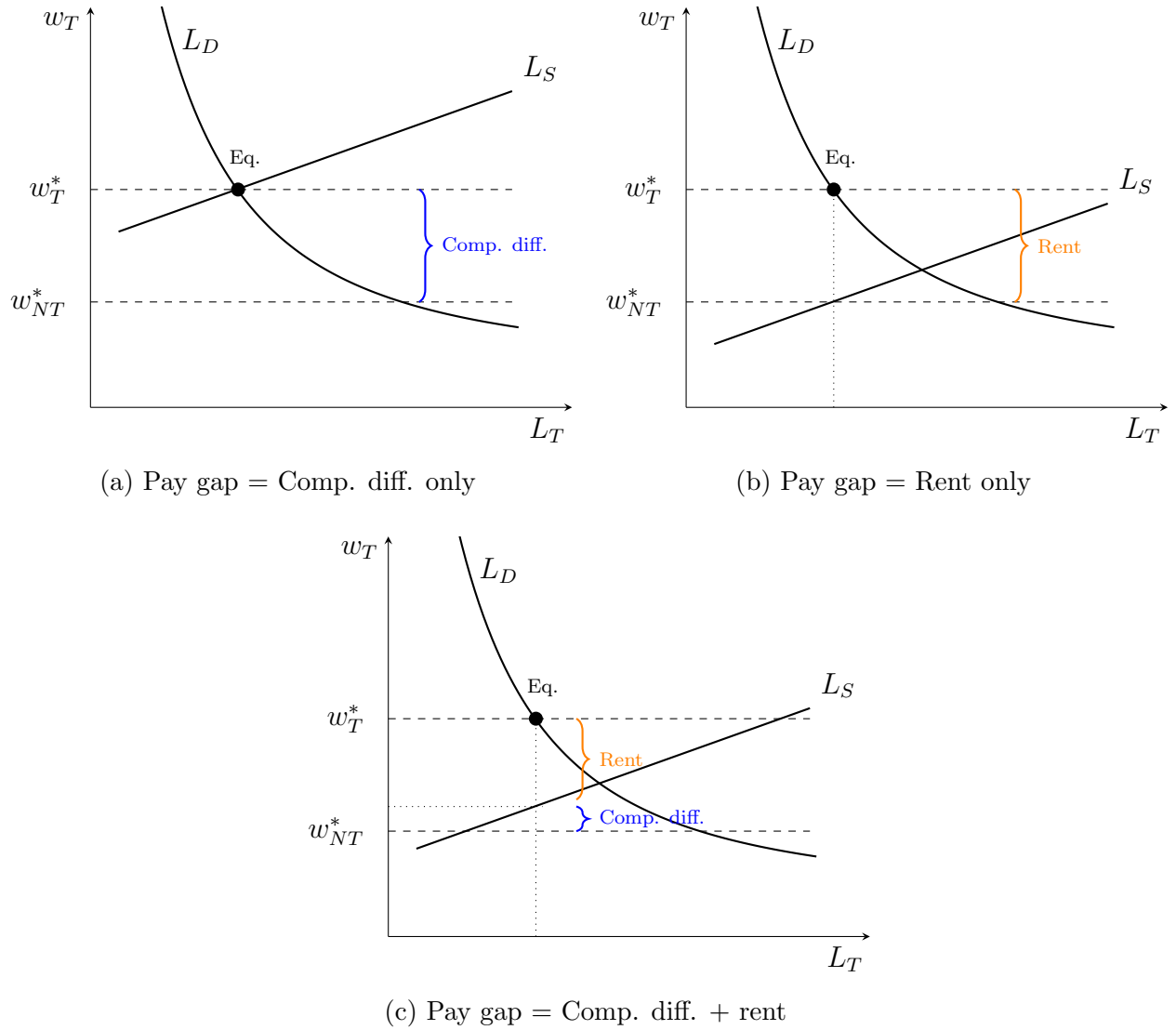
Figure 10: Illustrative examples showing that the teacher pay gap, the difference between the teaching wage w_T^* and the non-teaching wage w_{NT}^*, can be (a) only a compensating differential, (b) only a rent, or (c) both. In all three cases the pay gap is the same and positive, meaning in equilibrium there is a "teacher premium."

Because the number of teachers employed is dictated by the budget constraint for teachers, the equilibrium is given by demand where teacher pay w_T^*. However, whether teachers are earning a pure compensating differential, a pure rent, or a mix of both depends on where the equilibrium lies relative to the labor supply curve. An equilibrium that lies on (or below) the labor supply curve, as shown in (a), would indicate pure compensating differentials: despite the teaching premium, the marginal worker who enters teaching would not be willing to teach for less than w_T^*, because the pay premium exactly offsets the lower amenity value in teaching relative to teachers' next-best jobs. An equilibrium that lies above the labor supply curve, however, could constitute either a pure rent, as shown in (b), or a mix of rents and compensating differentials,

53

as shown in (c).

For completeness, we also illustrate the opposite case where there is a teaching penalty instead of a premium in Figure 11. If teaching offers sufficiently more attractive amenities than teachers' next-best job, such a teaching penalty could entirely be attributable to a compensating differential in the next-best job (or equivalently, a negative compensating differential in teaching). In fact, if teaching offers substantially more attractive amenities, teachers may extract rents even when a teacher penalty is in effect.

These figures illustrate that a pay premium in teaching on its own does not necessarily meant that teachers are extracting rents. In fact, rent extraction can also occur under a teacher penalty! Instead, assessing the extent to which teachers extract rents requires evaluating both the pay gap between teaching and teachers' next-best jobs as well as evaluating the difference in amenity value between teaching and teachers' next-best options. The following sections detail how we identify and estimate both quantities, starting with the pay gap.

(a) Pay gap = (Neg.) comp. diff. only   (b) Pay gap = (Neg.) comp. diff. + rent
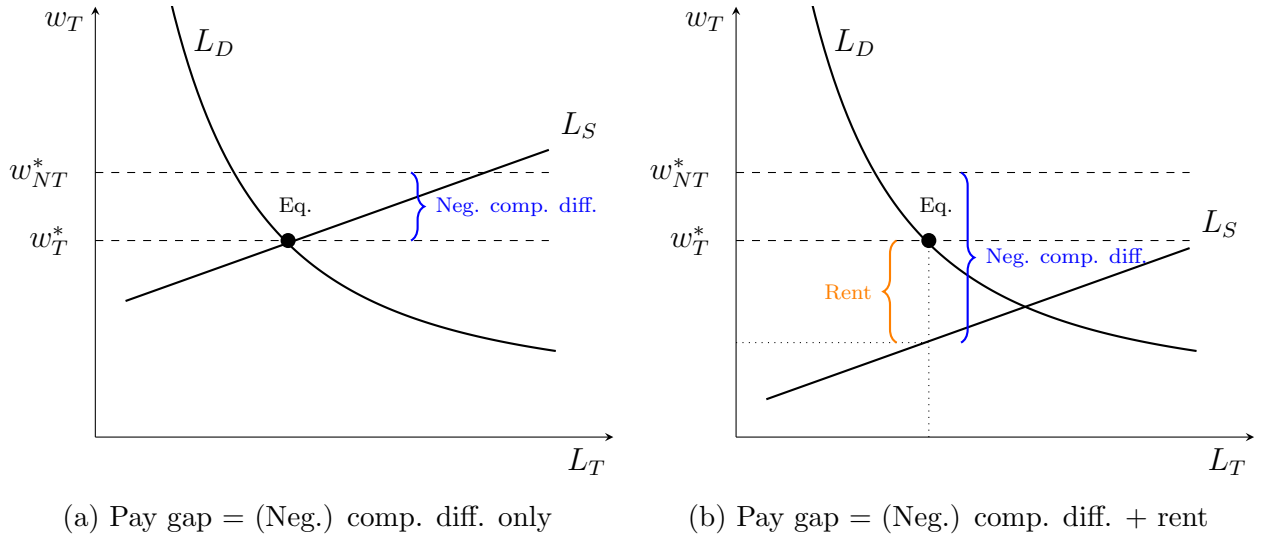
Figure 11: Illustrative examples showing that under a "teacher penalty," meaning an equilibrium in which $w_T^* < w_{NT}^*$, the teacher pay gap can be either (a) only a (negative) compensating differential, or (c) the combination of a negative compensating differential and a rent.

B   Identification Proofs

In this appendix, we provide proofs showing that under certain assumptions, the RD and RK estimands deliver weighted treatment effects of interest in settings where discontinuities arise (due to e.g. test score cutoffs) and retaking is allowed.

B.1 Notation

For exposition, assume that individuals can only retake the exam once. We use the following notation throughout the proofs that follow.

- U: person type (unobserved heterogeneity)

- S_{1i}: first-attempt score, drawn from $f_{S_1|U=u}(s)$

- R: indicator for whether one decides to retake the exam

- S_{2i}: second-attempt score, drawn from $f_{S_2|U=u,R=1}(s)$.

- $p_r(S_1, U)$: the probability that a person of type U retakes the test

- $p_c(U)$: the probability that a person of type U becomes certified, conditional on having passed the test at any point (first or second attempt)

- $y(C, U)$: potential earnings outcomes, where $Y_{C,i} = y(1, U)$ and $Y_{NC,i} = y(0, U)$.

This notation allows us to collapse the structural parameters explicit in the model framework (see Appendix E) that determine whether individuals retake the test or become certified into reduced form objects.

B.2 RD Design

Suppose there were a discontinuity in the observed likelihood of individuals becoming certified at the cutoff, resulting from the random assignment of individuals to the left or right of the cutoff. Then the 2SLS estimand that results from using whether the first-attempt score is on the right or left of the cutoff as an instrument for whether one becomes certified to teach is

$$\frac{\lim_{s \to 0^+} \mathbb{E}[y(C, U) \mid S_1 = s] - \lim_{s \to 0^-} \mathbb{E}[y(C, U) \mid S_1 = s]}{\lim_{s \to 0^+} \mathbb{E}[C \mid S_1 = s] - \lim_{s \to 0^-} \mathbb{E}[C \mid S_1 = s]}. \tag{10}$$

The following proof derives expressions for each of the four terms in (10) and shows that, when pieced together, they deliver a weighted average treatment effect.

Beginning with the numerator, observe that $\mathbb{E}[y(C, U) \mid S_1 = s]$ can be written as

$$\mathbb{E}[y(C, U) \mid S_1 = s] = \int \mathbb{I}\{s \geq 0\} \left[p_c(u)y(1, u) + (1 - p_c(u))y(0, u) \right]$$

$$+ \mathbb{I}\{s < 0\}(1 - p_r(s, u))y(0, u)$$

$$+ \mathbb{I}\{s < 0\}p_r(s, u)\Pr(S_2 < 0 \mid u, R = 1)y(0, u)$$

$$+ \mathbb{I}\{s < 0\}p_r(s, u)\Pr(S_2 \geq 0 \mid u, R = 1) \left[p_c(u)y(1, u) + (1 - p_c(u))y(0, u) \right] dF_{U|S_1=s}(u).$$

$$(11)$$

In words, (11) says that there are four possible routes one can take to reach a potential earnings outcome. First, they could pass the Praxis on their first attempt and either get certified or not. Second, they could fail the Praxis on their first attempt and choose not to retake the exam. Third, they could fail on their first attempt, choose to retake the exam, but fail again. Finally, they could fail on their first attempt, choose to retake, and pass. Each route leads either to the potential earnings outcome $y(1, U)$ or $y(0, U)$.

Breaking the expression up in this fashion makes it clear which terms are relevant when evaluating limits from the right and left of the cutoff. When taking the limit from the right of $s = 0$, we can ignore the three routes that involve $s < 0$ and only focus on the route that involves $s \geq 0$:

$$\lim_{s \to 0^+} \mathbb{E}[y(C, U) \mid S_1 = s] = \lim_{s \to 0^+} \int p_c(u)y(1, u) + (1 - p_c(u))y(0, u) \, dF_{U|S_1=s}(u)$$

$$= \lim_{s \to 0^+} \int y(0, u) + p_c(u)(y(1, u) - y(0, u)) \, dF_{U|S_1=s}(u) \qquad (12)$$

$$= \lim_{s \to 0^+} \int y(0, u) + p_c(u)(y(1, u) - y(0, u)) \frac{f_{S_1|u}(s)}{f_{S_1}(s)} dF_U(u),$$

where the substitution in the last line follows from Bayes rule.

Similarly, taking the limit from the left yields

$$\lim_{s \to 0^-} \mathbb{E}[y(C, U) \mid S_1 = s] = \lim_{s \to 0^-} \int (1 - p_r(s, u))y(0, u)$$

$$+ p_r(s, u)\Pr(S_2 < 0 \mid u, R = 1)y(0, u)$$

$$+ p_r(s, u)\Pr(S_2 \geq 0 \mid u, R = 1) \left[p_c(u)y(1, u) + (1 - p_c(u))y(0, u) \right] dF_{U|S_1=s}(u)$$

$$= \lim_{s \to 0^-} \int y(0, u)$$

$$+ p_r(s, u)\Pr(S_2 \geq 0 \mid u, R = 1)p_c(u)(y(1, u) - y(0, u)) \frac{f_{S_1|u}(s)}{f_{S_1}(s)} dF_U(u).$$

$$(13)$$

Then under the assumption that $f_{S_1|u}(s)$ and $f_{S_1}(s)$ are everywhere continuous and smooth through the cutoff $s = 0$, the numerator of the RD estimand is

$$\lim_{s \to 0^+} \mathbb{E}[y(C, U) \mid S_1 = s] - \lim_{s \to 0^-} \mathbb{E}[y(C, U) \mid S_1 = s]$$

$$= \int (y(1, u) - y(0, u)) \left[\lim_{s \to 0^+} p_c(u) - \lim_{s \to 0^-} p_c(u) p_r(s, u) \Pr(S_2 \geq 0 \mid u, R = 1) \right] \frac{f_{S_1|u}(0)}{f_{S_1}(0)} dF_U(u).$$
(14)

We can expand the denominator of (10) in a similar fashion. First, since C is a binary treatment, observe that we can write

$$\Pr[C = 1 \mid S_1 = s] = \int \mathbb{I}\{s \geq 0\} p_c(u)$$
$$+ \mathbb{I}\{s < 0\} p_r(s, u) \Pr(S_2 \geq 0 \mid u, R = 1) p_c(u) \, dF_{U|S_1 = s}(u).$$
(15)

Taking limits from the right and left of $s = 0$ yields

$$\lim_{s \to 0^+} \Pr[C = 1 \mid S_1 = s] = \lim_{s \to 0^+} \int p_c(u) \frac{f_{S_1|u}(s)}{f_{S_1}(s)} dF_U(u), \text{ and}$$

$$\lim_{s \to 0^-} \Pr[C = 1 \mid S_1 = s] = \lim_{s \to 0^-} \int p_c(u) p_r(s, u) \Pr(S_2 \geq 0 \mid u, R = 1) \frac{f_{S_1|u}(s)}{f_{S_1}(s)} dF_U(u).$$
(16)

Again under the assumption that $f_{S_1|u}$ and f_{S_1} are smooth around the cutoff $s = 0$, the denominator of the RD estimand is given by

$$\lim_{s \to 0^+} \Pr[C = 1 \mid S_1 = s] - \lim_{s \to 0^-} \Pr[C = 1 \mid S_1 = s]$$

$$= \int \left[\lim_{s \to 0^+} p_c(u) - \lim_{s \to 0^-} p_c(u) p_r(s, u) \Pr(S_2 \geq 0 \mid u, R = 1) \right] \frac{f_{S_1|u}(0)}{f_{S_1}(0)} dF_U(u).$$
(17)

Equation (17) has an important interpretation: it gives the magnitude of the discontinuity at the cutoff in the observed likelihood that one becomes certified. In other words, it is the effect of barely passing the Praxis on the first-attempt (versus barely failing) on the likelihood of becoming certified. This expression demonstrates that identification relies on there being a non-zero discontinuity; otherwise, the denominator would be zero and the estimand would be undefined.

From (17), we also see that depending on the setting, the discontinuity can be positive or negative. Although $p_c(u)$ is smooth through the cutoff, the retaking function $p_r(s, u)$ is not; those randomly assigned to the right of the cutoff retake with probability zero, while those randomly assigned to the left may retake with a positive probability. The discontinuity could

be positive if those to the left retake with some constant probability that is independent of u. But, the discontinuity could be negative if for those with $s < 0$, $p_r(s, u)$ is only positive for people with high $p_c(u)$, leading to

$$\lim_{s \to 0^-} p_c(u) p_r(s, u) \Pr(S_2 \geq 0 \mid u, R = 1) > \lim_{s \to 0^+} p_c(u).$$

Complete expression. Finally, we obtain an expression for the 2SLS estimand by dividing the numerator expression in (14) by the denominator expression in (17). The resulting expression is as previewed: a weighted average of the heterogeneous treatment effects $y(1, u) - y(0, u)$ over the distribution of u,

$$\frac{\lim_{s \to 0^+} \mathbb{E}[y(C, U) \mid S_1 = s] - \lim_{s \to 0^-} \mathbb{E}[y(C, U) \mid S_1 = s]}{\lim_{s \to 0^+} \mathbb{E}[C \mid S_1 = s] - \lim_{s \to 0^-} \mathbb{E}[C \mid S_1 = s]} = \int [y(1, u) - y(0, u)] \, \varphi_{RD}(u) \, dF_U(u), \tag{18}$$

where the weights are given by

$$\varphi_{RD}(u) = \frac{[\lim_{s \to 0^+} p_c(u) - \lim_{s \to 0^-} p_c(u) p_r(s, u) \Pr(S_2 \geq 0 \mid u, R = 1)] \frac{f_{S_1 \mid u}(0)}{f_{S_1}(0)}}{\int [\lim_{s \to 0^+} p_c(u) - \lim_{s \to 0^-} p_c(u) p_r(s, u) \Pr(S_2 \geq 0 \mid u, R = 1)] \frac{f_{S_1 \mid u}(0)}{f_{S_1}(0)} dF_U(u)}. \tag{19}$$

We can interpret the $\phi_{RD}(u)$ terms as weights as they are positive for all u. The RD estimand is most heavily weighted towards "compliers" as defined in Appendix D: individuals who either (1) pass on their first attempt and thus get certified, or (2) fail on their first attempt and do not get certified.[36]

B.3 RK Design

The above section shows that if there were an observed discontinuity in the likelihood of certification, an RD design would identify a weighted average treatment effect.

In our setting, the random assignment of individuals to the right or left of the cutoff does not induce a discontinuity in the observed likelihood of certification, rendering the RD design unusable. However, the same random assignment does induce a kink (i.e. a change in the slope) in the likelihood of certification at the cutoff. The following proof demonstrates that in such an environment, under a few additional assumptions, an RK design can be used to identify a weighted average treatment effect of interest.

---

[36] Note that, at a glance, the interaction between $p_c(u)$ and $(1 - \lim_{s \to 0^-} p_r(s, u) \Pr(S_2 \geq 0 \mid u, R = 1))$ makes it appear as though the individuals who will receive the greatest weights will be those who do not retake the exam *and* become certified. However, this combination of events cannot occur in our context. Put differently, there are "no defiers" in this environment.

The 2SLS estimand that arises from using (whether one scores to the right or left of the cutoff) $\times$ (the distance of their score from the cutoff) as an instrument for getting certified is given by

$$\frac{\lim_{s\to 0^+} \frac{d\mathbb{E}[y(C,U)|S_1=s]}{ds} - \lim_{s\to 0^-} \frac{d\mathbb{E}[y(C,U)|S_1=s]}{ds}}{\lim_{s\to 0^+} \frac{d\mathbb{E}[C|S_1=s]}{ds} - \lim_{s\to 0^-} \frac{d\mathbb{E}[C|S_1=s]}{ds}}. \tag{20}$$

The following proof follows the same general steps as the "discontinuity" proof above: we derive expressions for each of the four terms that appears in (20) and show that pieced together they deliver another weighted average treatment effect.

Notation. For notational convenience, we will use $\Lambda(s)$ to denote the following expression which appears throughout the proof:

$$\Lambda(s) \equiv \frac{\partial \left(\frac{f_{S_1|u}(s)}{f_{S_1}(s)} \right)}{\partial s} = \frac{f_{S_1|u}(s)}{f_{S_1}(s)} \left[\frac{f'_{S_1|u}(s)}{f_{S_1|u}(s)} - \frac{f'_{S_1}(s)}{f_{S_1}(s)} \right]. \tag{21}$$

Numerator. We begin by expanding the numerator. We already derived the general expression for $\mathbb{E}[y(C,U) \mid S_1 = s]$ in (11). Taking derivatives and limits of (11) from the right hand side, we have

$$\begin{aligned}
\lim_{s\to 0^+} \frac{d\mathbb{E}[y(C,U) \mid S_1 = s]}{ds} &= \lim_{s\to 0^+} \frac{d}{ds} \int [p_c(u)y(1,u) + (1 - p_c(u))y(0,u)] \frac{f_{S_1|u}(s)}{f_{S_1}(s)} \, dF_U(u) \\
&= \int [p_c(u)y(1,u) + (1 - p_c(u))y(0,u)] \lim_{s\to 0^+} \Lambda(s) \, dF_U(u),
\end{aligned} \tag{22}$$

and from the left hand side, we have

$$\begin{aligned}
\lim_{s\to 0^-} &\frac{d\mathbb{E}[y(C,U) \mid S_1 = s]}{ds} = \lim_{s\to 0^-} \frac{d}{ds} \int y(0,u) \\
&+ p_r(s,u)\mathrm{Pr}(S_2 \geq 0 \mid u, R = 1)p_c(u)\left(y(1,u) - y(0,u)\right) \frac{f_{S_1|u}(s)}{f_{S_1}(s)} dF_U(u) \\
&= \lim_{s\to 0^-} \int y(0,u)\Lambda(s) \, dF_U(u) \\
&+ \int [y(1,u) - y(0,u)]\, p_c(u)\mathrm{Pr}(S_2 \geq 0 \mid u, R = 1) \lim_{s\to 0^-} \left[\frac{\partial p_r(s,u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} + p_r(s,u)\Lambda(s) \right] dF_U(u).
\end{aligned} \tag{23}$$

59

Then the complete numerator, the difference between (22) and (23), is

$$\lim_{s\to 0^-}\frac{d\mathbb{E}[y(C,U)\mid S_1=s]}{ds}-\lim_{s\to 0^-}\frac{d\mathbb{E}[y(C,U)\mid S_1=s]}{ds}$$

$$=\int [y(1,u)-y(0,u)]\,p_c(u)\times$$

$$\left\{\lim_{s\to 0^+}\Lambda(s)-\Pr(S_2\geq 0\mid u,R=1)\lim_{s\to 0^-}\left[\frac{\partial p_r(s,u)}{\partial s}\frac{f_{S_1\mid u}(s)}{f_{S_1}(s)}+p_r(s,u)\Lambda(s)\right]\right\}dF_U(u)$$

$$=\int [y(1,u)-y(0,u)]\times\left(\lim_{s\to 0^+}p_c(u)\Lambda(s)\right.$$

$$-\lim_{s\to 0^-}p_c(u)\Pr(S_2\geq 0\mid u,R=1)p_r(s,u)\Lambda(s)\tag{24}$$

$$\left.-\lim_{s\to 0^-}\Pr(S_2\geq 0\mid u,R=1)\frac{\partial p_r(s,u)}{\partial s}\frac{f_{S_1\mid u}(s)}{f_{S_1}(s)}\right)dF_U(u)$$

$$=\underbrace{\int [y(1,u)-y(0,u)]\cdot\Lambda(0)\cdot\left(\lim_{s\to 0^+}p_c(u)-\lim_{s\to 0^-}p_c(u)\Pr(S_2\geq 0\mid u,R=1)p_r(s,u)\right)dF_U(u)}_{(a)}$$

$$-\underbrace{\int [y(1,u)-y(0,u)]\left(\lim_{s\to 0^-}p_c(u)\Pr(S_2\geq 0\mid u,R=1)\frac{\partial p_r(s,u)}{\partial s}\frac{f_{S_1\mid u}(s)}{f_{S_1}(s)}\right)dF_U(u)}_{(b)}.$$

$$\tag{25}$$

Consider expanding (a) by plugging in the full expression for $\Lambda(s)$:

$$\int [y(1,u)-y(0,u)]\cdot\frac{f_{S_1\mid u}(0)}{f_{S_1}(0)}\left[\frac{f'_{S_1\mid u}(0)}{f_{S_1\mid u}(0)}-\frac{f'_{S_1}(0)}{f_{S_1}(0)}\right]\left(\lim_{s\to 0^+}p_c(u)-\lim_{s\to 0^-}p_c(u)\Pr(S_2\geq 0\mid u,R=1)p_r(s,u)\right)d\!$$

$$\tag{26}$$

Under the assumption that s is randomly assigned around the cutoff, there is mean independence between

$$[y(1,u)-y(0,u)]\cdot\left[\frac{f'_{S_1\mid u}(0)}{f_{S_1\mid u}(0)}-\frac{f'_{S_1}(0)}{f_{S_1}(0)}\right]$$

and

$$\frac{f_{S_1\mid u}(0)}{f_{S_1}(0)}\cdot\left(\lim_{s\to 0^+}p_c(u)-\lim_{s\to 0^-}p_c(u)\Pr(S_2\geq 0\mid u,R=1)p_r(s,u)\right).$$

Since $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ for mean independent X and Y, (a) is equivalently

$$\int [y(1, u) - y(0, u)] \cdot \left[\frac{f'_{S_1|u}(0)}{f_{S_1|u}(0)} - \frac{f'_{S_1}(0)}{f_{S_1}(0)} \right] dF_U(u)$$
$$\times \int \left(\lim_{s \to 0^+} p_c(u) - \lim_{s \to 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) p_r(s, u) \right) \frac{f_{S_1|u}(0)}{f_{S_1}(0)} dF_U(u). \tag{27}$$

$$= \int [y(1, u) - y(0, u)] \cdot \left[\frac{f'_{S_1|u}(0)}{f_{S_1|u}(0)} - \frac{f'_{S_1}(0)}{f_{S_1}(0)} \right] dF_U(u)$$
$$\times \left(\lim_{s \to 0^+} \Pr[C = 1 \mid S_1 = s] - \lim_{s \to 0^-} \Pr[C = 1 \mid S_1 = s] \right), \tag{28}$$

where the last line follows from equation (17). With unrestricted heterogeneity in u, it is possible that the first term is positive or negative. However, recall that we assumed there is no discontinuity in $\Pr[C = 1 \mid S_1 = s]$ at the cutoff, which is only possible if the following two conditions hold for $s < 0$:

1. $\frac{\partial p_r(s, u)}{\partial s} > 0$: Scoring higher on the exam makes one more likely to retake ("encouragement" effect), and

2. $\mathrm{Corr}(p_c(u), p_r(s, u)) > 0$: Those with greater desire to become certified also are more likely to retake the exam ("intrinsic motivation" effect).

Thus, under the assumption there is no discontinuity in $\Pr[C = 1 \mid S_1 = s]$, the entire (a) term is zero.

Thus, the numerator is given by only (b),

$$\int [y(1, u) - y(0, u)] \left(0 - \lim_{s \to 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) \frac{\partial p_r(s, u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} \right) dF_U(u). \tag{29}$$

Denominator. We can similarly take derivatives and limits of the expression for the main component of the RD estimand denominator given by (15) to get the following expression for the RK estimand denominator:

$$\lim_{s \to 0^+} \frac{d\Pr[C = 1 \mid S_1 = s]}{ds} - \lim_{s \to 0^-} \frac{d\Pr[C = 1 \mid S_1 = s]}{ds}$$
$$= \int \Lambda(0) \cdot \left(\lim_{s \to 0^+} p_c(u) - \lim_{s \to 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) p_r(s, u) \right)$$
$$- \lim_{s \to 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) \frac{\partial p_r(s, u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} dF_U(u). \tag{30}$$

Following similar steps as in the numerator derivation (i.e. plugging in the expression for $\Lambda(s)$ and assuming there is no discontinuity in $\Pr[C = 1 \mid S_1 = s]$ at the cutoff), we can simplify the denominator to

$$\int \left(0 - \lim_{s \to 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) \frac{\partial p_r(s, u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} \right) dF_U(u). \tag{31}$$

Notice that the denominator is negative under the two conditions above that are necessary and sufficient for there to be no discontinuity in the likelihood of getting certified—that is, the slope in the likelihood that one becomes certified becomes "flatter" at the cutoff.

Complete expression. Finally we obtain an expression for the RK 2SLS estimand by dividing the numerator (25) by the denominator (31). The resulting expression is again a weighted average of the heterogeneous treatment effects $y(1, u) - y(0, u)$ over the distribution of u,

$$\frac{\lim_{s \to 0^+} \frac{d\mathbb{E}[y(C,U)|S_1=s]}{ds} - \lim_{s \to 0^-} \frac{d\mathbb{E}[y(C,U)|S_1=s]}{ds}}{\lim_{s \to 0^+} \frac{d\mathbb{E}[C|S_1=s]}{ds} - \lim_{s \to 0^-} \frac{d\mathbb{E}[C|S_1=s]}{ds}} = \int [y(1, u) - y(0, u)] \varphi_{RKD}(u) dF_U(u). \tag{32}$$

The weights are given by

$$\varphi_{RKD}(u) = \frac{\left(0 - \lim_{s \to 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) \frac{\partial p_r(s,u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} \right)}{\int \left(0 - \lim_{s \to 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) \frac{\partial p_r(s,u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} \right) dF_U(u)}. \tag{33}$$

We can interpret $\psi_{RKD}(u)$ as weights because $\psi_{RKD}(u)$ is non-negative for all u. The RK estimand gives no weight to individuals whose retaking behavior does not respond to their score, thus excluding individuals with such low intrinsic motivation that they would never retake the test, as well as individuals who would always retake the test no matter what. Instead, it gives weight to those who are are willing to retake the exam if they were to not pass on their first attempt, and gives relatively more weight to individuals who are more likely to be deterred from retaking if they score far below the cutoff initially (i.e. those who respond more strongly to the instrument, s).

C   Model simulations: From sharp RD to fuzzy RK

This appendix provides a series of simulations that illustrate when and why a kink might arise in place of a discontinuity under dynamic regression discontinuities. First, we show how a RD design approximates a randomized experiment if test-takers were not allowed to retake the test.

Second and third, we extend the framework to incorporate the fact that retaking is allowed, first by assuming everyone retakes the exam, and then relaxing this assumption to allow retaking behavior to respond to behavioral incentives. Fourth, we demonstrate that our data is consistent with the model. Fifth, we describe our estimators that give weighted average treatment effects of becoming certified on employment outcomes. We discuss the interpretation of the different weights given by the RD and RKD designs. Sixth, we describe how we implement estimation with the pooled sample of first-time test-takers in our data.

C.1 Regression discontinuity under one-shot exams

Our empirical design resembles that of Card et al. [2015] save for two important differences. First, our treatment of interest is binary (getting certified) rather than continuous. Second, in our context, whether one gets treatment is a function of behavioral choices (retaking the exam) rather than being the result of a one-shot evaluation of a deterministic rule.

We begin by describing the one-shot world and add in retaking in the next section. Consider the non-separable model

$$Y = y(C, R, \{S_1, S_2\}, U)$$

where Y denotes the labor market outcome of interest, C is an indicator for getting certified, S_1 and S_2 denotes the two possible scores on the Praxis on various attempts where S_2 is only observed if the individual chooses to retake (i.e. $R = 1$), and U captures individual heterogeneity. Individuals have some underlying test-taking ability μ_i such that their scores are determined by $S_{ai} = \mu_i + \varepsilon_i$ for all attempts a and mean zero ε_i. In other words, they do not have precise control over their score.

The treatment of interest is whether one gets certified: $C = C(\{S_1, S_2\}, R, \xi)$,[37] which can only happen if one passes on their first attempt ($S_1 \geq 0$) or on their second attempt ($R = 1$ and $S_2 \geq 0$). Even if one passes, there is additional randomness ξ_i that can influence certification completion, which may be correlated with U_i and thus also with Y_i.

Suppose for now that retaking is not allowed. The top row of Figure 12 shows simulations of what the first stage relationships between first-attempt scores (S_1) and the likelihood of passing the test or becoming certified would look like.[38] Those who do not pass can never pass

---

[37]The analogous expression in Card et al. [2015] is the "policy function" or "treatment assignment rule" underlying the fuzzy regression kink design. However, in the case of Card et al., the policy function is a deterministic rule that exhibits a kink at some point in the support of the underlying variable that determines C, whereas in our setting the function contains a random component.

[38]We assume that among those who pass, high-scorers are no more likely than barely-passers to enroll in and complete an EPP. However, as we know from the descriptives in Section ??, many who pass the Praxis do not continue through an EPP. We therefore choose parameters in our simulation to replicate this observation, with
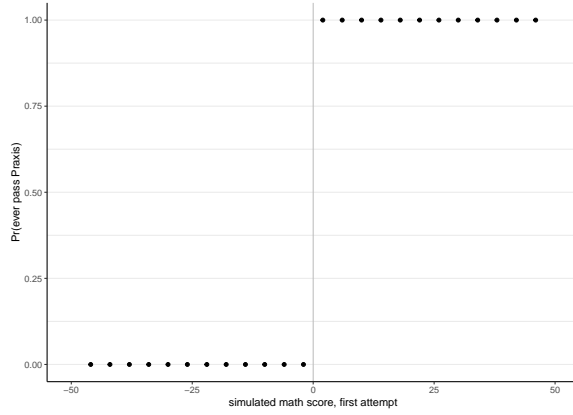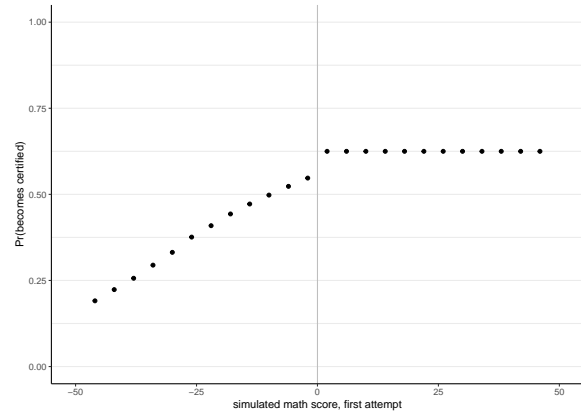
(a) Simulated Pr(pass), one-shot
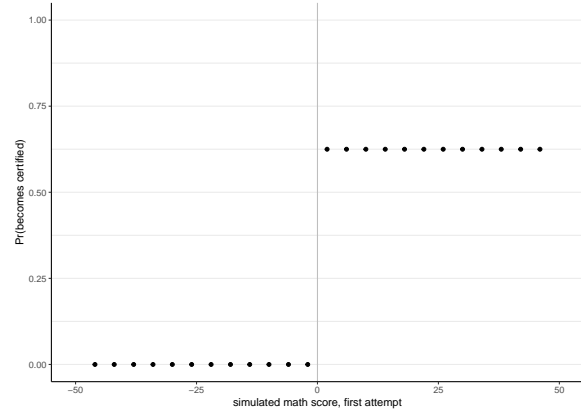
(b) Simulated Pr(certified), one-shot
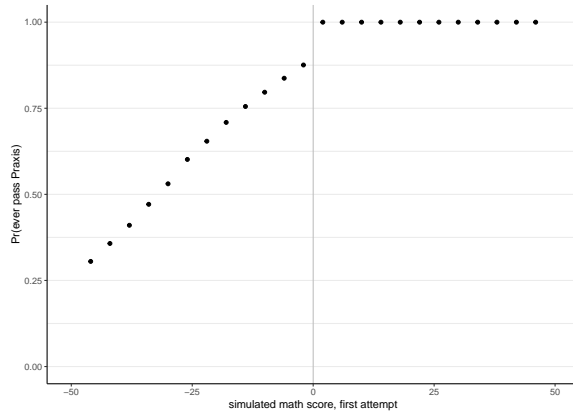
(c) Simulated Pr(pass), with automatic retaking

(d) Simulated Pr(certified), with automatic retaking
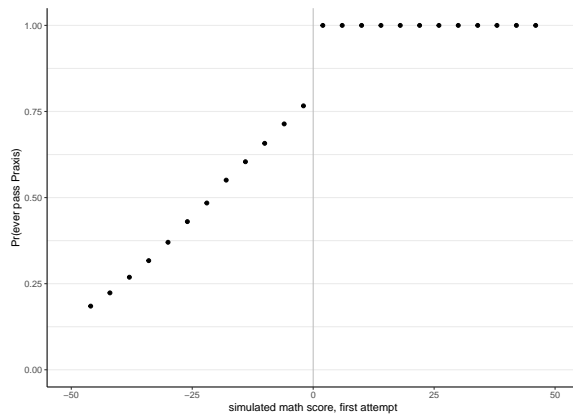
(e) Simulated Pr(pass), with behavioral retaking
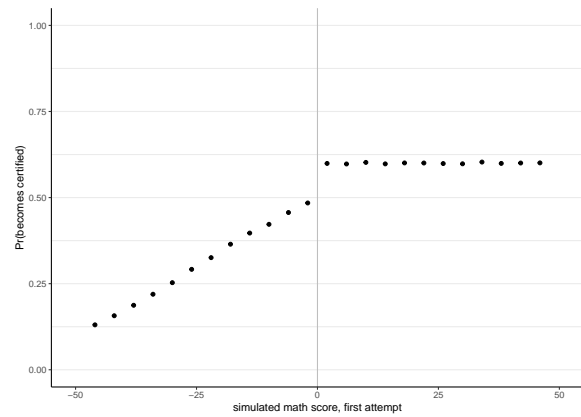
(f) Simulated Pr(certified), with behavioral retaking

Figure 12: Progression of first-stages from simulated one-shot test (1st row), to simulated dynamic test-taking with automatic retaking (2nd row), to simulated dynamic test-taking with behavioral retaking (3rd row).

the exam nor become certified, generating clean discontinuities in the likelihood of passing and of getting certified at $S_1 = 0$.

We are interested in the effect of getting certified on earnings. Regressing earnings on whether one gets certified would give a biased estimate of the parameter of interest, since the decision to get certified involves choices that may be made using information that is endogenous to earnings (i.e. $\mathbb{E}[\xi' U] \neq 0$).

However, notice that in this world, we can use whether one passes the Praxis on their first attempt, $\mathbb{I}(S_1 \geq 0)$, as a valid instrument for whether one gets certified. First, because the test can only be attempted once, passing the Praxis strongly affects the likelihood of certification. Secondly, the exclusion restriction holds: since individuals around the threshold are exogenously assigned permission to continue down the teacher pipeline and the Praxis is only relevant for teaching accreditation, the only channel through which we would expect passing the Praxis to affect earnings is through completing the teacher pipeline.

C.2 Dynamic regression discontinuities under automatic retaking

We now extend the one-shot framework to allow for the real-world possibility of retaking. For exposition, assume that individuals have are endowed with at most two retakes and that there is no cost to retaking the exam, such that all individuals who fail on their first try retake ($R = 1$) and realize their S_2.

The middle row of Figure 12 shows the resulting simulated first-stages under retaking. As in the one-shot case, those who pass on their first attempt have a constant probability of becoming certified. Furthermore, under the data-generating process $S_{ai} = \mu_i + \varepsilon_i$, those with higher first-attempt scores are more likely to have high μ_i. Thus, for retakers, the probability that they will pass on their second attempt and become certified is increasing in their initial score S_{1i}. The result visually is a discontinuity and now also a kink in the likelihood of passing and of getting certified at the cutoff.

Since the discontinuity is still present and pronounced, similarly to the one-shot world we can still use whether one passes the Praxis on their first attempt, $\mathbb{I}(S_1 \geq 0)$, as a valid instrument for whether one gets certified.

---

passers having a 2/3 probability of becoming certified.

C.3 Dynamic regression discontinuities under behavioral retaking

Finally, we extend the retaking framework once more by relaxing the assumption that individuals who fail automatically retake. We now allow for the possibility that not everyone who fails the exam initially will retake.

Assume that for those who fail initially, retaking behavior is influenced by two factors:

1. The "encouragement" factor: Getting a score closer to the cutoff increases the likelihood one retakes the exam.

2. The "intrinsic motivation" factor: Individuals who have a higher baseline desire to get certified are more likely to retake the exam.

Under these assumptions, individuals with low intrinsic motivation to get certified and/or individuals who are prone to being discouraged by getting a realized score below the cutoff will not retake the exam. Those who retake will be highly intrinsically motivated and/or encouraged by having scored close to the cutoff on their first attempt.

The last row of Figure 12 shows the resulting simulated first-stages under the two behavioral assumptions above. As before, those with higher first-attempt scores are more likely to have high μ_i. But in addition, because retakers are selected on their motivation to get certified and are influenced by the random assignment of their initial score, the slope in the likelihood of certification is steeper to the left of the cutoff. The result visually a stronger kink in the likelihood of passing and of getting certified at the cutoff. In particular, there is a diminished discontinuity in the likelihood of getting certified at the cutoff, which could be zero or even negative depending on the strength of the "intrinsic motivation" effect.

In this environment, the diminished or lack of discontinuity renders the instrument $\mathbb{I}(S_1 \geq 0)$ weak or invalid. However, we can construct a valid instrument that leverages both (1) the fact that scores are randomly assigned near the cutoff *and* (2) the actual value of the score influences retaking behavior: $\mathbb{I}(S_1 \geq 0) \times s$. As is evident by the kink, the instrument is relevant: near the cutoff, those who pass are randomly assigned a lower cost of continuing on to certification, while those who fail are randomly assigned a higher cost that increases with the distance of S_1 from the cutoff. The exclusion restriction would be threatened if the act of retaking the exam influences earnings through a channel other than certification, such as by improving one's human capital through the studying process alone. However, given that the exams test general core skills, it is unlikely that this channel exists. Thus, we assume the exclusion restriction holds as well.

C.4 Reconciling the model and data

The last row of Figure 12 can be contrasted with the actual first-stages in the data shown in Figure 3. The actual first-stages in our data are consistent with the model, especially in the kinked shape of the functions. In addition, the discontinuity in the likelihood of certification in the data is virtually zero, suggesting that retaking behavior is affected by both the "encouragement" effect and the "intrinsic motivation" effect.

D Re-casting the LATE framework with retaking

Because we estimate our RK estimand using a 2SLS estimator, it must have a LATE interpretation (Angrist et al.). In this section, we define what it means to be a "complier" when there can be multiple attempts at becoming eligible for treatment.

Typically, when treatment is assigned based on a one-shot deterministic rule, the union of actions taken by always-takers, never-takers, and compliers account for the full the set of possible actions. If individuals could only attempt the Praxis once, then compliers would be individuals who become certified if they pass on their first attempt but do not become certified if they fail on their first attempt. Always-takers in this setting are individuals who, regardless of whether pass or fail on the first attempt, will find a way to obtain certification. It is only possible to obtain certification by retaking the exam. Thus, always-takers are individuals who either pass on their first attempt or would retake the exam until they pass. Never-takers in our setting are individuals who will never become certified, regardless of whether they pass or fail on their first attempt.

However, when individuals have multiple chances to influence their eligibility to obtain treatment, the three classic types of behavior of always-takers, never-takers, and compliers no longer cover all possible actions. In fact, in settings like ours, individuals can take actions that do not fit any of the typical definitions of always-takers, never-takers, and compliers. For example, consider individuals who pass on their first attempt, retake once, but do not get certified. These individuals could be thought of as never-takers, but they could also be thought of as individuals who are unwilling to take the test more than once, or who are discouraged by their score on the second exam and therefore do not continue. There is no natural definition for such behavior in the classic LATE framework.

We therefore re-cast the concepts of compliers, always-takers, and never-takers with respect to retaking behavior. We define a "retaking complier" as someone whose retaking behavior is responsive to the first-attempt score they receive. Compliers do not retake if they pass on

their first attempt, but otherwise retake with positive probability that is increasing in their first-attempt score. We define a "retaking always-taker" as someone who always retakes if they fail on their first attempt. We define a "retaking never-taker" as someone who never retakes if they fail on their first attempt. Finally, we define a "retaking defier" as someone who only retakes if they pass on their first attempt. We (sensibly) do not observe defiers in our data.

The identification proofs in Appendix B show that the RK estimator we use only gives positive weight to "retaking compliers." The data suggests that the majority of test-takers are compliers: the retaking rate among those who fail on the first attempt is high (80%) and there is a significant positive relationship between the likelihood of retaking and initial scores.

E Model framework of selection into teaching

The following theoretical model framework traces out the teacher pipeline in Kentucky.

In the economy, the going salary rate for public school teachers is given by Y_T. Individuals differ on three margins: their potential "outside option" earnings $Y_{NC,i}$[39], their desire to teach v_i, and their mean test taking ability μ_i. We assume that the only channel through which certification affects earnings is via becoming a teacher; that is, the earnings one gets if they are not a teacher is given by $Y_{NT,i} = Y_{NC,i}$

The main expected outcome of going through the teacher pipeline is to eventually become a teacher. Thus, individuals choose whether to progress at each step of the pipeline based on their expected utility of becoming a teacher.[40] We denote whether individual i has reached each stage in the pipeline using a series of indicators: $P_i = 1$ if i attempts the Praxis, $R_i = 1$ if i retakes the Praxis, $C_i = 1$ if i completes their EPP and obtains certification, and $T_i = 1$ if i becomes a teacher. For exposition, we only entertain the possibility of one retake.

Individual i of unobservable type $U_i = (Y_{NT,i}, v_i, \mu_i)$ gets utility $\mathbb{U}_i$ from being a teacher, given by

$$\mathbb{U}_i = \gamma_0 + \gamma_1(Y_T - Y_{NT,i}) + v_i, \tag{34}$$

where γ_1 captures strength of preference for higher earnings and v_i captures idiosyncratic taste for teaching.[41] Those who become teachers are those for whom $\mathbb{U}_i > 0$.

However, individuals initially do not observe their true $Y_{NC,i}$ and v_i and never observe their

---

[39]Should this include non-employment?

[40]Conlon and Patel [2022] show that undergraduates tend to have biased beliefs about the occupation they will hold after graduation. However, how these beliefs relate to their earnings expectations we are unsure of. Here, we assume that they hold unbisaed beliefs about earnings.

[41]Note that taste for teaching can be influenced by a whole host of factors which we will later separate into factors that directly affect one's effectiveness as a teacher and factors that only affect one's enjoyment of the occupation, like working with children.

true μ_i. Instead, the timing of information revelation and decisions through the pipeline goes as follows:

1. **Draw types.** Individuals draw types $u_i = (Y_{NC,i}, v_i, \mu_i) \sim f(u)$. They observe noisy signals of their type given by $\tilde{Y}_{NC,i} = Y_{NC,i} + \nu_{1i}$, $\tilde{v}_i = v_i + \nu_{2i}$, and $\tilde{\mu}_i = \mu_i + \nu_{3i}$, where individuals' beliefs are correct in expectation (i.e. ν_{ki} are mean zero).

2. **Decision: Take the Praxis.** Individuals decide on their test-taking behavior with a two-stage decision. First, they decide whether to take the Praxis at all ($P_i = 1$ if they choose to attempt it, and $P_i = 0$ otherwise) based on their expected utility of becoming a teacher. Those who decide to take the Praxis then determine the number of attempts they will budget for themselves to pass the exam—we will refer to this as the "test-taking budget" or just "the budget" from here on. Then, individuals take and re-take the test until either their pre-determined budget is exhausted or they pass. For exposition, we assume that the budget can be one or two attempts at most.

Decision 2a: Take the Praxis at all. Those who choose $P_i = 1$ are those for whom

$$\mathbb{E}[\mathbb{U}_i \mid \tilde{v}_i, \tilde{Y}_{NT,i}] = \gamma_0 + \gamma_1(Y_T - \tilde{Y}_{NT,i}) + \tilde{v}_i > 0. \tag{35}$$

- Those who do not attempt the Praxis ($P_i = 0$) exit the pipeline and earn their realized outside option earnings, $Y_{NC,i}$.

Decision 2b. Compute one's budget. Individuals with $P_i = 1$ compute their budget of value one or two, which is determined by a reduced-form function $b(\tilde{Y}_{NT,i}, \tilde{v}_i, \tilde{\mu}_i, X_i, \eta_i)$. The function depends on the individuals' belief about their type u, their observable demographics X_i, and some random error η_i. The budget is binding unless individuals pass the exam before exhausting their budget.

Why is the budget a function of all these characteristics? Firstly, it is important that the budget is a function of both the individuals' belief about their outside option and their preferences, because the key trade-off individuals face when deciding whether to teach is between how much they think they will like the job (compared to other options) and how much they think they can be paid elsewhere. Those who think they will really like it, all else equal, will be more willing to budget for a retake than those do think they will not like it. However, secondly, the test is costly: individuals must pay a fixed fee for each attempt and are subject to a waiting period in between attempts. Thus, there may be individuals who think they will really like teaching but cannot afford to spend more money on a retake or to wait for the entire waiting period to take the test again—this is

captured by observable demographics, like students' socio-economic status. Finally, the ease with which one thinks they will be able to pass the test may also affect the choice of budget.

Action: Take (and perhaps retake) the Praxis. Individuals who have chosen $P_i = 1$ and their budget $b_i \in \{1, 2\}$ now carry out their test-taking. On the first attempt, individuals draw a first-attempt score $S_{1i} = \mu_i + \varepsilon_{1i}$. A score of $S_{1i} \geq 0$ is considered a "pass," while a score of $S_{1i} < 0$ is a "fail."

- Those with $S_{1i} \geq 0$ can move on to the next step of the pipeline (enrolling in and completing an EPP, if they so choose).

- Those with $S_{1i} < 0$ and $b_i = 1$ do not retake the exam, i.e. $R_i = 0$, and exit to their outside option.

- Those with $S_{1i} < 0$ and $b_i = 2$ may re-take the exam, i.e. $R_i = 1$.

On the second attempt, individuals draw a second-attempt (and, for exposition, final) score $S_{2i} = \mu_i + \varepsilon_{2i}$.[42]

- Those with $S_{2i} \geq 0$ can move on to the next step of the pipeline.

- Those with $S_{2i} < 0$ exit to their outside option.

Notice that because scores are drawn from the same distribution for each individual, among re-takers, those who had high "failing" scores are more likely to eventually pass than those who had low "failing" scores.

3. **Decision: Complete the EPP (i.e. obtain certification).** Those who pass the Praxis on either their first or second attempt must now decide whether or not to get certified. Certification involves a few steps beyond taking the Praxis, namely enrollment in an EPP and course completion, all of which we compress into a single step in this model. As individuals progress through the EPP, approaching graduation and the possibility of entering the labor market, individuals learn their true outside option earnings, $Y_{NC,i}$, and receive a shock related to graduation, ξ_i. We observe an individual completing their EPP and obtaining their certification, $C_i = 1$, when

$$\mathbb{E}[\mathbb{U}_i \mid \tilde{v}_i, Y_{NC,i}] = \gamma_0 + \gamma_1(Y_T - Y_{NC,i}) + \xi_i + \tilde{v}_i > 0 \tag{36}$$

---

[42]One possible extension is to add effort into the model to match the empirical fact that the distribution of scores on the second attempt is very similar to the distribution of scores on the first attempt, but is shifted to the right. Doing so does not change the model conclusions that we focus on.

- Those who do not complete their EPP ($C_i = 0$) exit to their realized outside option, $Y_{NC,i}$.

- Those who do complete their EPP and obtain their certificate ($C_i = 1$) progress to the next step.

Note that this specification allows for two distinct reasons that one might choose not to complete an EPP at this point. The first is that their outside option signal is sufficiently, outweighing their believed desire to teach. The second is that their outside option is low but it is too costly for them to finish the EPP, perhaps because they find it challenging to complete the coursework or because another life shock interferes with their coursework.

4. *Decision: Become a teacher in public schools.* Finally, after completing the EPP, individuals who are certified can become teachers in public schools.

The outcome of this stage—"becoming a teacher"—differs in an important way from the previous stages. Up until this point, all decisions along the pipeline have been the decision of the individual, i.e. "supply-side" decisions. However, whether a person becomes a teacher is a product of both demand and supply forces: the person must have obtained certification and applied to public school teaching positions, and a school principal/hiring committee must have chosen to hire said person.

Thus far, we have written down inequalities at each step of the pipeline that explain why individuals progress or not: whether $P_i = 1$, $R_i = 1$, and $C_i = 1$. However, we abstain from writing down a model of why one becomes a teacher because whether $T_i = 1$ (or $T_i = 0$) depends on how the teacher labor market clears.

More specifically, T_i is affected by three sets of information about preferences that we cannot disentangle by only observing T_i. Firstly, T_i is the product of application and hiring processes. It is rare to observe data on the application and hiring processes in U.S. public schools because the process is highly decentralized. The lack of information on this margin makes it difficult to say whether $T_i = 0$ if one did not apply or if one applied but was not chosen for the job.

Secondly, T_i is a function of principals' hiring ability. It is unclear whether hiring decisions are related to dimensions of teacher "quality" at all. If principals are effective at screening the most effective teachers out of the pool of applicants, we might imagine that $T_i = 1$ individuals are different in unobservable ways from $T_i = 0$ teachers who applied but did not get the job. The general consensus in the literature at the moment is that principals are hiring relatively blind to teacher "ability," but it remains an open question whether

71

this is truly the case or whether we do not yet have measures that capture what principals screen on.

Finally, T_i depends on the attractiveness of an unobserved market for teachers: private schools (although we have data that allows us to identify a significant share of private schools in the state). Although private schools do not require teacher certification, anecdotally it is common for private school teachers to obtain and maintain their state-approved teacher certification "just in case" they would like to switch to the public sector. Due to this data limitation, we may observe $T_i = 0$ when in fact the individual does take a teaching job, but in the private sector.

Thus, in place of writing a model for how T_i is determined, we simply acknowledge at this point that we observe whether individuals become teachers, T_i, and that this status is determined by a combination of application behavior, hiring prowess, or preferences for amenities in the private sector.

5. **Decision: Remain a teacher in public schools.** Those who do become public school teachers have a choice of whether to stay a teacher in public schools or to move on, either to a different position in the public school system ("up" to a principal or school administrator position, or "sideways" to another staff position) or out of the public school system all together. Whether an individual chooses to leave comes down to their realized true preference for teaching, v_i. Those who choose to stay are those for whom

$$\mathbb{E}[\mathbb{U}_i \mid v_i, Y_{NC,i}] = \gamma_0 + \gamma_1(Y_T - Y_{NC,i}) + \xi_i + v_i > 0. \tag{37}$$

Notice that individual's preference for teaching is not "changed" by the profession, but is instead only truly realized with experience as a teacher.

Do individuals appear to have accurate beliefs about their future earnings? The intuition being our "budgeting" model framework in Appendix E is that those who retake do so because they believe they have more to gain from retaking the exam, while those who do not retake are a mix of those who do not have much to gain or cannot afford to take the test again. Our descriptive analysis of retaking behavior weakly confirms this intuition. Figure XX shows that among attempters who fail the Praxis the first time, those who opt to retake the exam but fail have lower earnings on average than those who do not retake the exam. Furthermore, while the likelihood of retaking the test is increasing in one's initial failing score (Figure 4), the relationship is entirely driven by factors related to socioeconomic status: holding a Pell grant, being Hispanic, and earnings at the time of the test.

F   Model of selection out of teaching

The point of this exercise is to illustrate that in the current equilibrium, there are teachers for whom (1) they dislike teaching relative to their next-best option ($z < 0$, in notation to follow), and (2) the large pay differential exactly offsets by their distaste for teaching, such that lowering the teaching salary by some epsilon amount would tip them into leaving to a lower-paying job. One might expect $z > 0$ for all teachers, i.e. "teaching for the love of it." But $z < 0$ says that there exist people who strictly prefer another option but teach because the pay differential is high enough.

Let every test-taker have a next-best job option which pays $d \in \mathbb{R}$ less than teaching does. Let every test-taker also have a relative preference for teaching over their next-best option given by z (in dollars). Assume that test-takers get a noisy signal of how much they prefer teaching over their next-best option, $\tilde{z} = z + \varepsilon$, and that they only realize their true preference for teaching z once they become teachers. Finally, assume that the joint distribution of (d, z) is continuous.

Selection into and out of teaching goes as follows. Those who are interested in teaching are those for whom $\tilde{z} \geq -d$; they believe that their enjoyment of teaching will outweigh the pay differential they may be foregoing. Those who stay teachers are those who eventually realize that $z \geq -d$. Leaving teachers are those for whom $\tilde{z} \geq -d$ but $z < -d$.

Test-takers are by definition interested in teaching, meaning they have $\tilde{z} \geq -d$. However, the assumption that teachers are earning pure rents in our framework is to say that $z > 0$ for all test-takers. In contrast, if there are teachers who would not teach if it were not for the compensating differential they are paid, then there should exist teachers for whom $z < 0$.

Our estimates allow us to test the null hypothesis that $z > 0$ for all test-takers. Firstly, our RD estimates tell us that the average pay differential among test-takers, $\bar{d}$, is \$17,000, identifying the average value of d in the joint distribution over the entire support of (z, d). Secondly, our event study estimates allow us to identify the average d for only the test-takers who become teachers but realize that $z < -d$: for leaving teachers, $\bar{d}_l = \$20,000$, similar in magnitude to $\bar{d}$. As such, by the continuity of the joint distribution, there must be teachers for whom $z = \bar{d}$ and hence $z < 0$.

The takeaway is that there are not only teachers who would not be willing to work for less, but that there are teachers who are not willing to work as teachers for less and would take an enormous pay cut.

G   Additional Figures
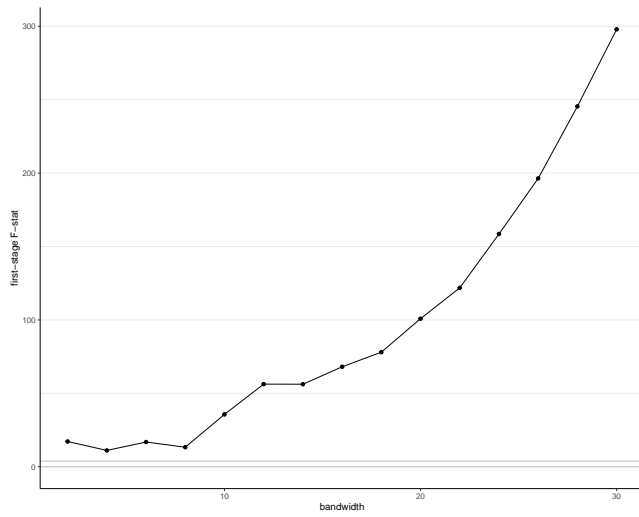
(a) Ever-passing the math test.
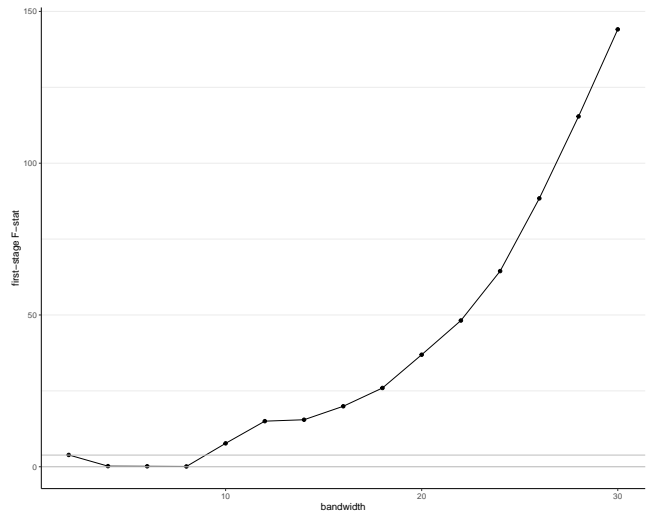
(b) Ever-passing all 3 tests.
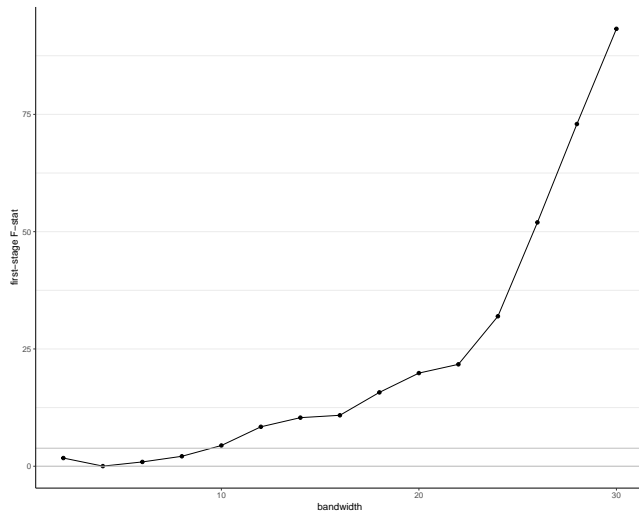
(c) Enrolling in an EPP within 3yrs.
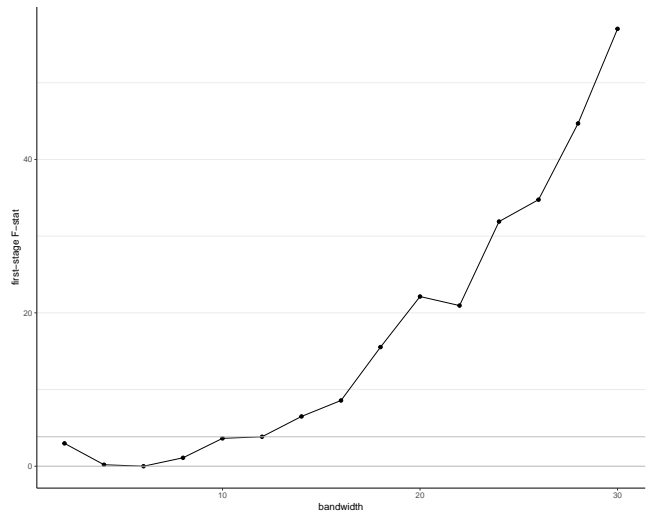
(d) Becoming certified within 4yrs.
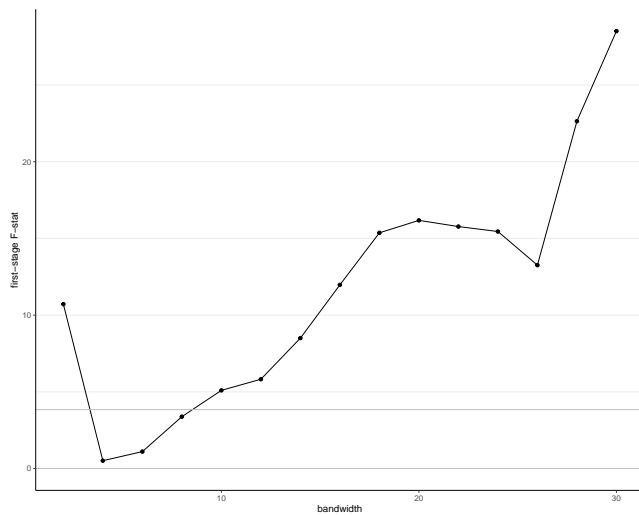
(e) Being a teacher in 4yrs.

Figure G.1: Sensitivity of first-stage RKD F-stats to bandwidth selection, for each stage along
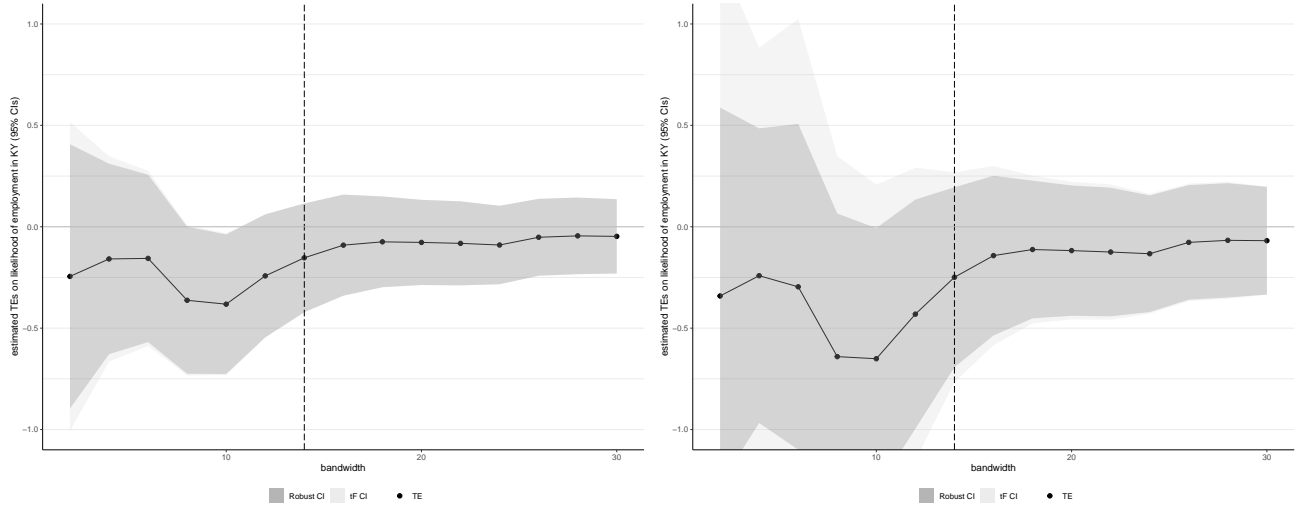
(a) Effect of ever-passing math on Pr(employed).   (b) Effect of ever-passing all 3 on Pr(employed).

Figure G.2: RD estimates of the effect on the likelihood of being employed (in the state of KY) of passing the Praxis, over bandwidths ranging from 2-30.
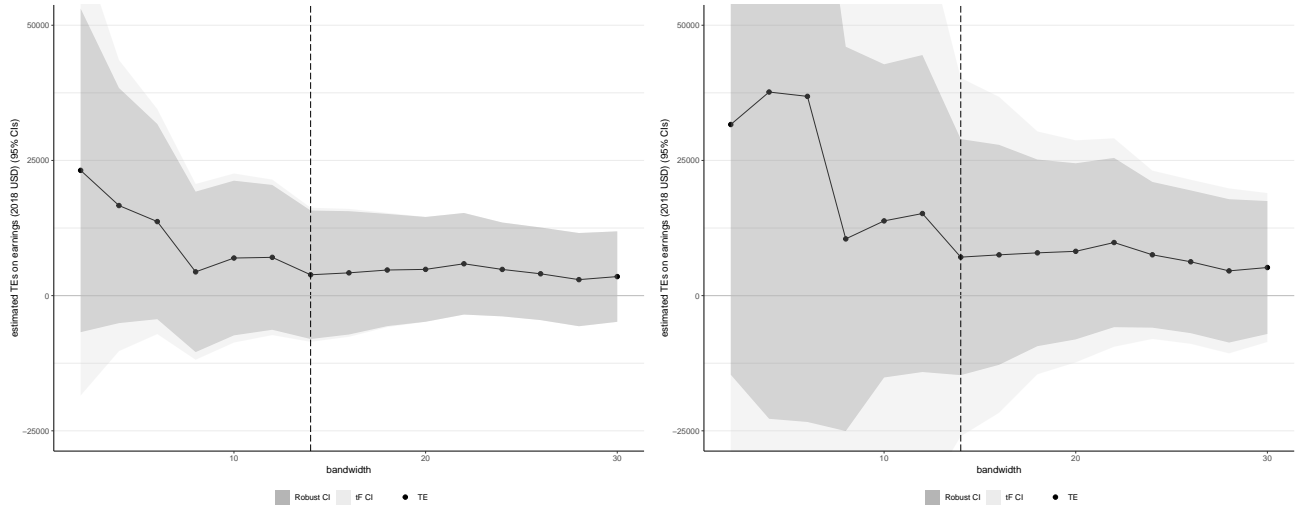
(a) Effect of ever-passing math on earnings.   (b) Effect of ever-passing all 3 on earnings.

Figure G.3: RD estimates of the effects on earnings in 4yrs of passing the Praxis, over bandwidths ranging from 2-30.
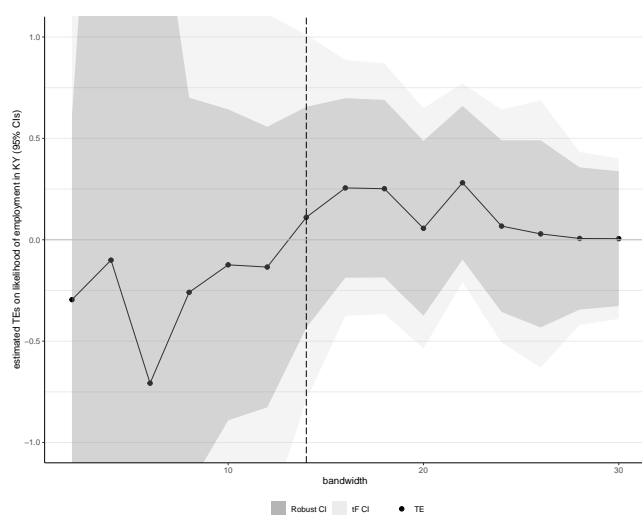
(a) Effect of becoming a teacher on Pr(employed).

Figure G.4: RKD estimates of the effect on the likelihood of being employed (in the state of KY) of reaching various stages of the teacher pipeline over bandwidths ranging from 2-30.

* Imagine you are offered the two jobs shown below. (1 of 3)

The differences between the jobs are highlighted in yellow. **The jobs are otherwise exactly the same**, even on characteristics not listed in the table.

Please review the jobs and indicate below whether you prefer Job A or Job B.

	Job A	Job B
Hours	Full-Time - 40 hours per week	Full-Time - 40 hours per week
Control Over Hours	Set your own schedule	Schedule set by manager
Option to Telecommute	Yes	Yes
Physical Demands	Heavy physical activity	Heavy physical activity
Pace	Relaxed	Relaxed
Independence	You can choose how you do your work	You can choose how you do your work
Paid Time Off (Vacation and Sick Days)	20 days	10 days
Working With Others	Team-based and evaluated on performance of team	Team-based and evaluated on performance of team
Training	You already have the skills for this job	You already have the skills for this job
Impact on Society	Occasional opportunities to make a positive impact on your community or society	Occasional opportunities to make a positive impact on your community or society
Pay	$56,500 per year	$62,800 per year

	Strongly Prefer Job A	Prefer Job A	Prefer Job B	Strongly Prefer Job B
Which job do you prefer?	○	○	○	○

Figure G.5: Screenshot of hypothetical job pair from survey to Kentucky teachers

H Additional Tables

Table H.1: Summary statistics comparing teachers to non-teachers

	In 2010			In 2018		
	Mean	S.D.	N	Mean	S.D.	N
Sample A: People who were not teachers in 2010						
Age	34	16	3341386	42	16	3341386
Female	0.5	0.5	3465894	0.5	0.5	3465894
White	0.65	0.48	3465894	0.65	0.48	3465894
Black	0.08	0.27	3465894	0.08	0.27	3465894
Other race	0.27	0.44	3465894	0.27	0.44	3465894
Hispanic	0.04	0.19	3465894	0.04	0.19	3465894
Highest degree earned since 2005			115152			321718
... Bachelor's	0.71			0.68		
... Master's	0.22			0.25		
... Doctorate	0.06			0.07		
Nr of quarters employed	3.41	1.02	1874444	3.44	1	2086198
Total yearly wages	33941.67	34416.44	1874444	36032.27	36652.57	2086198
Working as teacher	0	0	3465894	0.01	0.07	3465894
Sample B: People who were teachers in 2010						
Age	40.6	10.9	43350	48.6	10.9	43350
Female	0.78	0.41	43473	0.78	0.41	43473
White	0.95	0.21	43473	0.95	0.21	43473
Black	0.04	0.18	43473	0.04	0.18	43473
Other race	0.01	0.1	43473	0.01	0.1	43473
Hispanic	0.02	0.13	43473	0.02	0.13	43473
Highest degree earned since 2005			14198			18043
... Bachelor's	0.38			0.07		
... Master's	0.61			0.91		
... Doctorate	0.00			0.01		
Nr of quarters employed	3.97	0.22	43159	3.87	0.51	33084
Total yearly wages	57357.05	13495.48	43159	54479.75	22481.98	33084
Working as teacher	1	0	43473	0.55	0.5	43473

Note: Sample includes all individuals with non-zero earnings in the Kentucky UI data in 2010.

Table H.2: OLS estimates of cross-sectional differences in earnings between teachers and non-teachers among KY college graduates

| | *Dependent variable: Annual total wages (2018 USD)* | | | | | |
| | Sample: 18-60yo | | Sample: 18-60yo, working at least 4qtrs | | Sample 18-30yo, working at least 4qtrs | |
	(1)	(2)	(3)	(4)	(5)	(6)
Is teacher	4,977.318***	−6,187.064***	−1,649.273***	−7,252.242***	3,218.845***	2,545.662***
	(71.502)	(61.972)	(69.929)	(66.209)	(90.019)	(83.769)
Female		−7,643.454***		−8,566.723***		−4,744.887***
		(43.785)		(50.745)		(58.738)
Black		−4,958.967***		−6,057.636***		−3,967.194***
		(141.189)		(168.967)		(201.492)
White		463.919***		19.382		1,204.935***
		(114.653)		(138.385)		(162.909)
Hispanic		−2,796.032***		−3,140.671***		−1,512.377***
		(132.946)		(155.009)		(180.965)
Age (years)		589.780***		654.235***		2,337.656***
		(2.885)		(3.307)		(16.385)
Nr quarters employed		14,685.510***				
		(26.678)				
Holds at least a Master's		14,090.470***		15,369.000***		10,167.130***
		(49.858)		(56.726)		(83.056)
Mean	46591.01					
Observations	1,201,888	1,201,888	991,951	991,951	387,848	387,848
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.004	0.355	0.001	0.218	0.003	0.175

Note: Sample includes people 18-60yo with at least a BA earned in KY since 2005. *$p<0.1$; **$p<0.05$; ***$p<0.01$

Table H.3: Working Conditions in the U.S. in 2015, for Teachers and Comparison Demographic Groups

	Overall (1)	**Teachers** (2)	Women (3)	College-Educ. (4)
Mean hours per week	39.62	39.03	37.32	40.23
	(9.89)	(10.09)	(9.58)	(10.17)
Mean wage (in 2015 $)	30.30	25.89	25.99	38.00
	(36.55)	(15.02)	(27.95)	(30.10)
% with each working condition				
Sets own schedule	56.5	14.7	58.8	67.5
Telecommute	36.4	13.4	37.0	54.9
Moderate physical activity	38.4	88.2	38.1	36.7
Mostly sitting	42.9	5.8	49.4	57.9
Relaxed pace	29.8	13.9	32.6	35.5
Choose how do work	86.4	93.5	86.4	91.8
1-14 Paid Time Off (PTO) days	26.0	61.2	25.1	21.3
15+ PTO days	59.7	28.6	61.2	68.7
Team-based, evaluated on own	49.1	68.1	51.3	52.1
Work by self	32.4	26.0	35.1	33.5
Training opportunities	70.0	63.3	65.4	74.0
Frequent opp. to serve	34.5	61.8	39.5	39.0
Observations	1738	46	968	908

Table H.4: Regressions of Working Conditions on Demographic Variables (part 1)

	(1) Sets own sched.	(2) Telecommute	(3) Heavy Physical Activity	(4) Moderate P.A.	(5) Mostly sitting	(6) Relaxed pace	(7) Choose how do work
Teacher	-58.25	-42.98	1.17	55.92	-57.09	-21.62	1.46
	(8.44)	(8.02)	(6.49)	(8.53)	(8.40)	(8.02)	(6.03)
Female	3.30	-1.08	-8.01	-3.75	11.76	5.56	-0.50
	(2.49)	(2.36)	(1.91)	(2.51)	(2.47)	(2.36)	(1.78)
Nonwhite	-7.66	-6.86	0.57	2.85	-3.42	-0.18	-2.47
	(3.17)	(3.01)	(2.44)	(3.21)	(3.16)	(3.02)	(2.27)
HS or less	-28.35	-39.48	24.88	6.35	-31.23	-12.69	-12.53
	(3.00)	(2.85)	(2.30)	(3.03)	(2.98)	(2.85)	(2.14)
Some college	-17.05	-30.37	13.55	8.42	-21.97	-5.97	-4.17
	(3.07)	(2.92)	(2.36)	(3.10)	(3.06)	(2.92)	(2.19)
Under 35	-2.64	-6.86	2.66	-4.83	2.17	-21.58	12.82
	(4.78)	(4.54)	(3.68)	(4.83)	(4.76)	(4.55)	(3.42)
Aged 35-49	-4.82	-3.04	3.76	-14.20	10.44	-11.46	5.47
	(4.53)	(4.31)	(3.49)	(4.58)	(4.51)	(4.31)	(3.24)
Aged 50-61	-3.39	-3.26	1.53	-8.37	6.84	-9.28	8.36
	(4.63)	(4.40)	(3.56)	(4.68)	(4.61)	(4.40)	(3.31)
N	1,492	1,492	1,492	1,492	1,492	1,492	1,492

Table H.5: Regressions of Working Conditions on Demographic Variables (part 2)

	(1) No PTO	(2) 1-14 PTO Days	(3) 15+ PTO days	(4) Team Eval.	(5) Team-based, Own Eval.	(6) Work by self	(7) Training opps.	(8) Freq. opp. to serve
Teacher	0.16	41.94	-42.11	-7.22	16.27	-9.05	-10.05	21.56
	(5.94)	(7.82)	(8.58)	(6.93)	(8.89)	(8.35)	(7.94)	(8.38)
Female	-0.79	-1.36	2.14	-9.17	2.89	6.28	-11.39	9.39
	(1.75)	(2.30)	(2.53)	(2.04)	(2.62)	(2.46)	(2.34)	(2.47)
Nonwhite	-4.31	6.12	-1.81	3.95	0.49	-4.44	8.90	-2.66
	(2.23)	(2.94)	(3.22)	(2.60)	(3.34)	(3.14)	(2.98)	(3.15)
HS or less	6.07	15.00	-21.07	5.73	-5.28	-0.45	-12.54	-9.74
	(2.11)	(2.77)	(3.04)	(2.46)	(3.15)	(2.96)	(2.82)	(2.97)
Some college	3.33	9.38	-12.70	7.09	-4.33	-2.75	-6.16	-4.85
	(2.16)	(2.84)	(3.12)	(2.52)	(3.23)	(3.04)	(2.89)	(3.05)
Under 35	-6.91	7.20	-0.29	3.65	4.69	-8.34	19.57	-1.65
	(3.36)	(4.43)	(4.86)	(3.93)	(5.03)	(4.73)	(4.50)	(4.75)
Aged 35-49	-9.59	3.60	5.99	7.83	-5.89	-1.93	10.72	5.24
	(3.19)	(4.20)	(4.61)	(3.72)	(4.77)	(4.48)	(4.27)	(4.50)
Aged 50-61	-8.18	-1.47	9.65	3.95	-4.55	0.59	7.34	1.63
	(3.26)	(4.29)	(4.70)	(3.80)	(4.87)	(4.58)	(4.36)	(4.60)
N	1,492	1,492	1,492	1,492	1,492	1,492	1,492	1,492

Table H.6: Comparison of Working Conditions in Teaching to Below-Average Paying Jobs in Other Industries in 2015 (AWCS)

	Teacher (1)	Industry in AWCS						
		Ed/Health (2)	Prof (3)	Leisure (4)	Govt (5)	Finance (6)	Trade (7)	Info (8)
Next-best industry %	-	35.9	12.5	11.5	6.8	4.1	3.9	3.6
Mean hours per week	39.0	37.9	39.5	33.8	38.2	39.0	40.2	36.9
Mean hourly wage (in 2015)	25.9	18.1	18.8	10.2	21.8	20.2	17.1	18.4
% with each working condition								
Sets own schedule	14.7	48.1	54.7	69.0	52.4	71.1	45.2	67.9
Telecommute	13.4	23.3	37.5	12.7	29.9	46.3	16.1	52.3
Moderate physical demands	88.2	56.6	32.3	54.4	40.6	19.5	42.6	37.6
Mostly sitting	5.8	28.8	43.7	3.9	52.8	75.7	26.8	60.4
Relaxed pace	13.9	37.4	33.8	5.5	30.7	31.8	22.3	18.3
Choose how do work	93.5	91.6	83.6	87.0	92.8	88.8	77.5	97.2
1-14 PTO days	61.2	31.0	37.4	25.5	26.8	21.1	44.4	8.9
15+ PTO days	28.6	54.2	50.4	20.5	65.8	69.9	43.1	62.4
Team-based, evaluate own	68.1	52.1	40.6	63.4	54.0	40.4	45.0	54.3
Work by self	26.0	27.8	41.1	25.0	38.8	41.7	30.3	43.2
Training opportunities	63.3	67.0	67.6	50.4	73.8	65.3	68.7	71.3
Frequent opp. positive impact	61.8	51.2	20.8	29.4	51.1	38.6	24.1	21.9
N	46	271	169	30	96	100	188	25