

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

Carolyn Tsao[†]

November 8, 2024

Updated frequently — [\[click here for latest version\]](#)

Abstract

This paper quantifies the extent to which teachers earn rents, explicitly taking both pay and non-wage job attributes into account. 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 gaps to obtain an estimate of the teaching rent. Employing a fuzzy regression kink design that leverages variation in who becomes certified to teach in Kentucky, I show that teaching pays a premium of \$18–21,000/yr (33–40% of the teaching salary) more than teachers' next-best options. A similar pay gap arises from event studies around teacher exits. Importantly however, I also find that working conditions in teaching are relatively poor. Fielding a stated preference experiment to Kentucky teachers, I estimate that teachers are willing to pay 29–35% of their salary to switch to their next-best job, solely for better working conditions. My results indicate that teaching offers negligible rents; instead, teaching pays a large premium that mostly functions as a compensating differential. Extending my point estimates to other states suggests that over \$145 billion is being spent annually to compensate teachers for poor working conditions in the US, consistent with reports of teacher shortages today.

*I am indebted to David Lee, Alex Mas, and Ilyana Kuziemko for their guidance and unwavering support throughout this project. This work also would not have been possible without the invaluable support, dedication, and data expertise of Angie Tombari and the entire KYSTATS team. I thank Lukas Althoff, Zach Bleemer, Leah Boustan, Chris Campos, Janet Currie, Oren Danieli, Tim Ederer, Jen Jennings, Adam Kapor, Karl Schulze, Garima Sharma, Sharada Dharmasankar, Jesse Silbert, Christiane Szerman, Camille Terrier, Neil Thakral, Linh Tô, and participants of the Industrial Relations Seminar at Princeton, the Education Research Seminar at Princeton, and the CESifo/ifo Junior Workshop on Economics of Education for their insights and suggestions. I am grateful for the financial support of the Industrial Relations Section at Princeton, the Prize Fellowship in Social Sciences at Princeton, the Program for Research on Inequality at Princeton, and the Social Sciences and Humanities Council of Canada (SSHRC). Disclaimer: All errors are my own. The views and opinions expressed in this paper are my own and do not necessarily reflect the views of KYSTATS.

[†]Department of Economics, Princeton University. ctsao@princeton.edu

1 Introduction

Teacher salaries account for over half of public school spending in the US, totaling over \$500 billion in 2020–21. Given that state budgets are constrained, the opportunity cost of “over-compensating” teachers—that is, paying teachers *rents*—is high. However, there are reasons both to believe and to doubt that teachers earn rents. On one hand, public school teaching jobs are associated with pensions, job stability, and strong unions, which existing studies find may raise teacher compensation without raising student achievement (e.g., [Hoxby \[1996\]](#)). On the other hand, widespread reports of “teacher shortages” [[Sutcher et al., 2016](#)] and evidence that teachers earn less than the average college-educated worker [[Allegretto and Mishel, 2018](#)] suggest that teachers may not be capturing rents at all.¹ Consistent with this uncertainty, existing studies find mixed evidence of rent extraction in teaching.²

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 compensation intended as a *compensating differential* could be misinterpreted as a rent. Recent evidence suggests that teaching may be one such occupation, with over 50% of teacher strikes in the past decade demanding better working conditions [[Lyon et al., 2024](#)] amid increasing reports of stress, burnout, and student behavioral issues on the job.³

This paper quantifies the extent to which teachers earn rents, using a new approach that explicitly takes both pay and the value of working conditions into account. My approach is to compare 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: unobserved selection into

¹For example, see Eliza Fawcett and Jacey Fortin, “How Bad Is the Teacher Shortage? Depends Where You Live,” *The New York Times*, June 21, 2023; Moriah Balingit, “Teacher shortages have gotten worse. Here’s how schools are coping,” *The Washington Post*, August 24, 2023; Ben Chapman, “Teacher, Staff Shortages Reported by About Half of Schools, Survey Finds,” *The Wall Street Journal*, September 27, 2022.

²For example, see: [Podgursky and Tongrut \[2006\]](#), [Taylor \[2008\]](#), [Lovenheim \[2009\]](#), [Richwine and Biggs \[2011\]](#), [Cowen and Strunk \[2015\]](#), [Allegretto and Mishel \[2018\]](#).

³For example, see Sejla Rizvic, “Teachers, Facing Increased Level of Stress, Are Burned Out,” *The New York Times*, March 13, 2023; Bernard Mokam, “Rising Discipline Problems in Schools,” *The New York Times*, March 20, 2024.

jobs and a lack of data on teachers’ job switches and working conditions. 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. It is even more difficult to acquire the data needed to estimate the gap in working conditions, since 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 working conditions.⁴ Indeed, the first dataset to systematically document differences in working conditions across U.S. workers, the American Working Conditions Survey, only came into existence in 2015 and contains only a handful of teachers [Maestas et al., 2023].

In this paper, I address all of the challenges mentioned above. I estimate the pay gap using two quasi-experimental designs and a novel panel dataset covering the near-universe of workers in the state of Kentucky. I estimate the gap in the value of working conditions by surveying a representative sample of Kentucky teachers about their job attributes and eliciting their willingness to pay for each attribute through a series of stated preference experiments. My results provide the most comprehensive estimate to date of the teaching rent in a U.S. context.

In the first half of the paper, 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 design 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 panel dataset constructed by linking five administrative datasets from the state of Kentucky.

In the fuzzy RK, I circumvent selection into teaching by leveraging quasi-experimental variation in the process that determines who obtains teacher certification in Kentucky. 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, consistent with the score being used only 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 plausibly exogenous sources of variation in who becomes a teacher. 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-

⁴The most closely related study is Johnston [2023], who conducts a discrete choice experiment to estimate teachers’ preferences for school-specific attributes such as class size, merit pay, student poverty, along with some general attributes such as commute time.

random variation in how likely individuals are 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 the teaching job offers a large pay premium compared to prospective teachers’ next-best job options. From the RK design, 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.⁵ 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.

In the second half of the paper, I field a survey to a representative sample of teachers from Kentucky to elicit information about, and estimate their valuations of, the working conditions in teaching. To ensure that my estimates can be benchmarked to willingness to pay estimates in the literature, I replicate the survey instrument from the American Working Conditions Survey. The first part of the survey asks workers about nine key job attributes, 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.⁶ The stated-preference experiments 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 other jobs. Compared to the job of the average college graduate in the AWCS, teaching is significantly less likely to offer all of the working conditions in the survey, except for the opportunity to make a positive impact on one’s community or society.⁷ Strikingly, teaching is also less likely to offer nearly all desirable conditions compared to teachers’ “next-best jobs”—that is, lower-paying jobs in the education and healthcare

⁵This 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. Even with the most conservative standard errors we are able to rule out large pay differences in the opposite direction.

⁶The nine attributes are schedule flexibility, telecommuting opportunities, physical demands, pace of work, autonomy, paid time off, working with others, job training opportunities, and impact on society.

⁷Teaching is less likely to offer: 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%). However, teaching is more likely to offer is the opportunity to make a positive impact on one’s community or society (64% vs. 39%).

sectors in the AWCS, as identified through the earlier RK and event study analysis.

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. I obtain this estimate by identifying the job attributes that are more likely to be offered by teachers’ next-best jobs than by teaching, and estimating teachers’ total willingness to pay for this set of conditions. Based on the results from the pay gap analysis, I use lower-paying (below median pay) jobs in the education and healthcare sectors in the AWCS as a proxy for teachers’ next-best jobs. Using the choice data from the experiments and a simple model of job choice, I estimate teachers’ willingness to pay for each job attribute. Finally, I transform the estimates to obtain the average teacher’s total willingness to pay for all the attributes teachers’ next-best jobs are more likely to offer than teaching.

Overall, my findings suggest that Kentucky teachers earn negligible rents. Instead, by summing the estimated gaps, I find that teaching offers a large pay premium that is almost entirely compensation for undesirable working conditions.⁸ 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 teaching conditions in addition to pay.

This paper contributes to three strands of literature. The first strand investigates whether public sector workers earn rents, with a focus on public-private wage differentials (e.g., [Freeman \[1984\]](#), [Krueger \[1988\]](#), [Moulton \[1990\]](#), [Terrell \[1993\]](#), [Dustmann and Van Soest \[1998\]](#), [Gittleman and Pierce \[2012\]](#), [Brueckner and Neumark \[2014\]](#)). Among the studies on teachers, many examine how teacher unions impact salaries and student outcomes, especially through collective bargaining (e.g., [Hoxby \[1996\]](#), [Lovenheim \[2009\]](#), [Strunk \[2011\]](#), [Brunner and Squires \[2013\]](#), [Cowen and Strunk \[2015\]](#), [Matsudaira and Patterson \[2017\]](#), [Lovenheim and Willén \[2019\]](#), [Brunner and Ju \[2019\]](#), [Cook et al. \[2021\]](#), [Shi and Singleton \[2023\]](#)). Relative to this literature, I demonstrate a new approach to establishing whether public sector workers earn rents that does not require data on unions or bargaining, which can be hard to find. The second strand examines the relationship between teachers’ alternative job opportunities and selection into teaching (e.g., [Loeb and Page \[2000\]](#), [Corcoran et al. \[2004\]](#), [Hoxby and Leigh \[2004\]](#), [Bacolod \[2007\]](#), [Nagler et al. \[2020\]](#)), studying how factors like changes in pay, unionization, or labor

⁸By my most conservative estimates, 4% of the teaching salary is a rent, but the point estimate is indistinguishable from zero. In contrast, 29–35% of the teaching salary is a compensating differential.

market opportunities for women explain who works as a teacher over time.⁹ Many of these studies leverage specific changes to teaching institutions or broad changes in the labor market for variation in who becomes a teacher. Relative to the literature, I use new experimental variation to show that non-wage attributes of teachers’ alternative jobs play an important role in whether one works as a teacher as well. 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 (e.g., [Stinebrickner \[2002\]](#), [Scafidì et al. \[2006\]](#), [Champion et al. \[2011\]](#), [Goldhaber et al. \[2022\]](#)). I contribute to the literature new estimates of the pay gap using variation in who exits teaching *and* new variation in who becomes a teacher.

In terms of methodology, my regression kink design is related to existing studies that use regression discontinuity designs leveraging test score cutoffs to estimate returns to certification and degree completion in the US (e.g., [Zimmerman \[2014\]](#), [Clark and Martorell \[2014\]](#), [Jepsen et al. \[2016, 2017\]](#), [Goodman et al. \[2017\]](#)).¹⁰ My setting differs in that I find evidence of a kink and no jump at the cutoff, and I estimate the effects of obtaining a specific job.^{11,12} The survey experiment I use is related to a growing literature that uses choice experiments to estimate workers’ preferences for non-wage amenities (e.g., [Mas and Pallais \[2017\]](#), [Wiswall and Zafar \[2018\]](#), [Drake et al. \[2022\]](#), [Maestas et al. \[2023\]](#), [Dube et al. \[2024\]](#)).¹³ Few papers have used this methodology to estimate teachers’ preferences, however, except for [Johnston \[2023\]](#).¹⁴

The rest of the paper proceeds as follows. Section 2 describes how I estimate the pay gap, from data and methods to results. Similarly, Section 3 describes how I estimate the gap in the value of working conditions. Section 4 discusses the implications of my estimates, as well as how the estimates can be reconciled with previous findings. Section 5 concludes.

⁹While I focus on the extensive margin of selection into teaching, a vast related literature studies the determinants of which schools teachers work at *within* a school system, such as: local school competition ([Jackson \[2012\]](#), [Bruhn et al. \[2022\]](#)), student performance ([Feng \[2010\]](#), [Boyd et al. \[2005\]](#)), accountability reforms ([Kraft et al. \[2020\]](#)), and mentoring ([Smith and Ingersoll \[2004\]](#)). A large body of work also examines characteristics of the teaching job that predict who teacher exits, such as: student achievement ([Hanushek et al. \[2004\]](#)), teacher “voice” ([García et al. \[2022\]](#)), fertility decisions ([Stinebrickner \[1998\]](#)), and perceptions of working conditions ([Ladd \[2011\]](#)).

¹⁰These studies are part of a larger literature evaluating the labor market returns to post-secondary degrees or high school completion using various methods: see for example [Heckman et al. \[2011\]](#) on returns to the GED.

¹¹Because the kink is driven by retaking behavior, the variation I use is most similar to that used by [Cellini et al. \[2010\]](#), who use repeated referenda to study the effect of school facility investments on schools and localities using a dynamic regression discontinuity design.

¹²A few studies similarly use a regression discontinuity design to study the returns to becoming a teacher outside of the US, including [Saavedra et al. \[2022\]](#) in Colombia and [Barton et al. \[2017\]](#) in Kenya.

¹³These papers are a subset of an established literature that studies the value of non-wage amenities using other methods, such as from data on worker transitions between firms (e.g., [Sorkin \[2018\]](#)).

¹⁴Many papers use other methods to study teachers’ preferences and their role in teacher sorting across schools, including, but not limited to: [Boyd et al. \[2013\]](#), [Rothstein \[2015\]](#), [Biasi et al. \[2021\]](#), [Bates et al. \[2022\]](#), [Brown and Andrabi \[2020\]](#), [Bobba et al. \[2021\]](#).

2 Estimating the Pay Gap

2.1 Identification Challenge

In this section, the parameter of interest is 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., to earn a teacher certification), is endogenous to earnings potential.

I circumvent this challenge in two ways. First, I exploit 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, I leverage job switches induced by exits from teaching to other jobs. This approach allows me to estimate the initial gap in pay that arises after a teacher moves—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 my identification strategy, as well as the data that I use, before detailing the empirical strategies and results.

2.2 Key Features of the Teacher Certification Process

In most U.S. states, the traditional path to becoming eligible to teach in public schools is a multi-year process colloquially known as the “teacher pipeline.”¹⁵ 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 earn 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 combine with the last two years of their four-year college degree.¹⁶

¹⁵Over 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.

¹⁶Some 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.

The majority of state-approved EPPs use Praxis exams as EPP entry and/or exit exams.¹⁷ “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 (“Core Academic Skills for Educators”) 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 to test individuals for admission to, or graduation from, teaching preparation programs.¹⁸ However, states vary in which Praxis exams they require at what stages of the pipeline and in what constitutes a passing score. State departments of education and ETS are transparent about these differences.¹⁹ In Kentucky, 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, 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. Most test-takers who attempt one of the three exams choose 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.^{20, 21}

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

¹⁷As 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.

¹⁸This 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.

¹⁹See 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.

²⁰Magoosh, 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.

²¹A sizeable minority of programs 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.

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.

2.3 Matched Employer-Employee Data

My main administrative data is comprised of five statewide datasets from Kentucky linked using SSNs. I start with data on Praxis test-takers and track their progress through the teacher pipeline with data 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, I 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, I 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, I again merge these data with the Kentucky UI database for all workers between the ages of 18 to 60, which contains 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. I 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.²² The Kentucky UI wage records have the standard UI data limitation in that they do not separately identify individuals who are unemployed from individuals who might be working outside of the state; as a result, I refer to individuals in either category as “non-employed.”

Finally, to obtain details on public school employment, I 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 me to observe when individuals move between schools or districts, or change occupations within public schools.

The final sample includes the union of all individuals who ever attempted the Praxis CASE 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, I obtain age, gender, race, and ethnicity from whichever database the person first appears in, with near full coverage.

Because the linked administrative data does not contain information on occupations outside of the public school system, I supplement my analyses of the 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. I am able to observe the second occupation for nearly all of these individuals.

Table 1 shows sample means describing all individuals who attempted the Praxis CASE for the first time between 2013 and 2017 in Kentucky. I focus on the start date of 2013 since the Praxis CASE was last revised then, and on the end date of 2017 to ensure that the data can capture employment and earnings outcomes four years after attempting the exam. Panel A shows key demographic characteristics. First-time test-takers are on average in their third year of college and are mostly white women, in similar proportions to the demographic composition of teachers in the state.²³ One-fifth hold a Pell grant—lower than the state average—and two-thirds are working in the year prior to taking the exam.

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

²²Details in the appendix.

²³Although 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.

Table 1: Summary statistics on main Praxis sample

	Mean	S.D.	N
<i>Panel A: Sample characteristics at first attempt</i>			
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
<i>Panel B: Praxis performance and retaking</i>			
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...			
% 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

Note: Summary statistics on Praxis first-attempters between 2013 and 2017.

majority (80%) of test-takers eventually pass the entire Praxis CASE. However, these passes are not all obtained on the first attempt: for the least-passed component, math, just under two-thirds of 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. Only a third pass on their first retake. The majority of first-time attempters retake the exam no more than three times.

Finally, Figure 1 shows that there is significant “leakage” of individuals from the pipeline over time. Only 53% of 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.

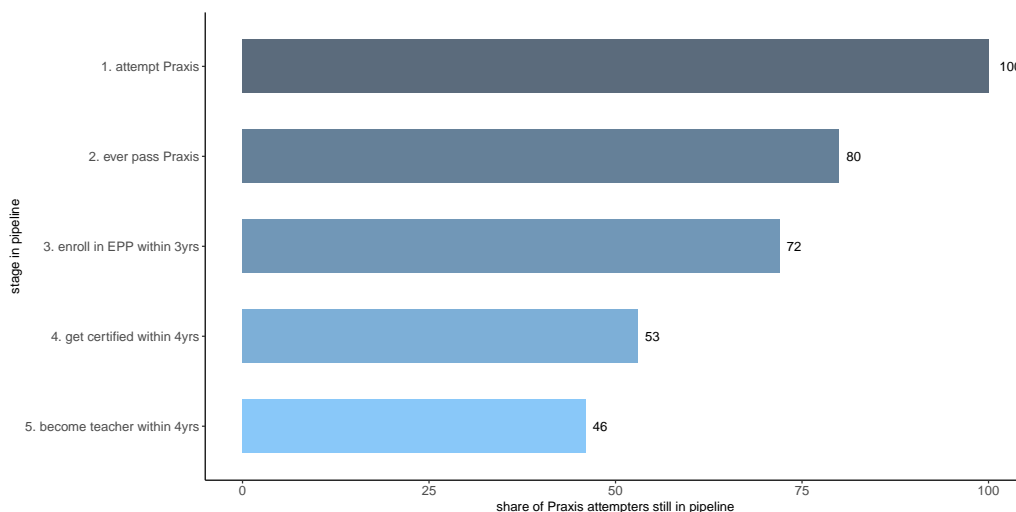
2.4 Empirical Strategy I: Regression Kink Design

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

2.4.1 Identification

The identification argument proceeds as follows. Assume that test-takers cannot precisely manipulate their test scores, and that there is an “encouragement-to-retake effect” of scoring near the cutoff that is increasing in one’s initial score. Under the first assumption, those who

Figure 1: Leakage from the teacher pipeline



Note: Share of first-time Praxis test-takers who persist to each stage of the pipeline. Sample includes individuals who attempt the Praxis for the first time from 2013 to 2022.

score near the cutoff on their first attempt are comparable in every way except that (1) those who score above the cutoff can continue directly on to certification, while (2) those who score below the cutoff can only continue on if they retake and pass the exam. Under the second assumption, those who score just below the cutoff will be more likely to retake the exam than those who score further below the cutoff.

If these assumptions hold, then 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. The interaction between whether one passes on their first attempt and how far they score from the cutoff on their first attempt would therefore be a valid instrument for whether one gets certified.

I demonstrate that both assumptions hold and that the instrument is valid in Section 2.4.4.

2.4.2 Why an RK and not an RD?

It is not immediately obvious why it is appropriate to use an RK and not appropriate to use an RD in this setting. The underlying principle of the fuzzy RD design is that as long as there is randomization in the running variable around the cutoff, then even with imperfect compliance, the running variable can be used as an instrument for treatment (i.e., becoming a teacher) [Hahn et al., 2001]. In theory, the same principle can be applied to our setting. Each attempt at the exam is an RD, where fuzziness is introduced by (i) retaking among those who score

below the cutoff, and (ii) attrition along the pipeline in who becomes as teacher, as shown in Figure 1. Existing work uses precisely this approach: for example, Jepsen et al. [2016] 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, my setting deviates from the classic fuzzy RD setting in that I find a kink and *no discontinuity* in the likelihood of certification at the cutoff. Applying a statistical result from Dong [2018], I formally demonstrate that the kink combined with the lack of discontinuity allows me to identify the treatment effect of certification on earnings. The proofs in Appendix D 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, I use a fuzzy RK in my main analysis.

In the case of the RK, the additional identifying assumptions needed 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 shown in Appendix E, 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 is zero if and only if retaking behavior is determined by both the “encouragement” and “intrinsic motivation” effects described above. I find supporting evidence that the “encouragement” effect is at play: 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 everyone below the cutoff retaking the exam, it must be that the “intrinsic motivation” factor is at play as well.

2.4.3 Estimating equations

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

$$\begin{aligned} \text{Structural eqn.} \quad Y_{i,t+4} &= \beta_0 + \beta_1 C_{i,t+4} + \beta_2 S_{it} + \epsilon_{it} \\ \text{First stage.} \quad C_{i,t+4} &= \alpha_0 + \alpha_1 \underbrace{(D_{it} \times S_{1i})}_{Z_{it}} + \alpha_2 S_{it} + \eta_{it}, \end{aligned} \tag{1}$$

where $C_{i,t+4}$ is an indicator for whether one is certified within 4 years of their first attempt at the Praxis CASE, S_{it} is their first attempt score on the math portion of the Praxis CASE, and $D_{it} = \mathbb{I}(S_{1i} \geq 0)$. The main variable of interest is β_1/α_1 , a weighted average treatment effect of becoming certified on the outcome $Y_{i,t+4}$. The main outcomes of interest, $Y_{i,t+4}$, are earnings four years after the first test attempt, and whether one is employed four years after the first test attempt.

LATE interpretation. Like the RD estimand, the RK estimand can be interpreted as the weighted average treatment effect across the distribution of individual types [DiNardo and Lee, 2011]. Each estimand uses different weights, suggesting different interpretations.

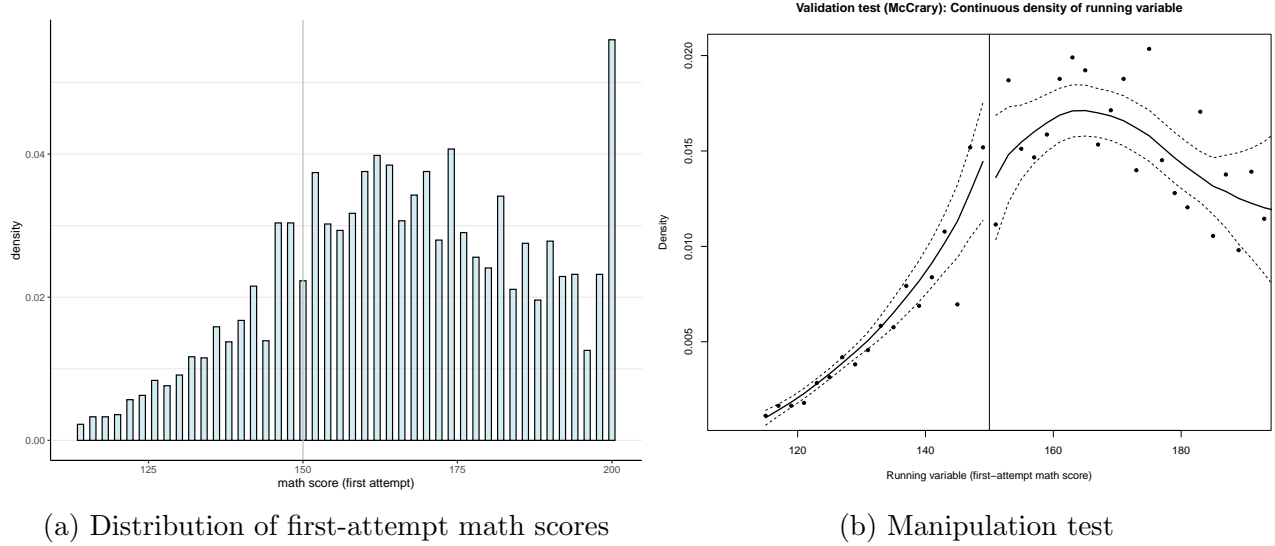
To demonstrate the difference between the RK and RD weights, I turn to the alternative interpretation of the 2SLS estimator as the “local average treatment effect” (LATE). The LATE is the average treatment effect for “compliers,” those who let the instrument dictate their receipt of the treatment.

First, consider an 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 an 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 to 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? I therefore borrow the classic concepts of compliers/always-takers/never-takers/defiers and re-cast the definitions to account for retaking behavior (Appendix F). Combined with the identification proofs in Appendix D, I find that the RK estimator is a weighted average treatment effect for “re-taking compliers”: individuals whose choice to retake is respon-

sive to their first-attempt score.

Figure 2: Tests for manipulation of first-attempt math scores



Note: Panel (a) shows a histogram of first-attempt math scores for the sample that passed reading and writing on the first attempt. Panel (b) conducts the test in [McCrary \[2008\]](#) and shows fitted polynomials to the right and left of the test score cutoff on the same sample used to plot the histogram in Panel (a).

2.4.4 Evaluating the quasi-experiment

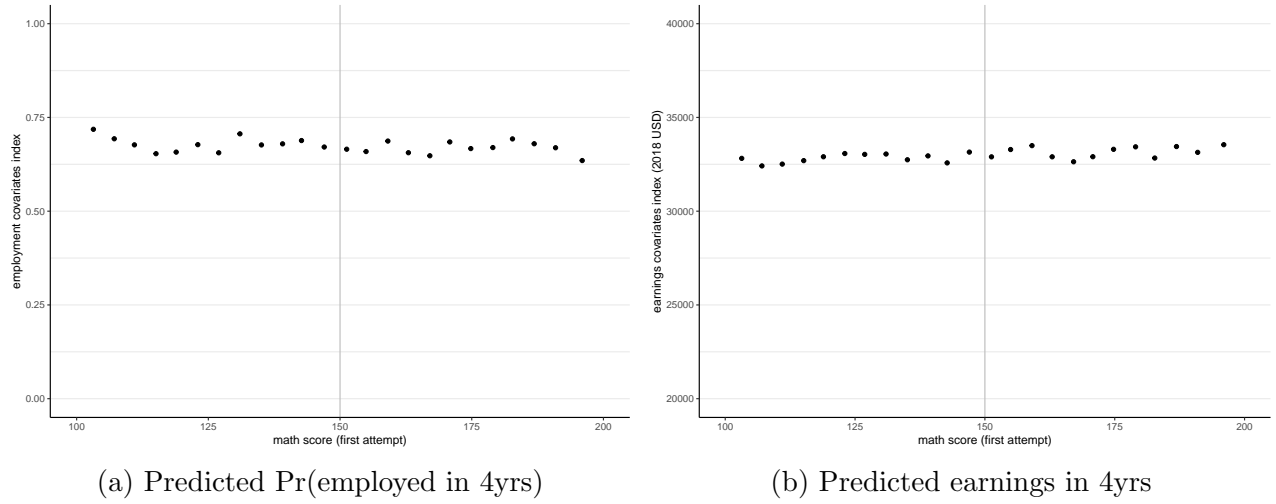
My empirical strategy focuses on individuals who score near the passing threshold on the Praxis CASE and treats whether they pass or fail as a near experiment. This strategy requires that scores are as good as randomly assigned near the passing threshold. In this section I test these assumptions.

Smoothness in the running variable. A key assumption for valid inference in the RK design is that the density of the running variable—the first-attempt math score—is smooth through the cutoff. A kink or a jump in the density of the sample at the cutoff would suggest endogenous sorting of individuals around the passing cutoff, which would invalidate the design. Figure 2 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.²⁴ There appears to be no jump or kink in the distribution at the cutoff.²⁵

²⁴Each bin represents a single score; gaps arise between the bars because only even number scores are possible on all components of the Praxis CASE.

²⁵Because 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.

Figure 3: Tests of selection around the cutoff, using covariates index



Note: Each panel presents binscatters for an “index” variable that is predicted using only covariates at “base-line,” i.e. in the year of or one year before attempting the Praxis for the first time. The index plotted in Panel (a) is the predicted share of individuals who are employed four years after the test based on the following covariates: gender, race/ethnicity, age, a dummy for whether the person was enrolled in post-secondary at the time of the test, a dummy for whether the person was enrolled in public post-secondary at the time of the test, a dummy for whether the person held a Pell grant at the time of the test, a dummy for whether the individual already possessed a post-secondary degree at the time of the test, earnings one year before the first attempt, a dummy for whether the person was employed one year before the attempt, dummies for the industry the person worked in one year before the attempt, and the year and month of the test. The index plotted in Panel (b) is the predicted earnings four years after the first attempt using the same set of baseline covariates.

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. I check this implication in three ways: (1) by visually examining the distributions 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 3), and (3) by conducting paired *t*-tests comparing differences in baseline characteristics of individuals who score within 10–15 points of the cutoff. All three tests suggest that there is balance on baseline characteristics.

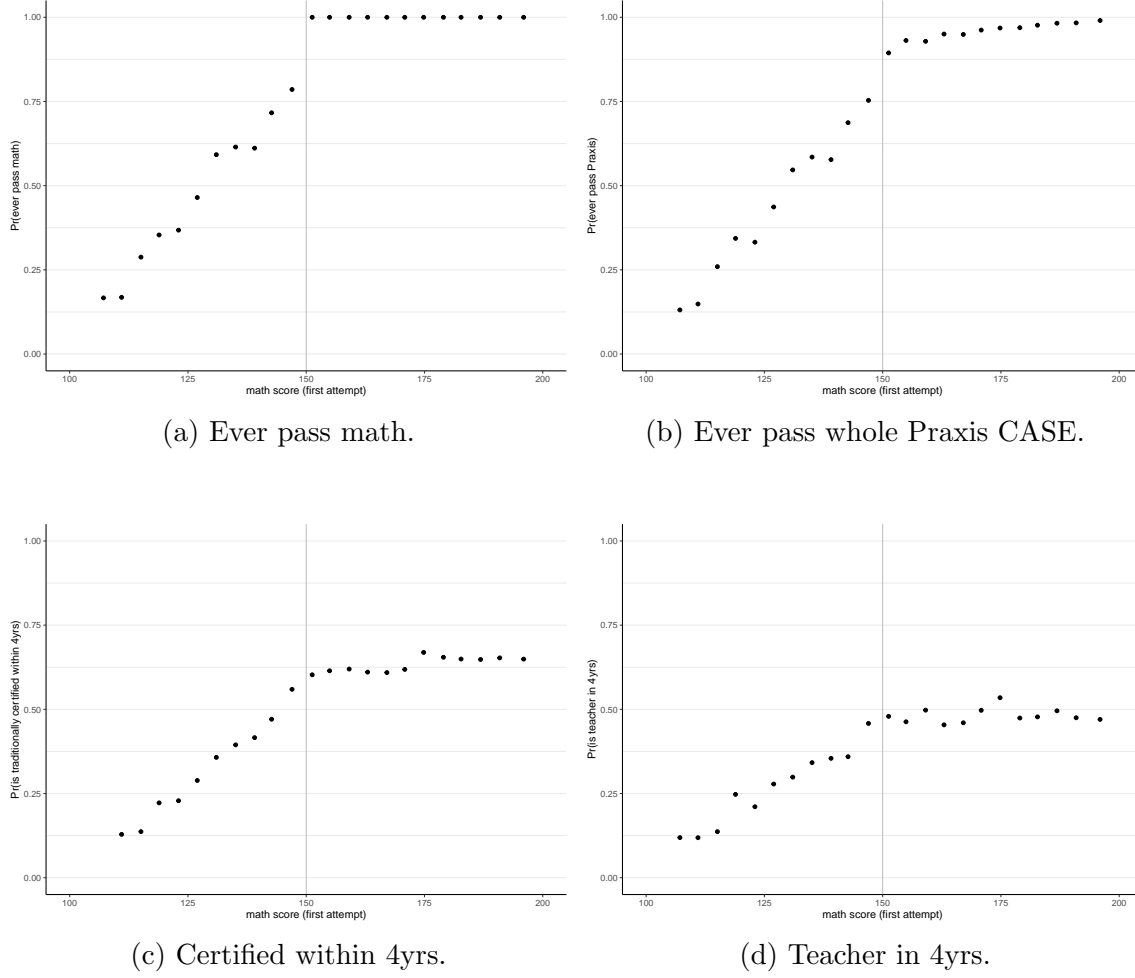
2.4.5 Evaluating instrument validity

I argue that the instrument is both relevant and excluded.

Relevance is satisfied by the pattern repeated in Figure 4: 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 5 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. I 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.

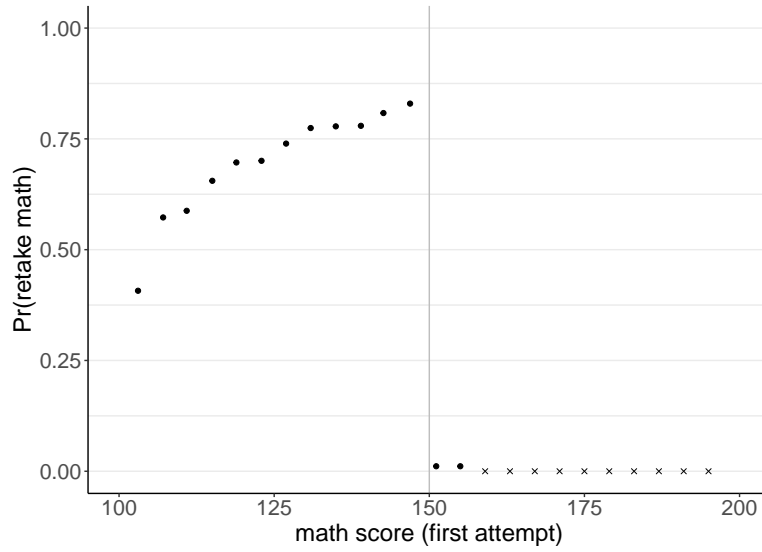
For the exclusion restriction to be satisfied, it must be the case that the first attempt score cannot affect earnings through any channel other than through its effect on the likelihood one gets certified. Perhaps the most compelling story that would violate the exclusion restriction would be if those 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 seems unlikely since test-takers need not wait long to retake the test (the modal wait time is 28 days), and the entry Praxis exam tests middle-school math and reading skills. I therefore conclude that it is reasonable to assume that the exclusion restriction holds.

Figure 4: First stage relationships



Note: Each panel shows the first stage relationship between the running variable (the first attempt math score) and the likelihood that individuals reach various stages along the teacher pipeline. Panel (a) shows the share of individuals who pass the math portion of the Praxis CASE by 2022, plotted over bins of the first-attempt math scores. Similarly, Panel (b) shows the share of individuals who pass all of math, reading, and writing on the Praxis CASE by 2022; Panel (c) shows the share of individuals who obtain their teacher certification within 4 years of their first attempt, and Panel (d) shows the share of individuals who are observed working as a teacher in Kentucky public schools in the fourth year after their first attempt.

Figure 5: Retaking behavior



Note: This figure is a 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. The increasing rate of retaking approaching the cutoff from the left, and the discontinuity in retaking after the cutoff, drives the kink.

2.5 Regression Kink Estimates of the Pay Gap

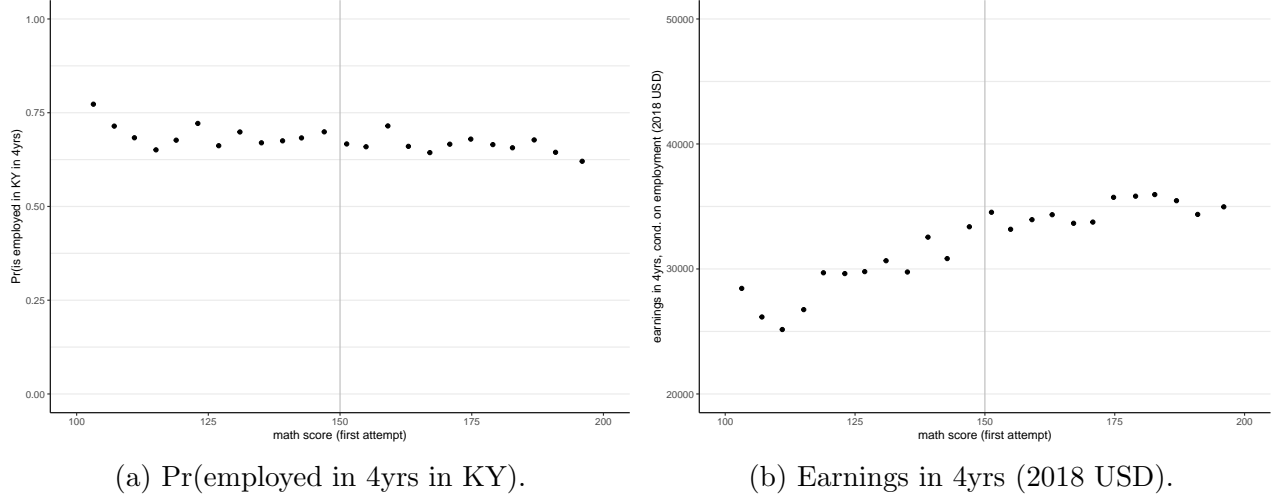
2.5.1 First-stage and reduced form plots

Figure 4 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. The main treatment of interest is getting certified, but for completeness, I show the progression of first-stage relationships for various outcomes along the pipeline.

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 A.1).

I also show graphical evidence of the “reduced-form” relationship between the first-attempt test scores and the outcomes of interest—earnings and attachment to the labor force four years after attempting the Praxis for the first time—in Figure 6. There appears to be no discontinuity

Figure 6: Reduced form relationships



Note: Binscatters of the 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.

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.

2.5.2 Treatment effect estimates

I now present my main estimates. I maintain a hands-above-the-table approach when it comes to bandwidth selection: I 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. I 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 7.

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

	Outcome:		Outcome:	
	Pr(Employed in 4yrs)		Earnings in 4yrs, cond. on emp.	
	First Stage (1)	Struct. Model (2)	First Stage (3)	Struct. Model (4)
<i>Treatment A: Passing the entire Praxis</i>				
RD estimate (conventional std. error)	0.12 (0.02)	-0.25 (0.23)		
VtF confidence interval		[-.7, .2]		
<i>Treatment B: Becoming Certified</i>				
RK estimate (conventional std. error)			-0.010 (0.004)	21,685.2 (11,380.9)
VtF confidence interval				[-621.4, 43,991.8]
<i>Treatment C: Becoming a Teacher</i>				
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 [Lee et al. \[2024\]](#).

I begin by showing the effect of passing the Praxis CASE and getting certified on the likelihood of employment. Leveraging the discontinuity in the pass rate at the cutoff, Table 2, Panel A and Appendix Figure A.2 show that passing the entire Praxis has no effect on the likelihood of short-run employment. Similarly, leveraging the kink in the likelihood of certification at the cutoff, Figure A.4 shows that getting certified has no effect on the likelihood of short-run employment. Given this null effect, in my analysis of treatment effects on earnings, I restrict the sample to those who are employed (i.e., those who have non-zero earnings).

Figure 7 shows the RK treatment effect estimates of becoming certified and becoming a teacher on earnings, across a range of bandwidths. Panels (a) and (b) show that the first stages begin to gain statistical power around a bandwidth of 12 to 14, coinciding with the MSE-optimal bandwidth recommended by [\[Calonico et al., 2020\]](#), indicated by a vertical dotted line at 14.²⁶

²⁶Consequently, the estimates shown in Table 2, Panels B and C are for a fixed bandwidth of fourteen, at

The main RK estimates in Panels (c) and (d) suggest there is a large, positive effect of getting certified and becoming a teacher on earnings, around \$18,000–\$21,000. The estimates gain precision and are relatively stable across all bandwidths in Panel (d), suggesting that becoming a teacher is the main mechanism underlying the overall effect on earnings.

2.5.3 Interpretation

The differences between the RD and RK estimates are intuitive given the context. The skills developed along the pipeline should 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.

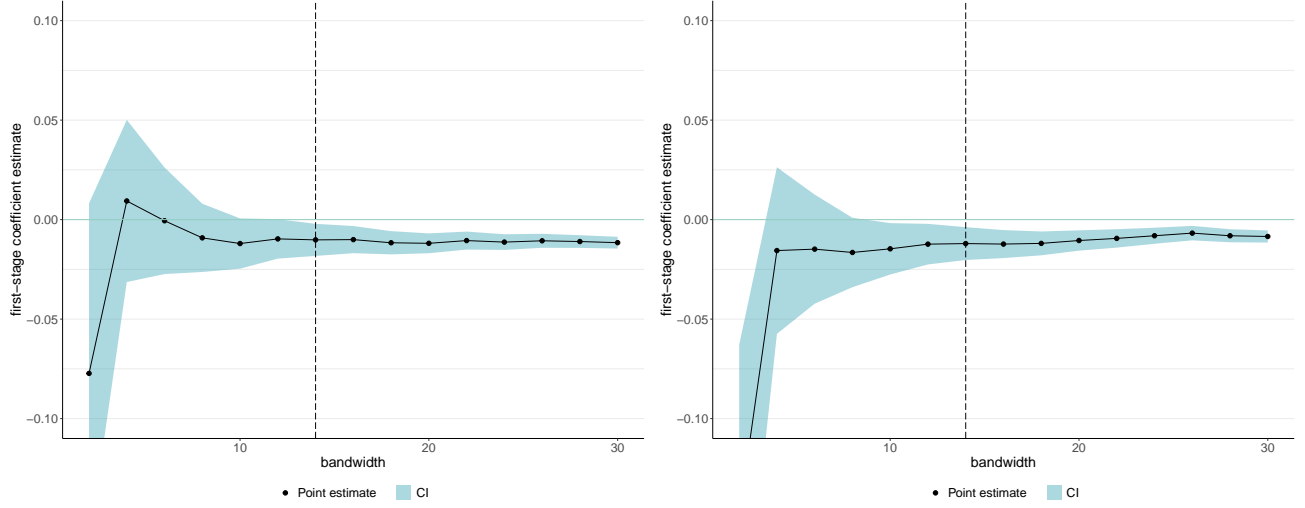
The RK results indicate that for “retaking compliers” who let their first-attempt score dictate whether they retake the exam, there is a large positive premium of getting certified: around \$18,000–\$21,000 per year. Implicit in this interpretation is an assumption about monotonicity in the retaking behavior: I 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.

2.5.4 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, I omit one college institution from my 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.

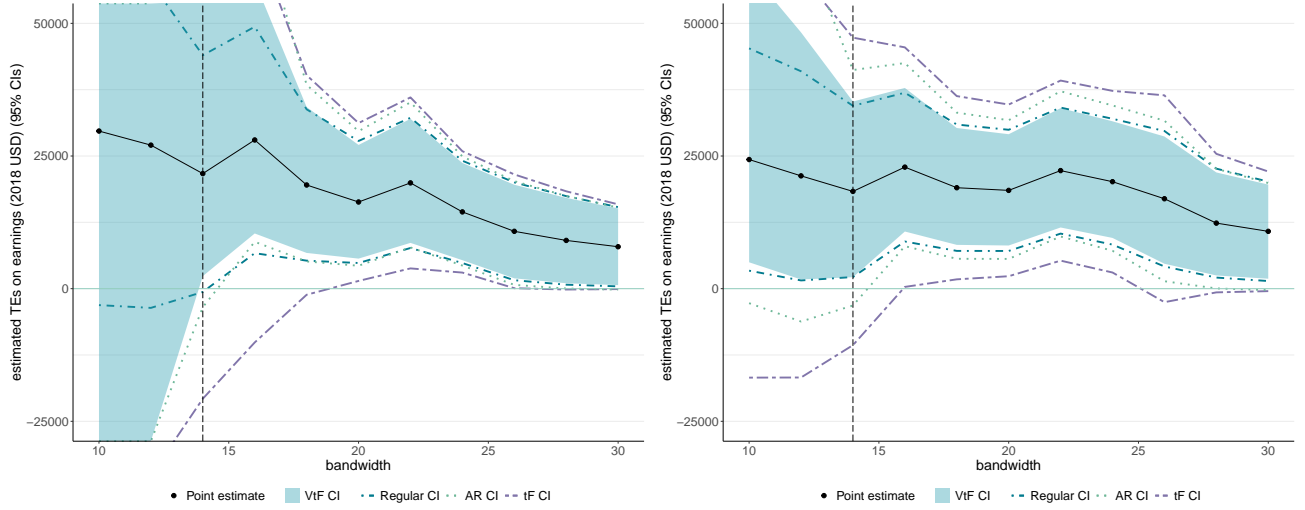
In addition, because of small sample sizes and the resulting weak first-stages in the RK, I present the classic robust 2SLS standard errors as well as Anderson-Rubin confidence intervals (CIs), tF CIs resulting from implementing the [Lee et al. \[2022\]](#) tF procedure, and the VtF CIs from implementing the [Lee et al. \[2024\]](#) VtF procedure. The tF confidence intervals are by which the estimates begin to stabilize.

Figure 7: First stage and 2SLS estimates from the RK design



(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.

Note: The first row shows estimates from first stage regressions over bandwidths ranging from 2 to 30. The treatment variable in Panel (a) is the likelihood of becoming certified in 4 years, meaning the coefficients can be interpreted as the change in the slope of $\Pr(\text{certified})$ at the cutoff. The treatment variable in Panel (b) is the likelihood of becoming a teacher in 4 years after the first attempt. The second row shows the 2SLS coefficients over bandwidths from 10–30; smaller bandwidths are omitted because the first stage estimates become stably significant starting at a bandwidth of 10. Panel (c) plots the 2SLS estimates of the effect of becoming certified on earnings, while Panel (d) plots the 2SLS estimates of the effect of becoming a teacher on earnings, 4 years after taking the Praxis for the first time. Multiple 95% confidence intervals (CIs) are shown following recommended standard error corrections for weak instruments, including: Anderson-Rubin CIs, tF CIs [Lee et al., 2022], and VtF CIs [Lee et al., 2024].

far the widest, encompassing zero for bandwidths less than fourteen. However, my 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 I investigate the effect on longer-run earnings five or six years after taking the test, although the sample size becomes considerably smaller.

2.6 Empirical Strategy II: Event Study

The second strategy I use to examine the pay gap between teaching and teachers’ next-best options is an event study design. I examine how pay changes for teachers who exit teaching and change jobs in two data sources: the Kentucky administrative data, and the Current Population Survey (CPS) matched sample. Relative to the RK, the event study permits me to “get away from the (test score) cutoff,” in that one would not expect that exiting teachers are necessarily those who would have scored near the passing cutoff on the Praxis. One caveat to this design however is that the resulting estimated pay changes may not be causal, since moves may be endogenous to pay. To address this concern, I apply the event study design to teacher exits that were likely driven by factors unrelated to individual teachers’ preferences, including “mass layoff” (school closure) events and principal turnover. These results provide complementary evidence to the RK estimates using a different source of variation.

2.6.1 Event study specification

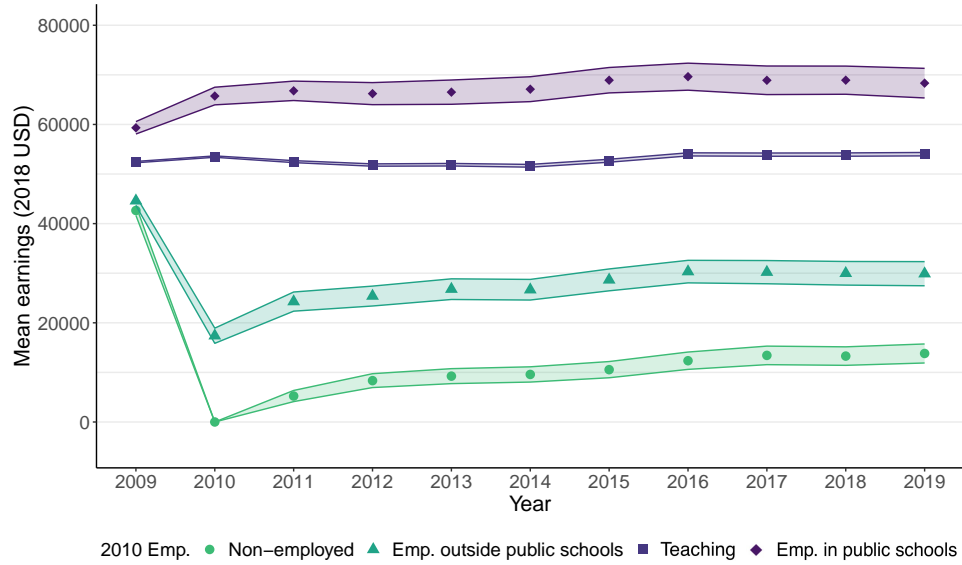
Using the administrative data, I identify the sample of teachers who move out of teaching at some point between 2009 and 2018, whose move cannot be attributed to retirement, and who are employed as a teacher for at least three years prior to the move. I stack their administrative data so that their dates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

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

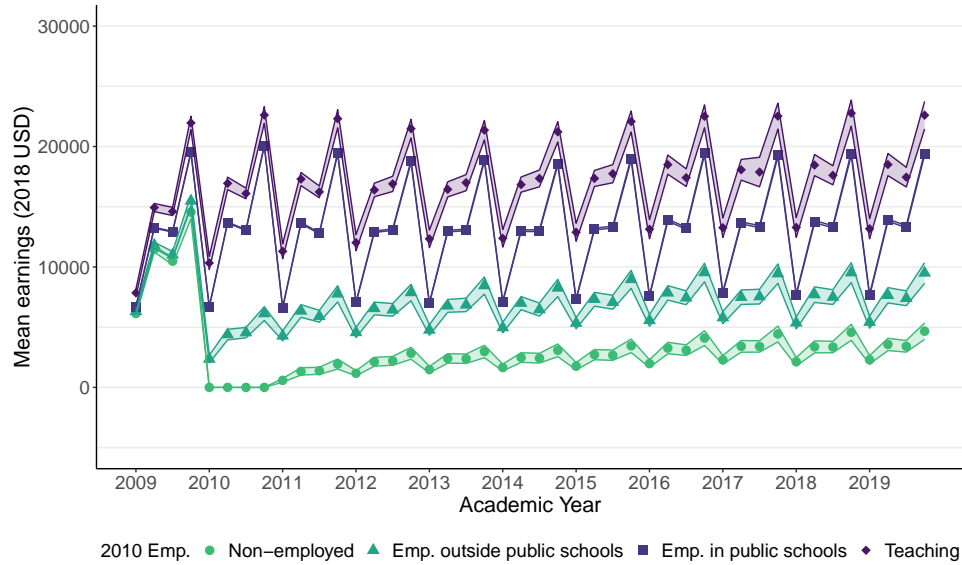
where the coefficients α_{τ} capture the adjusted earnings difference between year $t = -1$ and year τ , and $X_{i,t=-1}$ are controls describing the teachers’ characteristics one year prior to exit.

I 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, this is a useful validation exercise to check whether the large pay gap we observe in the administrative data is due to

Figure 8: Teachers' earnings trajectories



(a) Annual earnings.



(b) Quarterly earnings.

Note: Each figure plots earnings trajectories for teachers in 2009, grouped by the type of transition they made in 2010. The four transition groups are: teachers who stay teachers in 2010 (squares), teachers who exit teaching but continue being employed in public schools in 2010 (diamonds), teachers who exit teaching but continue being employed in 2010 (triangles), and teachers who exit teaching and do not appear in the UI data in 2010 (circles). Earnings for those not in the UI data are coded as 0. Panel (a) shows annual earnings, constructed by aggregating quarterly earnings, which are shown in Panel (b). An academic year is defined as the first two quarters of the year and the last two quarters from the previous year.

features specific to the administrative data, such as its coverage of only within-state employment. Second, the CPS contains data on occupations and hours across industries, allowing me to examine the types of occupations and part-time/full-time work that teachers leave teaching for.

2.7 Event Study Estimates of the Pay Gap

Figure 8 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.²⁷

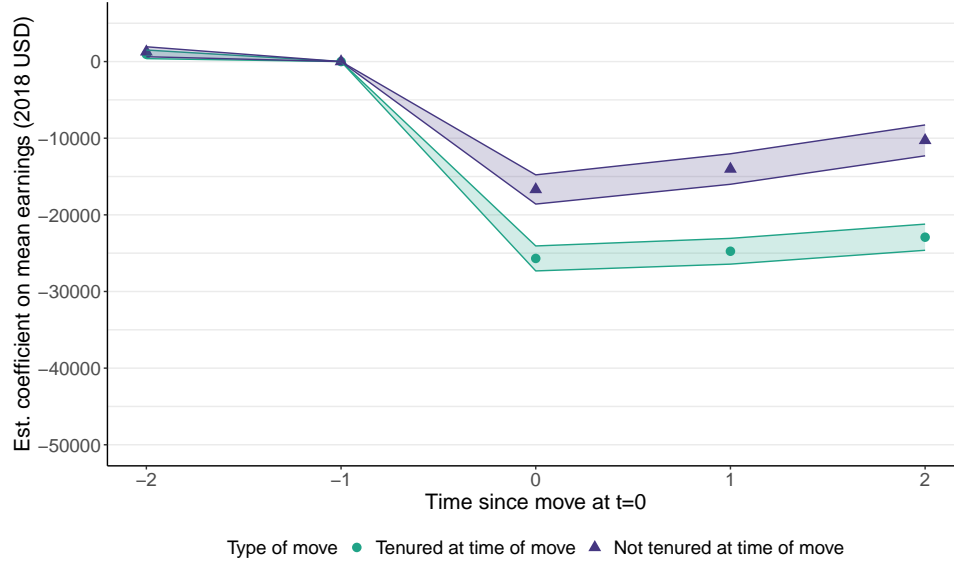
The raw means, while illustrative, may reflect pay differences biased by different initial levels of pay across districts or different ages among leaving teachers. I therefore turn to Figure 9, which shows the estimated coefficients from the event study specification when I control for pre-determined observables at the time of exit including gender, race, age and years of experience teaching at the time of move, and district fixed effects.

The regression-adjusted estimates suggest that leaving teachers experience a pay decrease of between \$18,000 and \$25,000 a year upon leaving teaching. I 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. The data are rich enough such that I can (1) distinguish voluntary exits from involuntary ones, and (2) identify when exits coincide with a more systematic turnover event at a school. First, to investigate whether voluntary separations might be driving the gap, I separately the earnings trajectories separately for teachers who were tenured versus not tenured at exit.²⁸ Voluntary moves made by tenured teachers coincide with larger earnings drops than plausibly involuntary moves 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, I

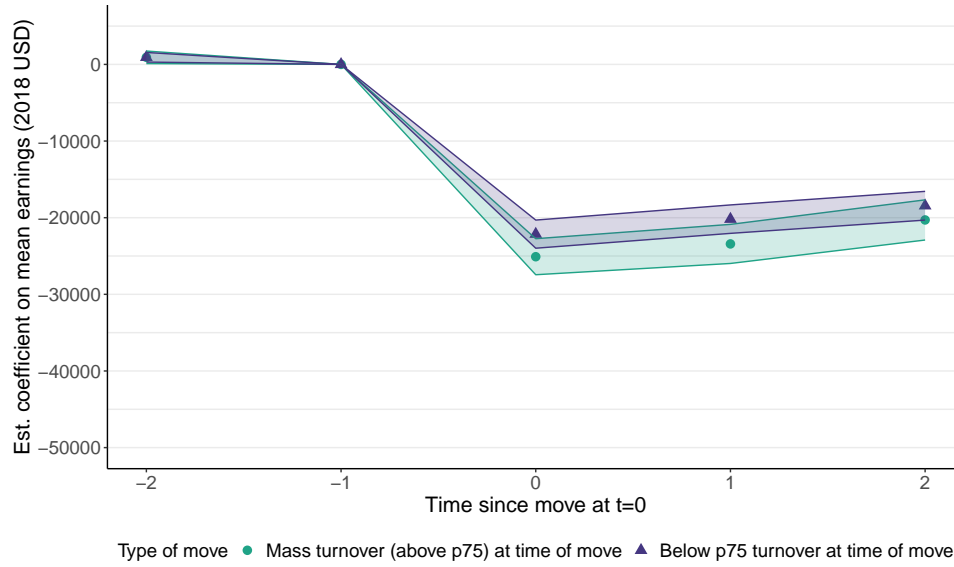
²⁷I code non-employment as zero earnings to maintain the same sample throughout.

²⁸Teachers 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.

Figure 9: Event study estimates around teacher exits



(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

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

also 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. I define a “mass turnover” as a school turnover rate that falls above the 75th percentile in a given year. I find similarly large earnings drops for teachers who leave from “mass turnover” schools.²⁹

I draw similar conclusions from my analysis of the matched CPS. From the matched CPS, I 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). Nearly half of leaving teachers exit to full-time jobs while one-fifth exits to part-time work.

Overall, the event study findings corroborate my RK estimates of a large estimated pay gap. My event study estimates are very similar in magnitude to my 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, I find suggestive evidence that the gap can persist even ten years after the initial exit from teaching.

2.8 Putting the pay gap estimates in context

Using two empirical strategies that leverage different sources of variation in who is a teacher in Kentucky, I find consistent evidence that the pay gap between teaching and teachers’ next-best jobs ranges between \$18,000 to \$21,000 a year, in 2018 US dollars. Can we think of this as the pay gap for the average teacher? And how does the pay gap compare to what teachers actually earn in Kentucky?

I interpret this range of estimates as providing an *upper bound* on the pay gap for the average teacher. The RK treatment effects are weighted towards marginal test-passers, amounting to nearly 40% of the sample of test-takers in the optimal bandwidth around the cutoff. My empirical 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, marginal and lower-scoring students may also be expected to have lower earnings potential; indeed, as shown in the reduced form analysis, short-term future income is weakly increasing in initial test scores. The test-takers driving the results are therefore those with relatively low

²⁹Additionally, I look at moves that occur at the same time as a principal change. My findings are similar.

earnings potential compared to the average teacher in the sample. Similarly, while I 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 (e.g., [Chingos and West \[2012\]](#)), 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 potentially low earnings potential, leading us to conclude that the estimated pay gaps serve as an upper bound. This is advantageous in our context: given that the primary goal of this paper is to assess the extent to which teachers earn rents, an upper bound estimate is more conservative and therefore preferred to a lower bound estimate of the pay gap.

Finally, I use a back-of-the-envelope calculation to put my estimates in context. [Table 3](#) shows the results of dividing the 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, I use the average teaching salary across districts for those with two years of experience and the most common set of credentials as a benchmark: approximately \$45,000 in 2018 USD.^{[30](#)},^{[31](#)} For the event study analysis, where the main outcome is earnings one year after leaving teaching, I use the average teaching salary one year prior to exit as a benchmark: around \$55,000 in 2018 USD.^{[32](#)} [Table 3](#), Column (2) shows my preferred calculation, 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.^{[33](#)}

³⁰These credentials can be thought of as courses that can be obtained within the first two years of teaching.

³¹There is essentially no variation in these average salaries between 2018 and 2022 after adjusting for inflation. 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 it is common for teachers to work a second job.

³²Based on the salary schedules, this implies that the average leaving teacher has between 5–10 years of experience.

³³My 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 7](#). 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 7](#) 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

	Salary (1)	Pay Gap / Salary	
		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: Column (1) shows salaries from the 2019 salary schedules for teachers, available online from the Kentucky Department of Education. Column (2) shows the ratio of the lower bound estimate of the pay gap to the salary in Column (1). Column (3) shows the ratio of the upper bound estimate of the pay gap to the salary in Column (1).

3 Estimating the Working Conditions Gap

3.1 Identification Challenge

In the second half of this paper, the 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. The first is to identify and estimate teachers’ willingness to pay for non-wage amenities in general. The second is to identify the non-wage amenities offered by teachers’ next-best jobs. Neither set of quantities can be inferred directly from observational data on job choices, since workers select into jobs for unobserved reasons, including their valuations over amenities. Even longitudinal data that tracks job transitions, which could feasibly be used to control for worker- or firm-specific attributes, rarely contains detailed information on the amenities each job provides.

To overcome the first challenge, I administer a set of stated-preference experiments to a representative sample of teachers in Kentucky. In the experiment, I present teachers with a series of job pairs with randomly varied job attributes and ask them to choose the job they most prefer. The resulting choice data allows me to observe the trade-offs teachers face and the choices they make. I use the resulting data to 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, I identify the most popular industries that teachers take their next-best jobs in and leverage the survey data from the American Working Conditions Survey (AWCS) [Maestas et al., 2023] to identify the amenities offered by the average job in these industries. I triangulate the “next-best” industries using multiple data sources: the industries of the next-best jobs from the RK design and administrative data; the industries of

the next-best jobs after exit from the event study design in the administrative data and the CPS; and the industries teachers report in the survey itself, where I directly ask teachers what their next-best job option would be. I then take the industry-level averages of the amenities reported in the AWCS.

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

3.2 New Data on Working Conditions

3.2.1 Survey of Kentucky teachers

To gain an understanding of how working conditions vary between teaching and other jobs, I field an online survey on a representative (once weighted) sample of teachers in the state of Kentucky. I conducted two rounds of the survey in May 2024, resulting in 587 surveys of teachers and 48 surveys of other school staff, and a total of 2,745 choices from the choice experiment.³⁴ The email response rate was 6%, exactly the average email survey response rate according to Qualtrics.

The purpose of my 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 working conditions. My approach is to replicate the American Working Conditions Survey (AWCS) as closely as possible: I ask exactly the same questions and follow the experimental design described in [Maestas et al. \[2023\]](#) in the stated preference module as well.³⁵

I conduct my own survey rather than use the data from the AWCS because my population of interest is specifically Kentucky teachers. To ensure that consistent comparisons can be made between my results from the RK, the event studies, and the survey, it is important that my pay gap and willingness-to-pay estimates are both derived from data on Kentucky teachers. However, the AWCS does not collect occupation for the majority of respondents. Furthermore, while a small sample of 46 (out of the total 1,748 respondents) report “teacher” as their occupation, these individuals reside in various places across the country.³⁶

³⁴Each 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.

³⁵The 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 jobs and asks respondents which job they prefer in each case.

³⁶This is a conservative estimate; some of these 46 respondents appear to have only typed in part of their

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

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

There are two key advantages to modeling my survey off of the AWCS: credibility and comparability. First, for the survey to be time-efficient and comprehensive, it is crucial that the survey asks about a small but core set of job attributes that workers deem important. As such, the AWCS was designed based on a thorough literature review across fields and numerous survey pilots to identify job attributes that respondents found important and that varied across jobs. By replicating the AWCS survey instrument, I can be confident that I am assessing general job attributes of the teaching job that workers would consider salient when choosing between teaching and other jobs. Second, since my 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. Here, the AWCS provides several natural sets of benchmark estimates: for women, college-educated workers, and workers in the education and health sector.

I only diverge from the AWCS at the end of my survey, when I ask respondents 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. The survey sample closely resembles the population 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, I 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

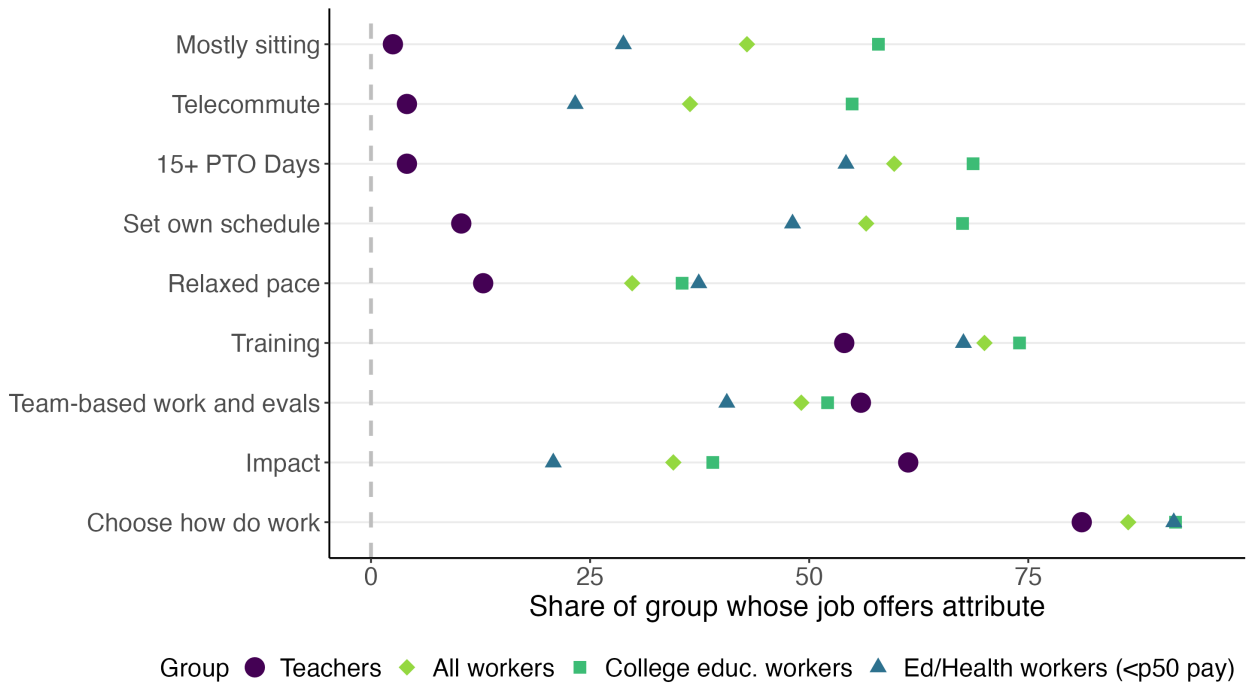
I begin by examining the incidence of the nine core non-wage job attributes among Kentucky teachers. Figure 10 presents a summary of the incidence of working conditions for four groups of workers: Kentucky teachers, all workers in the U.S., college-educated workers, and workers in the education and health industries with below median pay. All of the numbers displayed in the figure, along with more summary statistics on the job attributes in teaching compared to the jobs held by other workers in the AWCS, are provided in Appendix Tables B.1, B.2, and B.3. All means are weighted using the corresponding survey weights.

Figure 10 shows that teaching is significantly less likely than the average job held by a

 occupation, e.g. “teach” or “teac”.

college-educated worker to offer valued working conditions, even those that are commonplace in other jobs. Most college-educated workers 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).

Figure 10: The incidence of working conditions across jobs



Note: This figure shows the percentage of workers who report having each of 9 desirable job attributes, for 4 groups: Kentucky teachers (circles, from my survey), all workers in the US (diamonds, from the AWCS) college-educated workers (squares, from the AWCS), and workers in the education and health sector with below-median earnings in their sectors (triangles, from the AWCS). The attributes are sorted from least common in teaching (top) to most common in teaching (bottom).

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 summers are not technically paid time off. However, teaching also differs along other less obvious dimensions. Perhaps most starkly, teachers are significantly more likely than all other 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% in teaching vs. 74% for college-educated workers), 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% in teaching vs. 39% for jobs of college-educated workers).

The data on the small sample of teachers in the AWCS corroborate my survey findings. I identify teachers in the AWCS by merging occupation information from the RAND American Life Panel. Appendix Table B.4 shows summary statistics on the differences between the AWCS teachers and other workers.³⁷ Reassuringly, I reach the same qualitative conclusions using the AWCS data: that is, that teaching offers lower amenity value than the jobs of observationally similar workers.

Comparing my survey data to the AWCS also corroborate prior studies that document lower pay in teaching compared to jobs held by demographically similar workers (e.g., [Allegretto and Mishel \[2018\]](#)). Appendix Table B.1 shows that Kentucky teachers are contractually obligated to work a similar number of hours per week as the average U.S. worker, women worker, or college educated worker in the AWCS. However, the hourly wage in teaching is closer to the wage of the average U.S. worker or woman worker than it is to the average college educated worker.

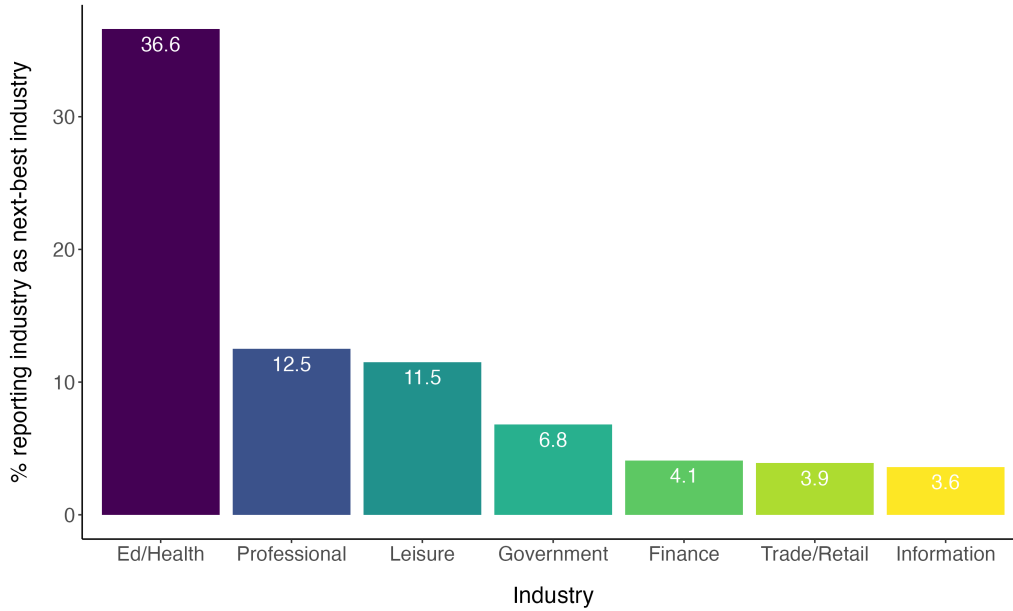
Next, I examine the characteristics of the industries that Kentucky teachers report they would work in for their next-best jobs.³⁸ Figure 11 gives the distribution of responses from my survey. Over a third of respondents indicate that their next-best job would be in either the education or healthcare 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. The survey responses are in line with the actual job choices that I observe in the RK and event study designs, where I find teachers' next-best jobs are in education, healthcare, administrative and support services (which is classified under professional and business services), and retail (which is classified under leisure and hospitality or trade).

How does teaching compare to the jobs in these industries? Appendix Table B.2 compares the job attributes in teaching to the attributes of the average job in each next-best industry. In terms of pay, the average job in all industries, aside from leisure and hospitality, pays more on average than teaching does. In addition, in terms of working conditions, the average job in all other industries is substantially more likely to offer many of the desirable job attributes that

³⁷I report statistical tests in Appendix Tables B.5 and B.6.

³⁸I am inquiring about specific occupations in ongoing iterations of the survey.

Figure 11: Most popular next-best industries reported by Kentucky teachers



Note: This figure shows the distribution of responses to the following survey question: “If you were not a teacher today, what alternative job might you be doing instead? Please visualize a job that you think is realistic, such as a job you may have applied to before, or a job you or your colleagues may have worked in besides teaching. Which industry would this job be in?”

teaching does not have, such as: scheduling flexibility, telecommuting options, relaxed pace of work, and more than 15 days paid time off.

That said, given that my pay gap estimates suggest that teachers’ next-best options pay substantially less than teaching does per year, I also examine how the attributes in teaching compare to the attributes of jobs in these next-best industries with low pay. Figure 10 shows only the incidence of job attributes among the most popular industries: education and health sector workers with below median pay. Appendix Table B.3 examines the attributes of the average job in each industry that pays below the industry median wage. The distribution of attributes is very similar to Table B.2. In particular, teaching is significantly less likely to offer most desirable attributes even when compared to jobs held by workers in the education and health sectors earning below median pay.

3.4 Estimating Teachers’ Willingness to Pay for Working Conditions

To estimate teachers’ willingness to pay for each job attribute, I conducted a series of three stated-preference experiments with each survey respondent. To maximize the size of the data

and minimize survey fatigue, I also gave respondents the option to respond to at most two additional stated-preference experiments, so that respondents could complete up to five experiments.³⁹ While the majority actually did respond to at least one more stated-preference experiment, my main results use the first three experiments to avoid any selection bias from additional responses.

I detail my experimental design and estimation strategy below, with additional details in Appendix G. I include “trick” questions to address the concern with that survey respondents may only skim the questions and thus provide inaccurate responses. My results are not sensitive to the inclusion or exclusion of inattentive respondents.

3.4.1 Experimental design

In the first part of the survey, respondents are 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.⁴⁰ These responses yield the shares shown in Figure 10. To obtain normalized measures of pay, I 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. I then use these responses to calculate each respondent’s hourly wage.

In the second part of the survey, respondents are presented with a minimum of three stated-preference experiments. In each experiment, respondents are asked to select between two jobs that partially vary in terms of job attributes, hours, and wages, exactly following the technical and visual design described in Maestas et al. [2023]. After the third experiment, all respondents are given the option to evaluate one to two additional job pairs, if willing.

I construct the hypothetical job profiles in two stages. First, I 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 functions as an initial set of values. I randomly vary subsets of these values to construct hypothetical job profiles in the second stage. Using the baseline job is advantageous because the resulting hypothetical job profiles resemble a job that the respondent is familiar with.⁴¹

³⁹No additional incentives were provided; survey respondents were simply told that completing the additional experiments would “be a huge help to the study.”

⁴⁰See Appendix Table 2 of Maestas et al. [2023] for a list of the job attributes and their potential values.

⁴¹However, to address concerns that respondents may prefer job profiles that they are familiar with, which could bias the willingness-to-pay estimates, I also define a “common” baseline job—essentially the average teaching job based on the survey pilot—and use it in one of the three experiments. The values of the common baseline job are shown in Appendix Table 2 of Maestas et al. [2023]. I find that the inclusion of the common

Second, I 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. I also vary the offered wage and hours per week between the two jobs, such that in total, each job is characterized by eleven total 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 respondent’s actual hourly wage. Given the respondent’s actual hourly wage w_i , the hypothetical wages for Job A and Job B 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, I truncate θ_A and θ_B at 0.75 and 1.25. To make these wage values easier for the respondent to interpret, I display the earnings in the hypothetical job choice in terms of the units in which the respondent chose to report their earnings. Finally, in an effort to maximize precision of my estimates given I only run three experiments per person, I redraw wages and/or attribute values in cases where one job dominates the other on all dimensions.⁴²

As a result, each pair consists of a Job A and a Job B that are identical in every way except for two non-wage attributes, the hourly wage, and the hours worked per week. Respondents are presented with a table that provides a side-by-side comparison of the two jobs’ attributes, and asked whether they “Prefer Job A,” “Prefer Job B,” “Strongly Prefer Job A,” or “Strongly Prefer Job B.” My presentation of the job pairs, shown in Appendix Figure A.5, is virtually identical to the presentation in Appendix Figure 3 of [Maestas et al. \[2023\]](#).

3.4.2 Estimation

Because I randomize the attributes across hypothetical job choices, I am able to use the choice data from the stated preference experiments to estimate Kentucky teachers’ willingness-to-pay for each job attribute. I assume that the respondents have the following indirect utility function:

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

where A_{ij} is the vector of job attributes offered by job j and w_{ij} is the corresponding wage. Under the assumption that ε_{ij} is i.i.d. EVI, the likelihood that worker i chooses teaching over job k is given by the expression

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

baseline jobs does not affect our willingness-to-pay estimates.

⁴²See Appendix G for details on randomization.

In theory, the β_i and δ_i subscripts allow for heterogeneity in preferences over amenities and pay. However, because I focus on Kentucky teachers, who are a specific and less variable group of individuals, I assume that $\beta_i = \beta$ and $\delta_i = \delta$ and use a standard logit model to estimate β and δ . I 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. I also conduct a battery of robustness checks, 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 my smaller sample, I find that my willingness-to-pay estimates are robust to these variations.

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

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

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 the object of interest:

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

For ease of interpretation, I report our estimates in terms of $1 - \exp \left(\frac{-\beta^r}{\delta} \right)$, so 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.

Finally, I use the preference estimates to estimate how much teachers are willing to pay for one job over another. I am interested in three such comparisons: (1) the average teacher’s total valuation for the “best” job over the “worst” job (i.e., the set of all desirable attributes over the set of none of them), (2) the average teacher’s total valuation of all the desirable amenities teaching does not offer, and (3) the total amenity value the average teacher would get from switching from teaching to the their next-best job.⁴³

I introduce some notation that will be used throughout the definitions for each of the three measures below. Let S denote the set of all attributes. Let $\beta^{s,1}$ denote the preference parameter for the highest attribute value of s . In cases where there are more than two values for attribute

⁴³I continue using low-paying jobs in the education and health sectors in the AWCS as a proxy for teachers’ next-best jobs.

s , let $\beta^{s,2}$ denote the preference parameter for the second highest attribute value of s .^{44,45}

The first measure is respondent i 's willingness to pay for the “best” job over the “worst” job, defined as:

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

The sum in this expression only puts positive weight on the estimated preferences for the most preferred value of each attribute.⁴⁶ This measure is equivalent to the best-to-worst measure used in Maestas et al. [2023], re-expressed for notational consistency.

The second measure is respondents' willingness to pay for all of the desirable amenities that teaching does not offer:

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

where $\mathbb{I}_{\text{teach}}^{s,j}$ is an indicator variable for whether the average Kentucky teaching job offers the j th highest attribute value of s or not. The sum in this expression only puts positive weight on the estimated preferences for the desirable attributes that teaching does *not* offer. Unlike Equation (7), this measure accounts for differences between the “best” job and a job that offers *some* desirable amenities, rather than none at all.⁴⁷

Finally, the third measure most closely captures the compensating differential between teaching and teachers' next-best jobs. I define the measure as the respondents' willingness to pay for their next-best job—that is, a below-median-pay job in the education or healthcare sector—over teaching:

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

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

⁴⁴If the attribute has only two values, $\beta^{s,2}$ is 0.

⁴⁵For example, if s is the physical demands attribute, $\beta^{s,1}$ is the preference parameter of moderate physical demands (the most preferred value) and $\beta^{s,2}$ is the preference parameter for mostly sitting.

⁴⁶Note that in this baseline measure, I am careful not to “double count” attribute values, which would inflate my summary measures. If an attribute has more than two values, I only use the coefficient associated with the attribute value with the highest willingness-to-pay estimate.

⁴⁷The 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 $(\beta^{s,1} - \beta^{s,2} \cdot \mathbb{I}_{\text{teach}}^{s,2})$.

of attribute s or not. The sum in this 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.

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

My 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, I 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. I also assume that there were no systematic differences in how respondents rated the subjective amenities like heavy versus moderate physical activity, which is reasonable given that our survey respondents are somewhat homogeneous (e.g., mostly women) and engage in similar work.

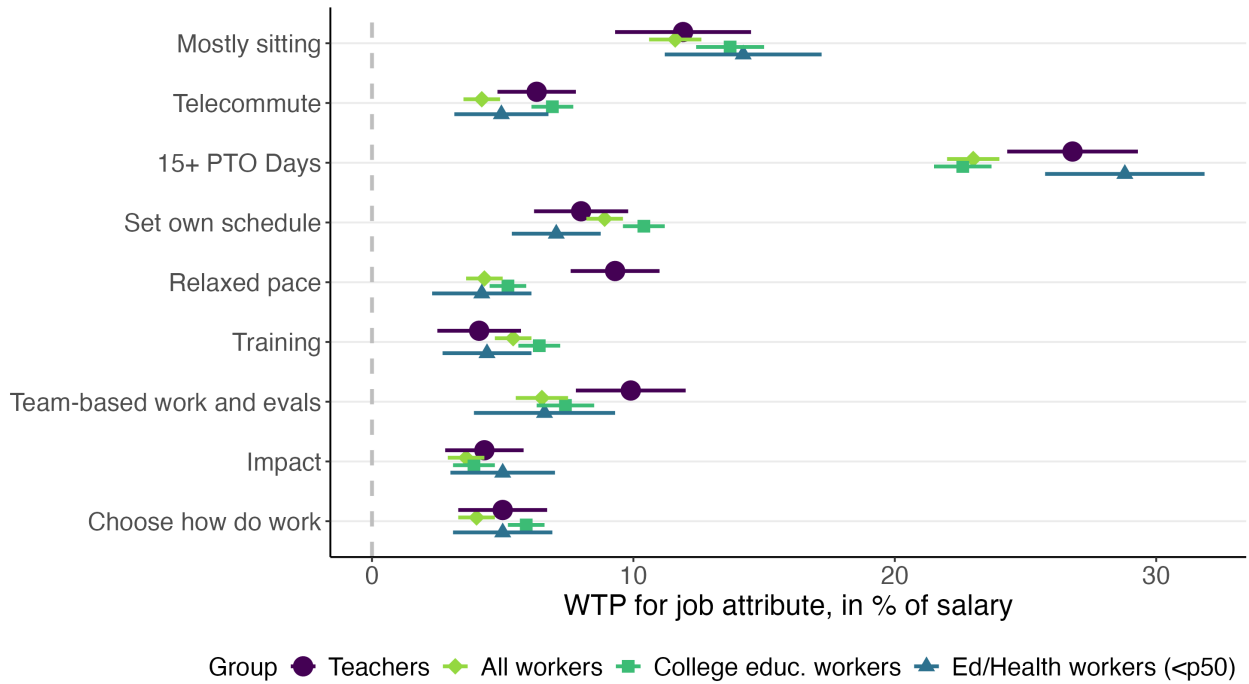
3.5 Main Estimates of Teachers' Willingness to Pay for Working Conditions

I present my main estimates of Kentucky teachers' willingness-to-pay for each job attribute in Figure 12, alongside estimates from the AWCS for all workers in the U.S., college-educated workers, and education/health sector workers with below median pay.⁴⁸ The estimates are transformed so that each point 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 estimate for teachers in the first row indicates that the average Kentucky teacher is willing to pay 11.9 percent of their salary to be able to mostly sit at a job rather than engage in heavy physical activity. Although my sample is significantly smaller than the AWCS, my estimates are precise because my population of study is specific and relatively homogeneous.

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

⁴⁸Estimates are also shown in Appendix Table B.7.

Figure 12: Willingness-to-pay estimates



Note: This figure shows willingness-to-pay estimates for each of the nine job attributes and for 4 different groups: Kentucky teachers (circles, from my survey), all workers in the US (diamonds, from the AWCS) college-educated workers (squares, from the AWCS), and workers in the education and health sector with below-median earnings in their sectors (triangles, from the AWCS). The attributes are sorted from least common in teaching (top) to most common in teaching (bottom).

Table 5: Estimates of total willingness-to-pay for one job over another

	AWCS			
	Teachers (1)	All workers (2)	College-educ. (3)	Ed/Health, < median wage (4)
Best job <i>vs. Worst job</i>	0.622 (0.033)	0.550 (0.016)	0.600 (0.016)	0.611 (0.034)
Best job <i>vs. Teaching</i>	0.347 (0.037)			
Next-best job <i>vs. Teaching</i>	0.286 (0.037)			
N	1,662	17,380	9,080	2,710

Note: Column 1 shows estimates of Kentucky teachers’ willingness to pay using data from the stated preference experiments. Column 2 shows the same estimates for all workers in the AWCS; Column 3 shows the estimates for college educated workers in the AWCS; and Column 4 shows the estimates for workers in the education or health sectors in the AWCS earning below the median wage in the two sectors. Standard errors, shown in parentheses, are estimated using the delta method and are clustered at the respondent level.

Second, I find that teachers’ preferences are very similar to those of demographically similar workers. The only job attribute that teachers value 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. One explanation for this difference is the prevalence of teacher “burnout”: surveys show that stress and burnout are especially common among teachers and contribute to high turnover rates in teaching, especially among starting teachers (e.g., [Herman et al. \[2018\]](#)). Workers who are burned out at the time of taking the survey may place a higher value on a relaxed pace of work.

Some of the insignificant differences in preferences are noteworthy as well. In particular, 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 Figure 10, 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. I also find that, although teachers are significantly more likely to work alone than the average American worker, teachers marginally prefer working in teams over working alone.

Third, because the average teaching job does not offer most of the job attributes I study, teachers are willing to pay a substantial amount of their salary to switch to a job with more

desirable amenities. Table 5 shows the resulting estimates. The first row shows that switching from the least desirable job to the most desirable job, in terms of amenities, is equivalent to a 62.2 percent wage increase for Kentucky teachers. 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). The second row shows that switching from teaching to the “best” job, in terms of amenities, is equivalent to a 36 percent wage increase. Finally, the third row shows 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

Applying independent quasi-experimental designs to novel administrative data from Kentucky, I find that teaching offers a large pay premium relative to teachers’ next-best options. I estimate a pay gap at between 33–40% of the average teachers’ annual earnings. I also find that Kentucky teachers are willing to pay 35% of their salary for a job with desirable attributes that teaching does not offer, or 29% of their salary for their next-best job.

Combining my most conservative estimates using the following expression,

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

I 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%). I cannot reject the hypothesis that teachers earn no rent at all.

My results suggest that, although teaching pays more than teachers’ next-best jobs, most of the pay premium is a compensating differential rather than a rent. In other words, the overall compensation in teaching is relatively similar to teachers’ next-best options once both pay and working conditions are taken into account—not because the jobs are similar, but because teachers are being compensated for poor working conditions.

This conclusion is in line with increasing reports of teacher shortages and poor working conditions in teaching, and has significant implications for education policy. Given that over \$500 billion are spent each year on teacher salaries, my estimates imply that over \$145 billion is being spent each year to compensate teachers for poor working conditions, raising the question of whether there may be more cost-effective solutions that could address the gap in working conditions between teaching and other jobs.

Assessing magnitudes. My estimates of the gap in the value of working conditions are large. One may be concerned that I am overestimating the gap because my survey does not cover all the non-wage aspects of teaching. For instance, the survey does not include desirable amenities common to teaching like “summers off” or shorter work days.

There are two reasons why my estimate of the working conditions gap is plausible and not an overestimate. First, consider the difference in the pay and working conditions between public and private schools. Private school teaching jobs are the most similar jobs 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.⁴⁹ 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](#)]⁵⁰—a remarkably similar number to the gap in the value of working conditions that I estimate. While only suggestive, the private school/public school pay gap is consistent with my finding that public school teachers may require a large compensating differential to make up for poor working conditions driven by student behavioral issues and stress.

Second, although my survey does not include desirable teaching amenities like “summers off,” it also does not include undesirable amenities common to teaching, such as student behavioral issues or exposure to violence in schools. These undesirable amenities are salient to teachers today: for example, in a Pew Research Center survey in 2024, 59% of public school K-12 teachers reported being worried about the possibility of a shooting happening at their school. It is therefore not obvious that my current estimate of the value of teaching conditions is biased upward.⁵⁰

My estimates of the pay gap are also large. I argue that the magnitude is plausible, especially when my point estimates are interpreted as an upper bound (as discussed in Section 2.8). The only remaining concern is that the estimates may be biased because my earnings data comes a state UI system, since 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 my estimates of the pay gap could be biased upward. However, my 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 (e.g., [Champion et al. \[2011\]](#), [Scafidi et al. \[2006\]](#),

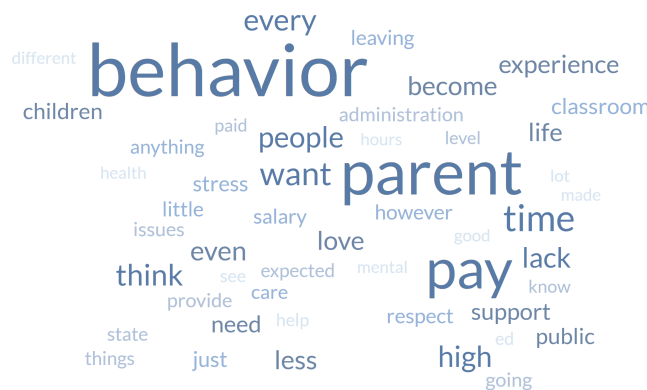
⁴⁹For example, see Ben Orin, “Why Are Private-School Teachers Paid Less Than Public-School Teachers?”, *The Atlantic*, October 24, 2013 .

⁵⁰An updated version of the survey that includes additional teaching conditions will be included in this paper in early 2025.

Goldhaber et al. [2022]). These papers find similarly sized declines in pay following exits from teaching.

Qualitative evidence. My conclusion is internally consistent with Kentucky teachers' views expressed in the open-ended portion of the survey portrayed in Figure 13. 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. This conclusion is also consistent with reports of teacher shortages and increasingly challenging working conditions in teaching, especially since COVID-19.⁵¹

Figure 13: “Word cloud” of factors affecting teachers’ job choice



Note: This figure depicts the common words used in response to the final (open-ended) question of the survey, excluding 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. My conclusions may initially seem at odds with prior work that finds evidence of rent extraction in teaching. However, once we take differences in context and approach into account, my findings 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 rents. 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: Kentucky public school teachers had the 16th lowest unionization rate in the country in 2008 (58.6%).^{52,53} In

⁵¹For example, see Sejla Rizvic, “Teachers, Facing Increased Level of Stress, Are Burned Out,” *The New York Times*, March 13, 2023.

⁵²The fifteen states with unionization rates lower than Kentucky, in decreasing order, were: Louisiana, Oklahoma, Florida, Georgia, Kansas, South Dakota, Wyoming, Virginia, North Carolina, Arizona, New Mexico, Mississippi, Arkansas, and South Carolina. The lowest rate of unionization was 26.9% in South Carolina.

⁵³Data from the School and Staffing Survey [National Center for Education Statistics, 2008].

2011, less than 30% of Kentucky teachers were covered by a union that actively participated in collective bargaining.⁵⁴ My findings are therefore applicable to many states with similar degrees of teacher union involvement, but not necessarily applicable to the average U.S. state.

This is also the first study to explicitly incorporate the value of non-wage amenities when evaluating whether workers extract rents. Most prior work uses total funding inputs for staff expenses to study rent extraction in teaching. However, most of the job attributes that I 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 established findings about teachers. The first is that teachers have become increasingly negatively selected on academic performance (e.g., [Bacolod \[2007\]](#), [Corcoran et al. \[2004\]](#), [Podgursky et al. \[2004\]](#)).⁵⁵ 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 negatively selected because the working conditions in teaching are declining. 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 relative job attributes may affect selection into teaching has received considerable attention (see e.g., [Hanushek and Rivkin \[2006\]](#)), this paper is one of the first to provide empirical evidence that supports this theory.

The second 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 (e.g., [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 since 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

⁵⁴See the report prepared by the Kentucky Legislative Research Commission [[Kentucky Legislative Research Commission, 2022](#)].

⁵⁵The findings of this literature raise a natural question: is academic performance in college, including on standardized tests such as the Praxis, related to teacher effectiveness? This question is outside the scope of this paper. However, prior work examining this question generally finds that academic performance is not highly predictive of performance, as in e.g., [Harris and Sass \[2011\]](#).

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 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 National Longitudinal Survey of Youth (NLSY) [[Stinebrickner, 2002](#)], Georgia [[Scafidi et al., 2006](#)], and Washington [[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: one explanation that would unify these findings would be if teaching were a high-paying option relative to teachers' next-best options *because* it offers a compensating differential.

Limitations. One limitation of my study is that my main results come from one state, which limits the generalizability of my findings. Since 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 few 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 I find for teachers.

Another limitation of my study is that, because my survey targets general, non-teaching-specific amenities with the intention of using the AWCS as a baseline, my results do not speak to a group of working conditions that are common in teaching but are uncommon in other jobs. The open-ended responses to my 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 my estimates. I aim to quantify these valuations in ongoing 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. This belief is at odds with recent reports of teacher shortages and challenging working conditions in teaching. To reconcile these conflicting reports, I investigate how the *overall* compensation in teaching today compares to 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 comparing 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 the United States.

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 series of stated preference experiments. My survey instrument is identical on key dimensions to the instrument used in the American Working Conditions Survey (AWCS) by [Maestas et al. \[2023\]](#), allowing me to compare my estimates to the representative sample of U.S. workers in the AWCS. My 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: interested teachers have substantially lower-paying next-best options than teaching. 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. The similarity in estimates across the two independent designs and datasets suggests that the large pay gap that exists is 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 that 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? My 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.
- N. Barton, T. Bold, and J. Sandefur. Measuring rents from public employment: Regression discontinuity evidence from kenya. *Center for Global Development Working Paper*, (457), 2017.
- M. D. Bates, M. Dinerstein, A. C. Johnston, and I. Sorkin. Teacher labor market equilibrium and the distribution of student achievement. Technical report, National Bureau of Economic Research, 2022.
- B. Biasi, C. Fu, and J. Stromme. Equilibrium in the market for public school teachers: District wage strategies and teacher comparative advantage. Technical report, National Bureau of Economic Research, 2021.
- M. Bobba, T. Ederer, G. Leon-Ciliotta, C. Neilson, and M. G. Nieddu. Teacher compensation and structural inequality: Evidence from centralized teacher school choice in perú. Technical report, National Bureau of Economic Research, 2021.
- D. Boyd, H. Lankford, S. Loeb, and J. Wyckoff. Explaining the short careers of high-achieving teachers in schools with low-performing students. *American economic review*, 95(2):166–171, 2005.
- 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.
- C. Brown and T. Andrabi. Inducing positive sorting through performance pay: Experimental evidence from pakistani schools. *University of California at Berkeley Working Paper*, 2020.
- J. K. Brueckner and D. Neumark. Beaches, sunshine, and public sector pay: theory and evidence on amenities and rent extraction by government workers. *American Economic Journal: Economic Policy*, 6(2):198–230, 2014.

- J. Bruhn, S. Imberman, and M. Winters. Regulatory arbitrage in teacher hiring and retention: Evidence from massachusetts charter schools. *Journal of Public Economics*, 215:104750, 2022.
- E. J. Brunner and A. Ju. State collective bargaining laws and public-sector pay. *ILR Review*, 72(2):480–508, 2019.
- E. J. Brunner and T. Squires. The bargaining power of teachers’ unions and the allocation of school resources. *Journal of Urban Economics*, 76:15–27, 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. R. Cellini, F. Ferreira, and J. Rothstein. The value of school facility investments: Evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics*, 125(1):215–261, 2010.
- S. Champion, A. Mastri, and K. Shaw. The teachers who leave: Pulled by opportunity or pushed by accountability?, 2011.
- M. M. Chingos and M. R. West. Do more effective teachers earn more outside the classroom? *Education Finance and Policy*, 7(1):8–43, 2012.
- D. Clark and P. Martorell. The signaling value of a high school diploma. *Journal of Political Economy*, 122(2):282–318, 2014.
- J. Cook, S. Lavertu, and C. Miller. Rent-seeking through collective bargaining: Teachers unions and education production. *Economics of Education Review*, 85:102193, 2021.
- 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.
- J. DiNardo and D. S. Lee. Program evaluation and research designs. In *Handbook of labor economics*, volume 4, pages 463–536. Elsevier, 2011.

- Y. Dong. Jump or kink? regression probability jump and kink design for treatment effect evaluation. *Unpublished manuscript*, 2018.
- M. Drake, N. Thakral, and L. Tô. Wage differentials and the price of workplace flexibility. Technical report, Mimeo, 2022.
- A. Dube, S. Naidu, and A. D. Reich. Power and dignity in the low-wage labor market: Theory and evidence from wal-mart workers. Technical report, National Bureau of Economic Research, 2024.
- C. Dustmann and A. Van Soest. Public and private sector wages of male workers in germany. *European Economic Review*, 42(8):1417–1441, 1998.
- L. Feng. Hire today, gone tomorrow: New teacher classroom assignments and teacher mobility. *Education Finance and Policy*, 5(3):278–316, 2010.
- R. B. Freeman. Unionism comes to the public sector. Technical report, National Bureau of Economic Research, 1984.
- E. García, E. Han, and E. Weiss. Determinants of teacher attrition: Evidence from district-teacher matched data. *Education Policy Analysis Archives*, 30(25):n25, 2022.
- M. Gittleman and B. Pierce. Compensation for state and local government workers. *Journal of Economic Perspectives*, 26(1):217–242, 2012.
- 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.
- J. Goodman, M. Hurwitz, and J. Smith. Access to 4-year public colleges and degree completion. *Journal of Labor Economics*, 35(3):829–867, 2017.
- J. Hahn, P. Todd, and W. Van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2001.
- E. A. Hanushek and S. G. Rivkin. Teacher quality. *Handbook of the Economics of Education*, 2:1051–1078, 2006.
- E. A. Hanushek, J. F. Kain, and S. G. Rivkin. Why public schools lose teachers. *Journal of human resources*, 39(2):326–354, 2004.

- D. N. Harris and T. R. Sass. Teacher training, teacher quality and student achievement. *Journal of public economics*, 95(7-8):798–812, 2011.
- J. J. Heckman, J. E. Humphries, and N. S. Mader. The ged. *Handbook of the Economics of Education*, 3:423–483, 2011.
- K. C. Herman, J. Hickmon-Rosa, and W. M. Reinke. Empirically derived profiles of teacher stress, burnout, self-efficacy, and coping and associated student outcomes. *Journal of Positive Behavior Interventions*, 20(2):90–100, 2018.
- 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. K. Jackson. School competition and teacher labor markets: Evidence from charter school entry in north carolina. *Journal of Public Economics*, 96(5-6):431–448, 2012.
- 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.
- C. Jepsen, P. Mueser, and K. Troske. Second chance for high school dropouts? a regression discontinuity analysis of postsecondary educational returns to the ged. *Journal of Labor Economics*, 35(S1):S273–S304, 2017.
- A. C. Johnston. Preferences, selection, and the structure of teacher pay. *Available at SSRN 3532779*, 2023.
- Kentucky Legislative Research Commission. Teachers in kentucky: Changes in turnover and workforce composition, 2022. URL <https://apps.legislature.ky.gov/lrc/publications/ResearchReports/RR377.pdf>. Accessed: 2024-11-02.
- M. A. Kraft, E. J. Brunner, S. M. Dougherty, and D. J. Schwegman. Teacher accountability reforms and the supply and quality of new teachers. *Journal of Public Economics*, 188: 104212, 2020.
- A. B. Krueger. Are public sector workers paid more than their alternative wage? evidence from longitudinal data and job queues. In *When public sector workers unionize*, pages 217–242. University of Chicago Press, 1988.

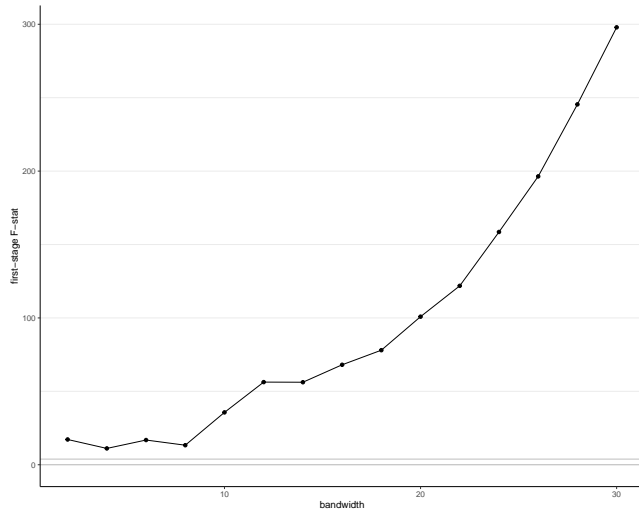
- H. F. Ladd. Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational evaluation and policy analysis*, 33(2):235–261, 2011.
- D. S. Lee, J. McCrary, M. J. Moreira, and J. Porter. Valid t-ratio inference for iv. *American Economic Review*, 112(10):3260–3290, 2022.
- D. S. Lee, J. Porter, J. McCrary, L. Yap, and M. J. Moreira. Quantile treatment effects and economic policy evaluation. (31893), 2024. <https://www.princeton.edu/~davidlee/wp/w31893.pdf>.
- S. Loeb and M. E. Page. Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics*, 82(3):393–408, 2000.
- 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. F. Lovenheim and A. Willén. The long-run effects of teacher collective bargaining. *American Economic Journal: Economic Policy*, 11(3):292–324, 2019.
- 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. D. Matsudaira and R. W. Patterson. Teachers' unions and school performance: Evidence from california charter schools. *Economics of Education Review*, 61:35–50, 2017.
- J. McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2):698–714, 2008.
- B. R. Moulton. A reexamination of the federal-private wage differential in the united states. *Journal of Labor Economics*, 8(2):270–293, 1990.
- 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. Table 043. public school teachers and percentage of teachers who are members of unions or employee associations, by state: 2007-08, 2008. URL https://nces.ed.gov/surveys/sass/tables/sass0708_043_t1s.asp. Accessed: 2024-11-02.
- 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. Podgursky and R. Tongrut. (mis-) measuring the relative pay of public school teachers. *Education Finance and Policy*, 1(4):425–440, 2006.
- M. Podgursky, R. Monroe, and D. Watson. The academic quality of public school teachers: An analysis of entry and exit behavior. *Economics of Education Review*, 23(5):507–518, 2004.
- 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.
- J. Richwine and A. G. Biggs. Assessing the compensation of public-school teachers. *Center for Data Analysis Report*, 11(03), 2011.
- J. Rothstein. Teacher quality policy when supply matters. *American Economic Review*, 105(1):100–130, 2015.
- J. E. Saavedra, D. Maldonado, L. Santibanez, and L. O. Herrera-Prada. Premium or penalty? labor market returns to novice public sector teachers. *Journal of Human Resources*, 2022.
- 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.
- Y. Shi and J. D. Singleton. School boards and education production: Evidence from randomized ballot order. *American Economic Journal: Economic Policy*, 15(1):438–472, 2023.
- T. M. Smith and R. M. Ingersoll. What are the effects of induction and mentoring on beginning teacher turnover? *American educational research journal*, 41(3):681–714, 2004.
- I. Sorkin. Ranking firms using revealed preference. *The quarterly journal of economics*, 133(3): 1331–1393, 2018.
- T. R. Stinebrickner. An empirical investigation of teacher attrition. *Economics of education review*, 17(2):127–136, 1998.

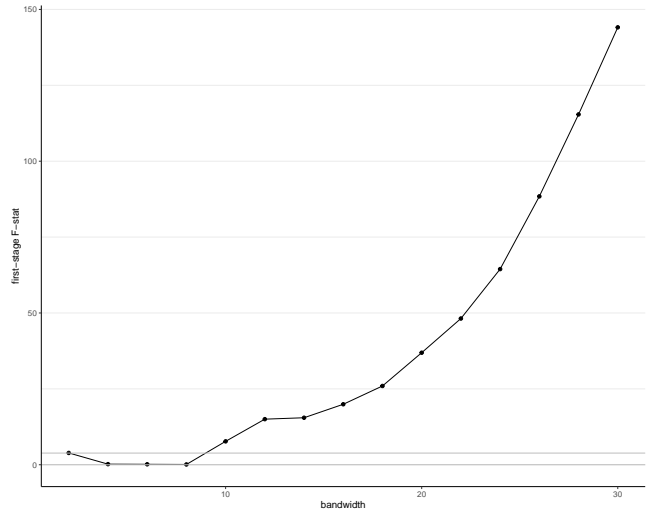
- T. R. Stinebrickner. An analysis of occupational change and departure from the labor force: Evidence of the reasons that teachers leave. *Journal of Human Resources*, pages 192–216, 2002.
- K. O. Strunk. Are teachers’ unions really to blame? collective bargaining agreements and their relationships with district resource allocation and student performance in california. *Education Finance and Policy*, 6(3):354–398, 2011.
- L. Sutch, L. Darling-Hammond, and D. Carver-Thomas. A coming crisis in teaching? teacher supply, demand, and shortages in the us. *Learning Policy Institute*, 2016.
- L. L. Taylor. Comparing teacher salaries: Insights from the us census. *Economics of Education Review*, 27(1):48–57, 2008.
- K. Terrell. Public-private wage differentials in haiti do public servants earn a rent? *Journal of Development Economics*, 42(2):293–314, 1993.
- M. Wiswall and B. Zafar. Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1):457–507, 2018.
- S. D. Zimmerman. The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711–754, 2014.

Appendices

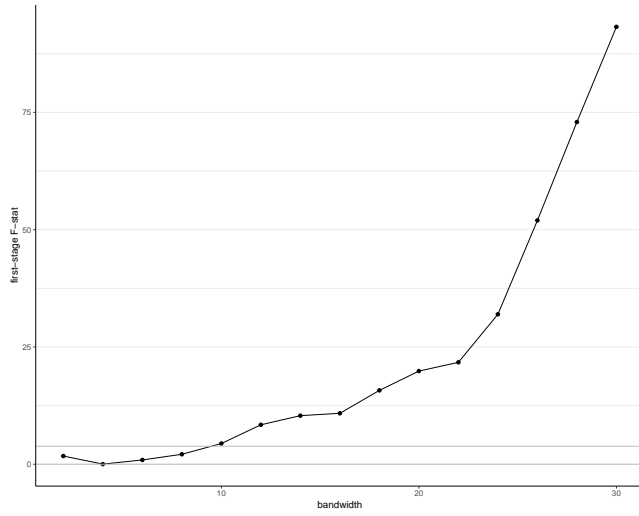
A Additional Figures



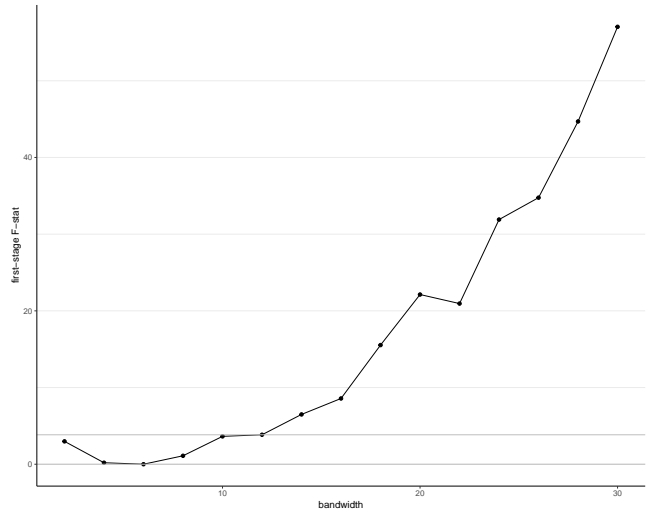
(a) Ever-passing the math test.



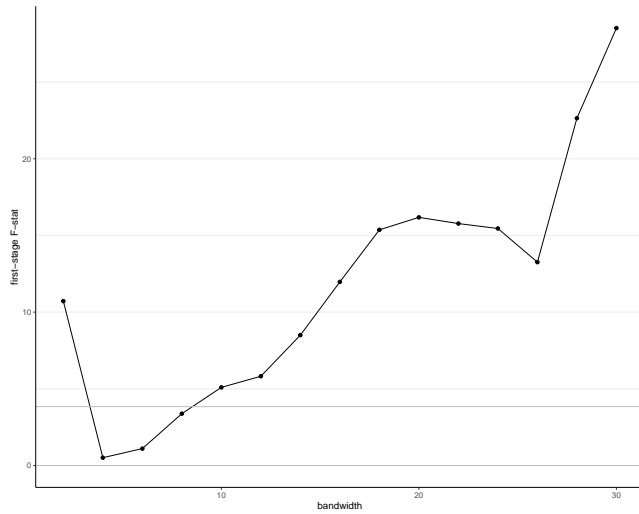
(b) Ever-passing all 3 tests.



(c) Enrolling in an EPP within 3yrs.

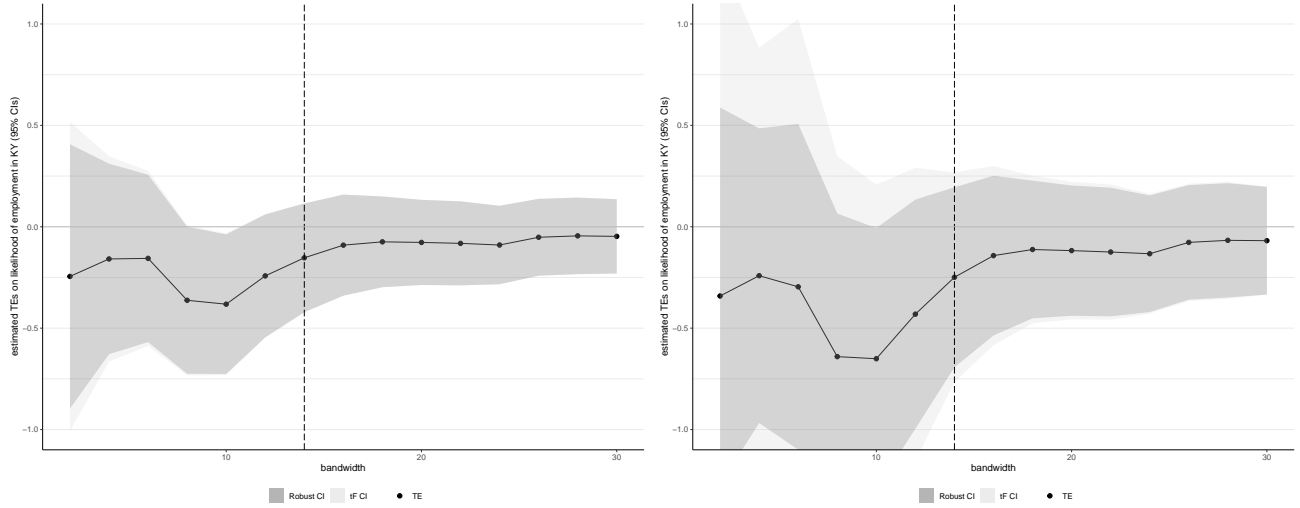


(d) Becoming certified within 4yrs.



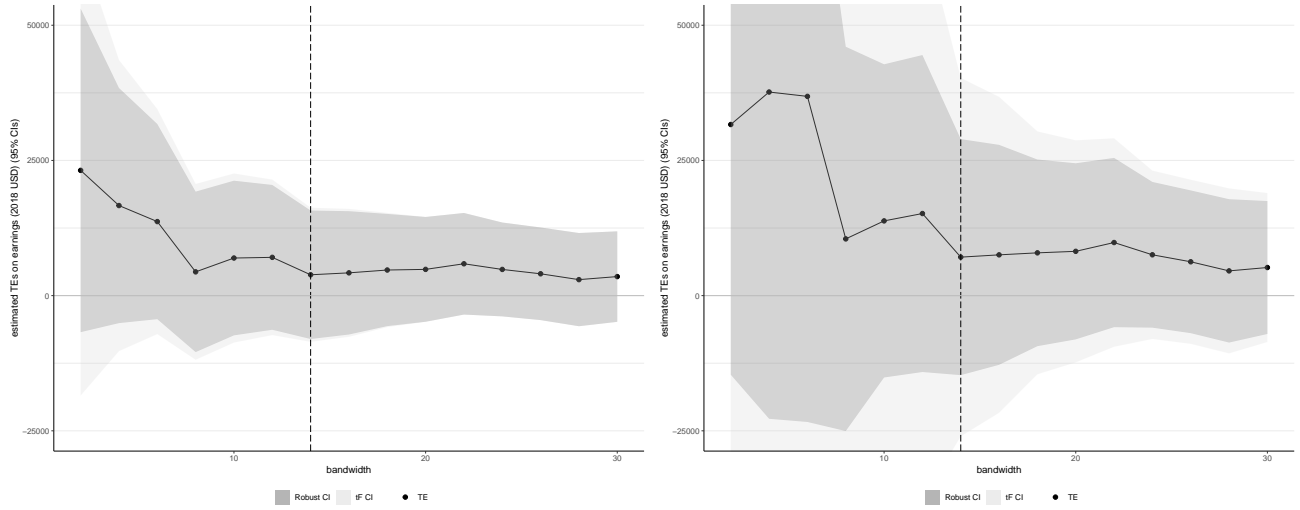
(e) Being a teacher in 4yrs.

Figure A.1: Sensitivity of first-stage RKD F-stats to bandwidth selection, for each stage along



(a) Effect of ever-passing math on $\Pr(\text{employed})$. (b) Effect of ever-passing all 3 on $\Pr(\text{employed})$.

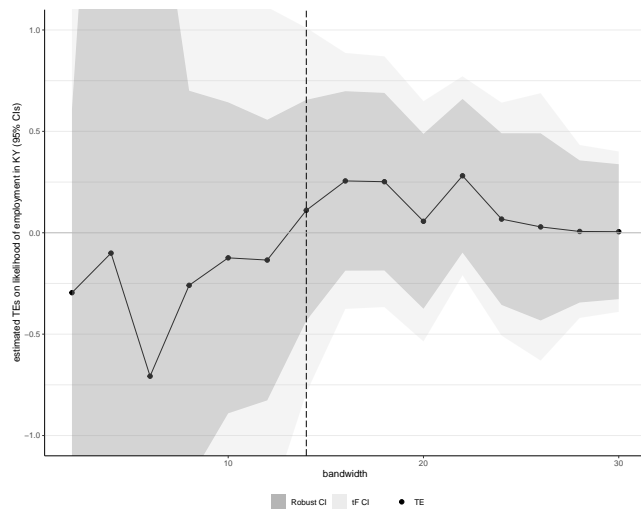
Figure A.2: RD estimates of the effect on the likelihood of being employed (in the state of KY) of passing the Praxis, over bandwidths ranging from 2-30.



(a) Effect of ever-passing math on earnings.

(b) Effect of ever-passing all 3 on earnings.

Figure A.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 $\Pr(\text{employed})$.

Figure A.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?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

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

B Additional Tables

Table B.1: 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 (4.5)	39.6 (9.9)	37.3 (9.6)	40.2 (10.2)
Total hours per week	49.9 (9.8)			
Hourly wage (in 2015 USD)	28.4 (11.9)	30.3 (36.6)	26.0 (28.0)	38.0 (30.1)
<i>% with each working condition</i>				
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

Note: This table presents summary statistics on the jobs held by survey respondents in my survey (Column 1) and in the AWCS (Columns 2–4). The only variable that does not have a direct comparison across columns is “total hours per week,” which was not asked in the AWCS. Standard deviations are shown in parentheses.

Table B.2: Comparison of Working Conditions in Teaching to Jobs in Other Industries

	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
<i>% with each working condition</i>								
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

Note: This table presents comparisons of the jobs held by Kentucky teachers in my survey (Column 1) and various industries in the AWCS in 2015 (Columns 2-8), similarly to Table B.1. The first row includes an additional statistic: the share of Kentucky teachers surveyed who reported that their next-best job would be in the column's industry.

Table B.3: 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
<i>% with each working condition</i>								
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

Table B.4: 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 (9.89)	39.03 (10.09)	37.32 (9.58)	40.23 (10.17)
Mean wage (in 2015 \$)	30.30 (36.55)	25.89 (15.02)	25.99 (27.95)	38.00 (30.10)
<i>% with each working condition</i>				
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 B.5: Regressions of Working Conditions on Demographic Variables (part 1)

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

Table B.6: Regressions of Working Conditions on Demographic Variables (part 2)

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

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

	Teachers (1)	AWCS		
		All workers (2)	College-educ. (3)	Ed/Health, < median wage (4)
Sets own schedule <i>vs. Schedule set by manager</i>	0.080 (0.018)	0.089 (0.007)	0.104 (0.008)	0.071 (0.017)
Telecommute <i>vs. No telecommuting</i>	0.063 (0.015)	0.042 (0.007)	0.069 (0.008)	0.050 (0.018)
Moderate physical activity <i>vs. Heavy physical activity</i>	0.159 (0.021)	0.145 (0.010)	0.168 (0.013)	0.197 (0.026)
Sitting <i>vs. Heavy physical activity</i>	0.119 (0.026)	0.116 (0.010)	0.137 (0.013)	0.142 (0.030)
Relaxed <i>vs. Fast pace</i>	0.093 (0.017)	0.043 (0.007)	0.052 (0.007)	0.042 (0.019)
Choose how do work <i>vs. Tasks well defined</i>	0.050 (0.017)	0.040 (0.007)	0.059 (0.007)	0.050 (0.019)
10 days PTO <i>vs. No days PTO</i>	0.196 (0.024)	0.164 (0.009)	0.158 (0.010)	0.197 (0.025)
20 days PTO <i>vs. No days PTO</i>	0.268 (0.025)	0.230 (0.010)	0.226 (0.011)	0.289 (0.031)
Team-based, evaluate own <i>vs. Team-based, evaluate team</i>	0.099 (0.021)	0.065 (0.010)	0.074 (0.011)	0.067 (0.027)
Work by self <i>vs. Team-based, evaluate team</i>	0.048 (0.022)	0.086 (0.010)	0.073 (0.012)	0.097 (0.026)
Training opportunities <i>vs. Already have skills</i>	0.041 (0.016)	0.054 (0.007)	0.064 (0.008)	0.044 (0.017)
Frequent opp. to serve <i>vs. Occasional opp. to serve</i>	0.043 (0.015)	0.036 (0.007)	0.039 (0.008)	0.050 (0.020)
N	1,662	17,380	9,080	2,710

Note: Column 1 shows estimates of Kentucky teachers' willingness to pay using data from the stated preference experiments. Column 2 shows the same estimates for all workers in the AWCS; Column 3 shows the estimates for college educated workers in the AWCS; and Column 4 shows the estimates for workers in the education or health sectors in the AWCS earning below the median wage in the two sectors. Standard errors, shown in parentheses, are estimated using the delta method and are clustered at the respondent level.

Table B.8: OLS estimates of cross-sectional differences in earnings between teachers and non-teachers among KY college graduates

	<i>Dependent variable: Annual total wages (2018 USD)</i>					
	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*** (71.502)	-6,187.064*** (61.972)	-1,649.273*** (69.929)	-7,252.242*** (66.209)	3,218.845*** (90.019)	2,545.662*** (83.769)
Female		-7,643.454*** (43.785)		-8,566.723*** (50.745)		-4,744.887*** (58.738)
Black		-4,958.967*** (141.189)		-6,057.636*** (168.967)		-3,967.194*** (201.492)
White		463.919*** (114.653)		19.382 (138.385)		1,204.935*** (162.909)
Hispanic		-2,796.032*** (132.946)		-3,140.671*** (155.009)		-1,512.377*** (180.965)
Age (years)		589.780*** (2.885)		654.235*** (3.307)		2,337.656*** (16.385)
Nr quarters employed		14,685.510*** (26.678)				
Holds at least a Master's		14,090.470*** (49.858)		15,369.000*** (56.726)		10,167.130*** (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

C 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 C.6. 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, 2023](#)].

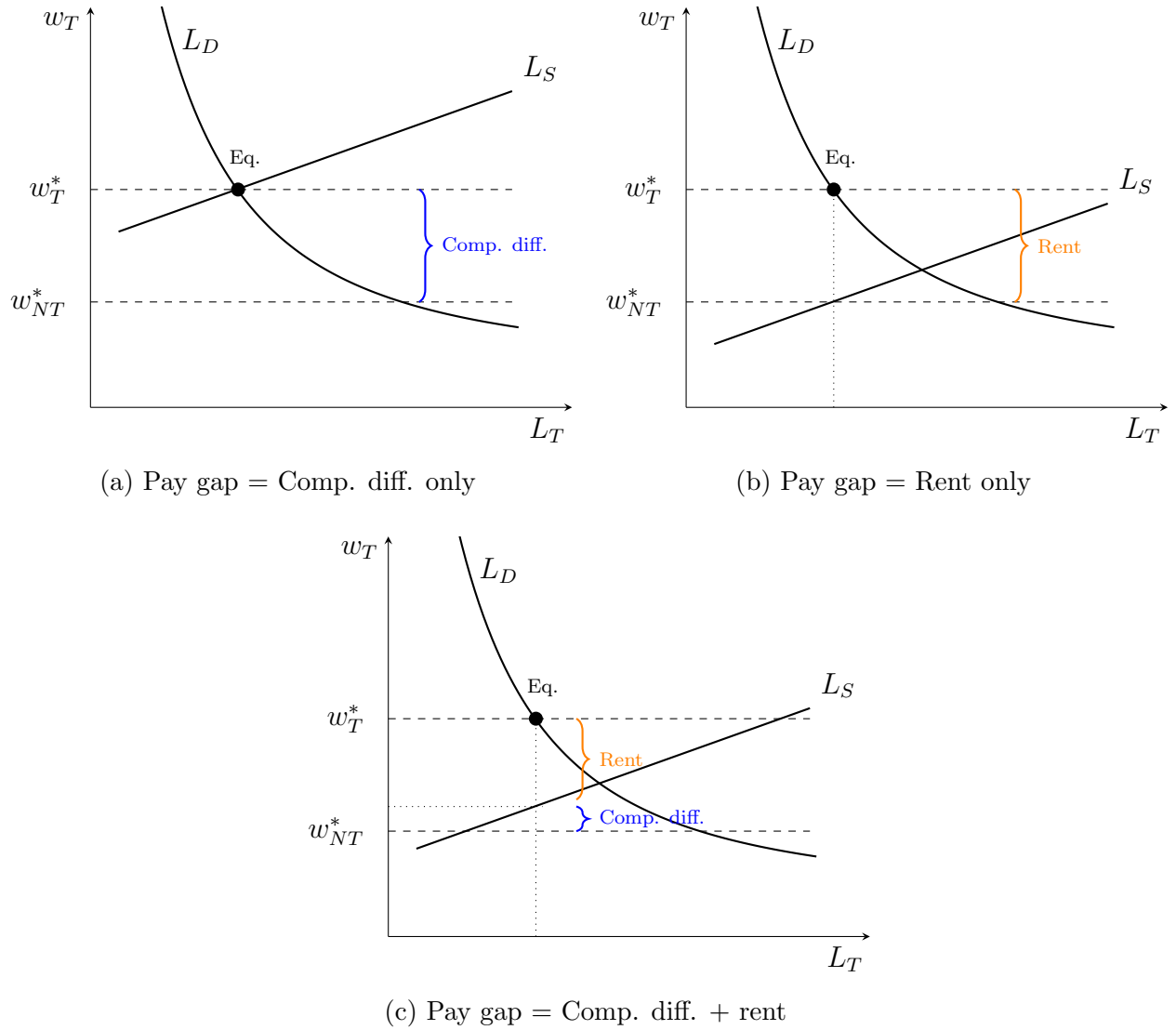


Figure C.6: Illustrative examples showing that the teacher pay gap, the difference between the teaching wage w_T^* and the non-teaching wage w_{NT}^* , can be (a) only a compensating differential, (b) only a rent, or (c) both. In all three cases the pay gap is the same and positive, meaning in equilibrium 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,

as shown in (c).

For completeness, we also illustrate the opposite case where there is a teaching penalty instead of a premium in Figure C.7. 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 mean 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.

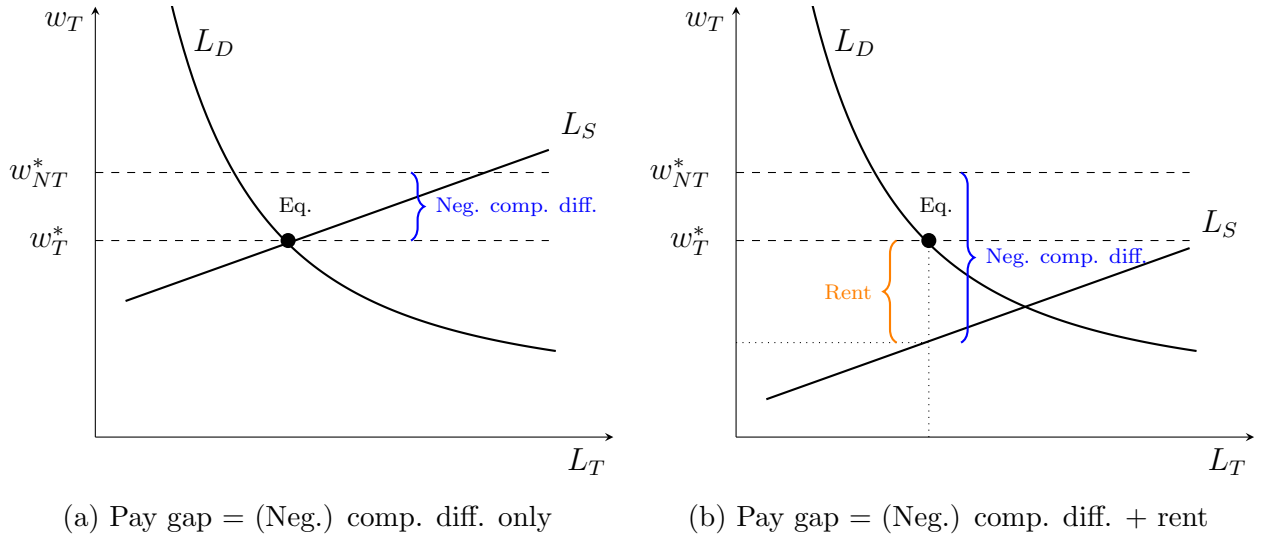


Figure C.7: Illustrative examples showing that under a “teacher penalty,” meaning an equilibrium in which $w_T^* < w_{NT}^*$, the teacher pay gap can be either (a) only a (negative) compensating differential, or (c) the combination of a negative compensating differential and a rent.

D 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.

D.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 ??) that determine whether individuals retake the test or become certified into reduced form objects.

D.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

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

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

$$\begin{aligned}\mathbb{E}[y(C, U) \mid S_1 = s] &= \int \mathbb{I}\{s \geq 0\} [p_c(u)y(1, u) + (1 - p_c(u))y(0, u)] \\ &\quad + \mathbb{I}\{s < 0\}(1 - p_r(s, u))y(0, u) \\ &\quad + \mathbb{I}\{s < 0\}p_r(s, u)\Pr(S_2 < 0 \mid u, R = 1)y(0, u) \\ &\quad + \mathbb{I}\{s < 0\}p_r(s, u)\Pr(S_2 \geq 0 \mid u, R = 1) [p_c(u)y(1, u) + (1 - p_c(u))y(0, u)] dF_{U|S_1=s}(u).\end{aligned}\tag{11}$$

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 $s \geq 0$:

$$\begin{aligned}\lim_{s \rightarrow 0^+} \mathbb{E}[y(C, U) \mid S_1 = s] &= \lim_{s \rightarrow 0^+} \int p_c(u)y(1, u) + (1 - p_c(u))y(0, u) dF_{U|S_1=s}(u) \\ &= \lim_{s \rightarrow 0^+} \int y(0, u) + p_c(u)(y(1, u) - y(0, u)) dF_{U|S_1=s}(u) \\ &= \lim_{s \rightarrow 0^+} \int y(0, u) + p_c(u)(y(1, u) - y(0, u)) \frac{f_{S_1|u}(s)}{f_{S_1}(s)} dF_U(u),\end{aligned}\tag{12}$$

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

Similarly, taking the limit from the left yields

$$\begin{aligned}\lim_{s \rightarrow 0^-} \mathbb{E}[y(C, U) \mid S_1 = s] &= \lim_{s \rightarrow 0^-} \int (1 - p_r(s, u))y(0, u) \\ &\quad + p_r(s, u)\Pr(S_2 < 0 \mid u, R = 1)y(0, u) \\ &\quad + p_r(s, u)\Pr(S_2 \geq 0 \mid u, R = 1) [p_c(u)y(1, u) + (1 - p_c(u))y(0, u)] dF_{U|S_1=s}(u) \\ &= \lim_{s \rightarrow 0^-} \int y(0, u) \\ &\quad + p_r(s, u)\Pr(S_2 \geq 0 \mid u, R = 1)p_c(u)(y(1, u) - y(0, u)) \frac{f_{S_1|u}(s)}{f_{S_1}(s)} dF_U(u).\end{aligned}\tag{13}$$

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

$$\begin{aligned} & \lim_{s \rightarrow 0^+} \mathbb{E}[y(C, U) \mid S_1 = s] - \lim_{s \rightarrow 0^-} \mathbb{E}[y(C, U) \mid S_1 = s] \\ &= \int (y(1, u) - y(0, u)) \left[\lim_{s \rightarrow 0^+} p_c(u) - \lim_{s \rightarrow 0^-} p_c(u) p_r(s, u) \Pr(S_2 \geq 0 \mid u, R = 1) \right] \frac{f_{S_1|u}(0)}{f_{S_1}(0)} dF_U(u). \end{aligned} \quad (14)$$

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

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

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

$$\begin{aligned} \lim_{s \rightarrow 0^+} \Pr[C = 1 \mid S_1 = s] &= \lim_{s \rightarrow 0^+} \int p_c(u) \frac{f_{S_1|u}(s)}{f_{S_1}(s)} dF_U(u), \text{ and} \\ \lim_{s \rightarrow 0^-} \Pr[C = 1 \mid S_1 = s] &= \lim_{s \rightarrow 0^-} \int p_c(u) p_r(s, u) \Pr(S_2 \geq 0 \mid u, R = 1) \frac{f_{S_1|u}(s)}{f_{S_1}(s)} dF_U(u). \end{aligned} \quad (16)$$

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

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

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

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

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 ,

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

where the weights are given by

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

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 F: 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.⁵⁶

D.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.

⁵⁶Note that, at a glance, the interaction between $p_c(u)$ and $(1 - \lim_{s \rightarrow 0^-} p_r(s, u) \Pr(S_2 \geq 0 \mid u, R = 1))$ 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

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

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:

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

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

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

and from the left hand side, we have

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

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

$$\begin{aligned}
& \lim_{s \rightarrow 0^-} \frac{d\mathbb{E}[y(C, U) \mid S_1 = s]}{ds} - \lim_{s \rightarrow 0^+} \frac{d\mathbb{E}[y(C, U) \mid S_1 = s]}{ds} \\
&= \int [y(1, u) - y(0, u)] p_c(u) \times \\
&\quad \left\{ \lim_{s \rightarrow 0^+} \Lambda(s) - \Pr(S_2 \geq 0 \mid u, R = 1) \lim_{s \rightarrow 0^-} \left[\frac{\partial p_r(s, u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} + p_r(s, u) \Lambda(s) \right] \right\} dF_U(u) \\
&= \int [y(1, u) - y(0, u)] \times \left(\lim_{s \rightarrow 0^+} p_c(u) \Lambda(s) \right. \\
&\quad \left. - \lim_{s \rightarrow 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) p_r(s, u) \Lambda(s) \right. \\
&\quad \left. - \lim_{s \rightarrow 0^-} \Pr(S_2 \geq 0 \mid u, R = 1) \frac{\partial p_r(s, u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} \right) dF_U(u) \\
&= \underbrace{\int [y(1, u) - y(0, u)] \cdot \Lambda(0) \cdot \left(\lim_{s \rightarrow 0^+} p_c(u) - \lim_{s \rightarrow 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) p_r(s, u) \right) dF_U(u)}_{(a)} \\
&\quad - \underbrace{\int [y(1, u) - y(0, u)] \left(\lim_{s \rightarrow 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) \frac{\partial p_r(s, u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} \right) dF_U(u)}_{(b)}. \tag{24}
\end{aligned}$$

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

$$\int [y(1, u) - y(0, u)] \cdot \frac{f_{S_1|u}(0)}{f_{S_1}(0)} \left[\frac{f'_{S_1|u}(0)}{f_{S_1|u}(0)} - \frac{f'_{S_1}(0)}{f_{S_1}(0)} \right] \left(\lim_{s \rightarrow 0^+} p_c(u) - \lim_{s \rightarrow 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) p_r(s, u) \right) dF_U(u) \tag{26}$$

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

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

and

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

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

$$\int [y(1, u) - y(0, u)] \cdot \left[\frac{f'_{S_1|u}(0)}{f_{S_1|u}(0)} - \frac{f'_{S_1}(0)}{f_{S_1}(0)} \right] dF_U(u) \quad (27)$$

$$\begin{aligned} & \times \int \left(\lim_{s \rightarrow 0^+} p_c(u) - \lim_{s \rightarrow 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) p_r(s, u) \right) \frac{f_{S_1|u}(0)}{f_{S_1}(0)} dF_U(u). \\ & = \int [y(1, u) - y(0, u)] \cdot \left[\frac{f'_{S_1|u}(0)}{f_{S_1|u}(0)} - \frac{f'_{S_1}(0)}{f_{S_1}(0)} \right] dF_U(u) \\ & \times \left(\lim_{s \rightarrow 0^+} \Pr[C = 1 \mid S_1 = s] - \lim_{s \rightarrow 0^-} \Pr[C = 1 \mid S_1 = s] \right), \end{aligned} \quad (28)$$

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. $\text{Corr}(p_c(u), p_r(s, u)) > 0$: 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),

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

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:

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

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

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

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 ,

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

The weights are given by

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

We can interpret $\varphi_{RKD}(u)$ as weights because $\varphi_{RKD}(u)$ 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 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).

E 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.

E.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

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

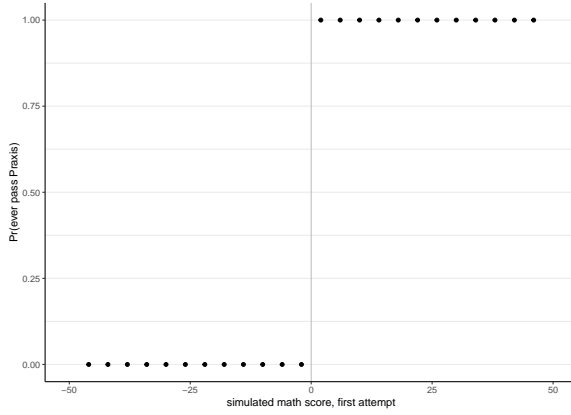
where Y denotes the labor market outcome of interest, C is an indicator for getting certified, S_1 and S_2 denotes the two possible scores on the Praxis on various attempts where S_2 is only observed if the individual chooses to retake (i.e. $R = 1$), and U captures individual heterogeneity. Individuals 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: $C = C(\{S_1, S_2\}, R, \xi)$,⁵⁷ 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 .

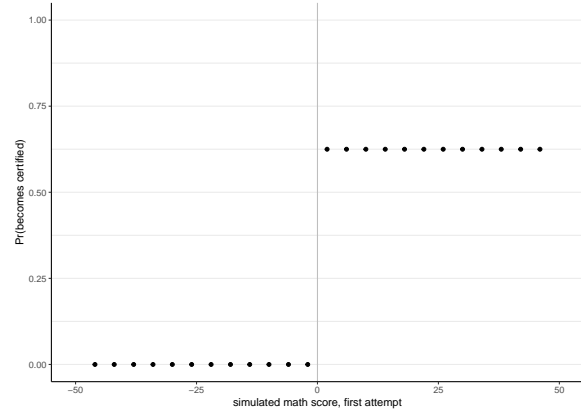
Suppose for now that retaking is not allowed. The top row of [Figure E.8](#) shows simulations of what the first stage relationships between first-attempt scores (S_1) and the likelihood of passing the test or becoming certified would look like.⁵⁸ Those who do not pass can never pass

⁵⁷The 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.

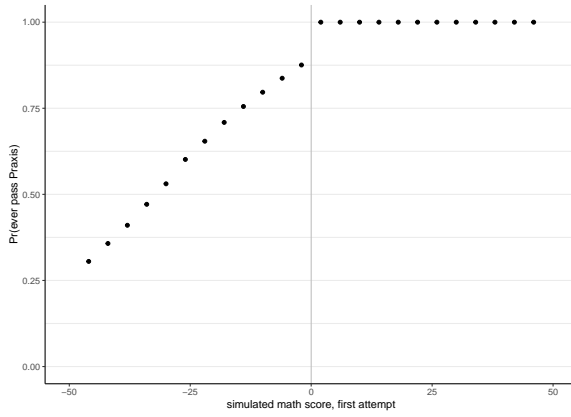
⁵⁸We assume that among those who pass, high-scorers are no more likely than barely-passers to enroll in and complete an EPP. However, as we know from the descriptives in [Section ??](#), many who pass the Praxis do not continue through an EPP. We therefore choose parameters in our simulation to replicate this observation, with



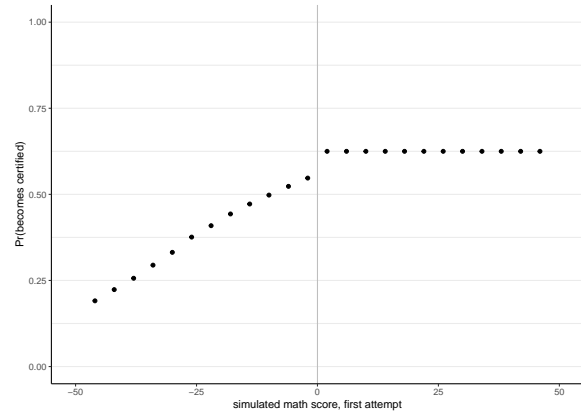
(a) Simulated $\Pr(\text{pass})$, one-shot



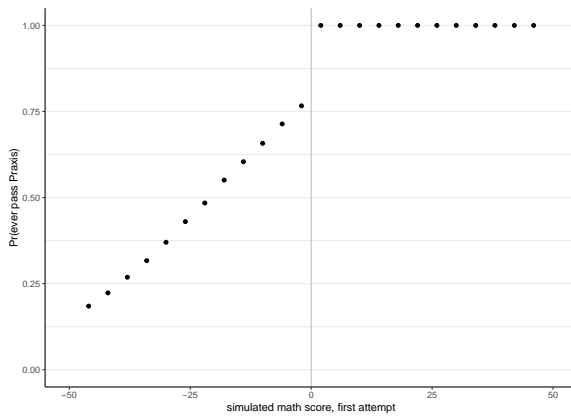
(b) Simulated $\Pr(\text{certified})$, one-shot



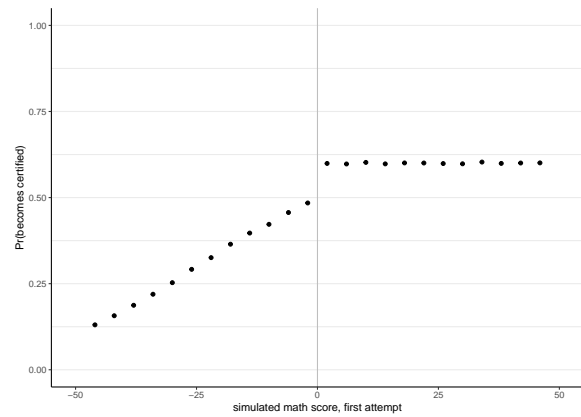
(c) Simulated $\Pr(\text{pass})$, with automatic retaking



(d) Simulated $\Pr(\text{certified})$, with automatic retaking



(e) Simulated $\Pr(\text{pass})$, with behavioral retaking



(f) Simulated $\Pr(\text{certified})$, with behavioral retaking

Figure E.8: 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.

E.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 E.8 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.

E.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 E.8 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.

E.4 Reconciling the model and data

The last row of Figure E.8 can be contrasted with the actual first-stages in the data shown in Figure 4. 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.

F 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 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 they 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 D 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.

G Additional Survey Details

Implementation. 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.

Trick questions. 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.”

Presenting the wages. 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.

Limiting cases where one jobs dominates another. 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.