IT'S NOT (JUST) ABOUT THE MONEY: PAY AND THE VALUE OF WORKING CONDITIONS IN TEACHING*

Carolyn Tsao†

Abstract

This paper quantifies the extent to which teachers earn rents by estimating both the pay gap and the working conditions gap between teaching and teachers' next-best jobs. Using a regression kink design and event studies applied to administrative data from Kentucky, I find that teaching pays a premium of around $20,000 per year over teachers' outside options. However, a choice experiment reveals that teachers—especially inexperienced ones—would pay a large share of their salary for the better working conditions available elsewhere. Combining these estimates, experienced teachers earn a moderate 16% rent, while inexperienced teachers earn no rent at all.

JEL codes: J31, J32, J45, J81, I21

I. Introduction

Teaching is the largest public sector occupation in the US, and teacher salaries account for over half of public school spending, totaling over $500 billion in 2020–21. Yet the extent to which this spending reflects rents— versus compensating differentials for relatively undesirable working conditions— is not well understood. Public school teaching is associated

*I am indebted to David Lee, Alex Mas, and Ilyana Kuziemko for their guidance, and to Angie Tombari and KYSTATS for providing data access and invaluable support. I thank Lukas Althoff, Zach Bleemer, Leah Boustan, Pauline Carry, Ellora Derenoncourt, Sharada Dharmasankar, Jacob Dorn, Adam Kapor, Pat Kline, Jesse Rothstein, Karl Schulze, Jesse Silbert, Christiane Szerman, Camille Terrier, and seminar participants at the CESifo/ifo Junior Workshop on Economics of Education, the Becker Friedman Institute Women in Applied Microeconomics Conference, Princeton University, Toronto Metropolitan University, Microsoft Research, University of Georgia, Tulane University, Williams College, Boston University, The Ohio State University, the London School of Economics, the University of Illinois Urbana-Champaign, the Brookings Institute, and the University of California Berkeley for their insights and suggestions. I am grateful for the financial support of the Industrial Relations Section, the Prize Fellowship in Social Sciences, and the Program for Research on Inequality at Princeton, and the Social Sciences and Humanities Council of Canada (SSHRC). All errors are my own. Small change. The views and opinions expressed in this paper are my own and do not necessarily reflect the views of KYSTATS.

†Microsoft Research (300 Lafayette St, New York, NY, 10012). Email: carolyntsao@microsoft.com

with job stability and strong unions, and existing studies find that these institutions may raise compensation without improving student achievement (Hoxby, 1996), consistent with rents. At the same time, widespread reports of stress, burnout, and teacher shortages—with over 50% of teacher strikes in the past decade demanding better working conditions (Lyon, Kraft, and Steinberg, 2024)— suggest that teaching conditions may be sufficiently undesirable to warrant a substantial pay premium (Sutcher, Darling-Hammond, and Carver-Thomas, 2016).[1] Either way, the extent to which teachers earn rents versus compensating differentials has direct implications for education spending: large rents would mean that a significant share of teacher compensation comes at the expense of other educational investments, while large compensating differentials would raise the question of whether improving working conditions may be a more cost-effective way to attract and retain teachers.

However, distinguishing rents from compensating differentials requires overcoming two empirical challenges: credibly identifying what teachers would earn in their next-best jobs, and measuring how working conditions in teaching compare to those in teachers' next-best jobs. While a number of studies compare teachers' pay to that of observationally similar workers (e.g., Allegretto and Mishel, 2018), such comparisons do not actually identify rents because workers select into occupations for unobserved, idiosyncratic reasons. For instance, if teachers are negatively selected on earnings potential, it is possible that they earn less than observationally similar workers (a negative observational pay gap) and earn more in teaching than in their next-best jobs (a positive rent). Furthermore, while recent work has made progress on documenting variation in job attributes across the U.S. (Maestas et al., 2023), there is no systematic data on how working conditions differ across occupations, making it difficult to assess how conditions in teaching compare to other jobs.[2]

This paper addresses both challenges. Using quasi-experimental variation in who becomes a teacher in Kentucky and a new survey on how working conditions vary across jobs, I quantify the extent to which teachers earn rents by separately estimating the gap in pay and the gap in the value of working conditions.[3] I estimate the pay gap using a fuzzy regression kink design and an event study combined with a novel panel dataset covering the near-universe of workers in Kentucky. The data track workers' earnings and employment across jobs over time, and— crucially— identify which individuals express interest in becoming teachers and whether they follow through. I estimate the working conditions gap

[1] For example, see Eliza Fawcett and Jacey Fortin, "How Bad Is the Teacher Shortage? Depends Where You Live." *The New York Times*, June 21, 2023.

[2] The most relevant dataset that documents differences in working conditions across U.S. workers, the American Working Conditions Survey (AWCS), contains few teachers (Maestas et al., 2023).

[3] Jäger et al. (2024) refer to these as the wage-based and non-wage-based components of the worker rent, following Lamadon, Mogstad, and Setzler (2022).

by surveying KY teachers and other workers, and using choice experiments to elicit their willingness to pay for a wide range of job attributes. Importantly, my survey replicates the instrument in the American Working Conditions Survey (Maestas et al., 2023), adapted to focus on teaching and the industries identified as teachers' next-best options in the quasi-experimental analysis. This allows me to directly compare working conditions in teaching to those in the jobs teachers actually move to. Taken together, my approach provides the most comprehensive estimate of an occupation-specific rent in the U.S. to date.

My main finding from the quasi-experimental analyses is that teaching offers a large pay premium compared to teachers' next-best jobs. The RK shows that for individuals who score near the teaching entry exam cutoff on their first attempt, their next-best in-state options four years later pay an estimated \$18,000–21,000 less per year than teaching does, and fall in one of four industries: education, health care, retail, or accommodation and food services. Similarly, event studies around teacher exits show that exiting teachers earn an average of \$20,000–25,000 less per year four years after exit and tend to leave to the same four industries. The estimates imply a large pay gap of 33%–40% of a KY teaching salary.

My main finding from the survey is that for teachers, teaching offers *less* desirable conditions than teachers' next-best jobs. While teaching does offer unique attributes that teachers value, such as working with children and "summers off," the job also does not offer important attributes that teachers value and that are common elsewhere, such as the absence of hostility and a relaxed pace of work. As a result, I find that experienced teachers in KY are willing to pay 17% of their salary to switch to their next-best job solely for more desirable working conditions, while inexperienced teachers are willing to pay much more—51% of their salary—to make the same switch. The difference in magnitudes is driven by inexperienced teachers facing less desirable job conditions than experienced teachers: in particular, inexperienced teachers do not have high job security ("tenure"), and they face more hostility and less administrative support than experienced teachers do in conflicts with parents and children.

Overall, my findings suggest that teachers earn a large pay premium that is significantly offset by the job's relatively undesirable working conditions, resulting in experienced teachers earning rents but inexperienced teachers earning no rents at all. My estimates suggest that experienced teachers earn a rent of 16% of their salary—surprisingly similar in magnitude to recent estimates of worker rents in the entire labor market (Lamadon, Mogstad, and Setzler, 2022; Jäger et al., 2024). However, for inexperienced teachers, my results suggest that their pay premium functions entirely as a compensating differential, consistent with the fact that younger teachers experience the highest turnover rates. Extending my results from KY to other states suggests that over \$120 billion is being spent annually to compensate teachers for relatively 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.

This paper primarily contributes to two strands of literature. The first strand investigates the extent to which workers earn rents. Most work in this area focuses on quantifying worker rents in the entire labor market (Manning, 2011; Lamadon, Mogstad, and Setzler, 2022; Jäger et al., 2024). However, few papers focus on measuring worker rents in specific industries or occupations, which may differ from the average and be more tangible for wage regulation, especially in public sector jobs.[4] This paper proposes an approach for quantifying an occupation-specific rent that can be implemented by leveraging institutional features (e.g., cutoffs on certification tests) to identify next-best jobs and using low-cost choice experiments to estimate workers' preferences for bundles of job attributes.[5] I apply the method to estimate the teaching rent, providing the most comprehensive estimate of an occupation-specific rent to date.

The second strand studies teacher labor markets. A significant body of work examines how teaching compensation and preferences over teaching-specific job attributes explain teacher sorting across schools (Boyd et al., 2005; Biasi, Fu, and Stromme, 2025; Bates et al., 2025; Johnston, 2025; Wiswall and Zafar, 2018).[6] Yet relatively few studies focus on documenting what teachers' *outside option* compensation and preferences over *outside option* working conditions are, both of which are key factors in workers' decision to select into teaching and are particularly important to understand given concerns about teacher shortages.[7] One reason for this dearth of evidence is that it is challenging to identify workers' next-best options in general. This paper provides the first quasi-experimental evidence characterizing teachers' next-best jobs in terms of pay, working conditions, and industry/occupation. The results suggest that relatively undesirable working conditions are a key reason why some who consider teaching may instead choose another occupation.

[4] A notable exception is Kroft et al. (2025), who focus on the construction industry and use features of the industry to estimate an equilibrium model of the labor and product market to evaluate rents. A large literature has sought to understand why wage premia differ across industries (e.g., Card, Rothstein, and Yi (2024)).

[5] A growing body of work uses, and demonstrates the validity of using, choice experiments to understand workers' preferences for various job attributes, from dignity at work (Dube, Naidu, and Reich, 2025) to workplace flexibility (Drake, Thakral, and Tô, 2022).

[6] For instance, a number of papers estimate teacher preferences over student characteristics, using teacher assignment data (Boyd et al., 2005; Biasi, Fu, and Stromme, 2025), application data (Bates et al., 2025), or choice experiments (Johnston, 2025; Wiswall and Zafar, 2018).

[7] The most closely related work examines how *changing* outside options affect selection into teaching. See, for example, Nagler, Piopiunik, and West (2020), who leverages variation in outside options caused by business cycles, and Bacolod (2007), who uses women's changing labor market opportunities over the twentieth century.

My study is most closely related to Maestas et al. (2023), Johnston (2025), and Jäger et al. (2024). Maestas et al. (2023) embed a choice experiment in the AWCS to assess how job attributes and preferences for working conditions vary across U.S. workers. I survey KY teachers using two adapted versions of the AWCS, allowing me to benchmark my findings for teachers to those of Maestas et al. (2023), and crucially, to compare teaching to teachers' next-best jobs and compare teachers' preferences to those of other workers. Johnston (2025) is the first to use a choice experiment to assess how teachers value teaching compensation and a wide range of job attributes, including key general attributes such as managerial (i.e., principal) support. I extend Johnston's approach to examine how much teachers value the bundle of attributes not just in teaching, but also in their next-best jobs, allowing me to estimate the working conditions gap. My estimates of a large compensating differential in teaching corroborate Johnston's conclusion that "shifting resources into compensation and amenities that teachers prefer relative to cost" may be an effective way to "improve the appeal of teaching." Finally, Jäger et al. (2024) use a survey to elicit individual workers' rents and the pay gap between their job and their next-best option in the German labor market. These estimates allow them to assess the accuracy of workers' beliefs about their outside options. I employ a related approach that uses the same rents definition and use a survey to back out a component of the rent. That said, my approach differs in two key ways. First, I use the survey to back out the non-wage component of the rent, rather than the wage component. Second, I identify workers' next-best options using not only job switches, but also quasi-experimental variation in who becomes a teacher, made possible by leveraging institutional features of the occupational licensing process.

The rest of the paper proceeds as follows. Section II describes how I estimate the pay gap, from data and methods to results. Similarly, Section III describes how I estimate the gap in the value of working conditions. Section IV discusses the implications of my estimates, reconciles them with previous findings, and concludes.

II. Estimating the Pay Gap

The central object of interest is the rent from teaching—the extent to which a teacher's total compensation exceeds what she would earn in her next-best option, accounting for differences in both pay and working conditions:

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

The first term is the pay gap: the difference in earnings between teaching and the next-best job. The second term is the working conditions gap: the monetized difference in non-pecuniary attributes. The remainder of this section estimates the pay gap using two quasi-experimental strategies; Section III estimates the working conditions gap using a choice experiment.

II.A. Identification Challenge

The main identification challenge in pinning down the pay gap between teaching and teachers' next-best job options is selection into teaching: the decision to become a teacher or earn teaching qualifications 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 exams required for entry into KY's teacher preparation programs. 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 describe the features of the teacher certification process most relevant for identification and the data I use before detailing the empirical strategies and results.

II.B. Key Features of the Teacher Certification Process

In most U.S. states, including KY, the traditional path to becoming eligible to teach in public schools 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.[8]

Rather than design their own exams, most EPPs use standardized tests designed by ETS, collectively referred to as the "Praxis" exams.[9] I focus on the Praxis CASE ("Core

[8]The majority of public school teachers today earned their teaching certificates through the traditional pipeline. However, over the past decade, political pressures and concerns of labor shortages in teaching have led to increased demand and supply of "alternative route certifications," which allow individuals to bypass the traditional pipeline to become "emergency certified."

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

Academic Skills for Educators"), which tests core proficiency in math, reading, and writing, and is a common entry requirement for EPPs. Importantly, the CASE math exam tests middle-school level mathematics, not pedagogical knowledge, and unlike the GRE or SAT, the Praxis is only ever used to test individuals for entry into or exit from teaching preparation programs.[10]

In KY, individuals must pass the Praxis CASE to be admitted to an EPP. Most test-takers who attempt one of the three exams choose to sit all three tests at once, a strategy that is also incentivized by ETS' pricing scheme: test-takers can take one exam for $90 or all three 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.[11, 12]

The way the test is run makes it difficult for test-takers to target a specific score. Like other ETS exams, the Praxis CASE is multiple-choice, timed, and administered on computers in monitored exam rooms with no Internet access. Each exam question carries its own weight corresponding to its difficulty level, which is not revealed to the test-taker, but instead is used to convert the raw scores into standardized scores which range from 100 to 200. Test-takers thus see their scores immediately at the end of the exam.

Between 2013 and 2023, KY maintained common cutoff scores on the Praxis CASE across all EPPs, with no changes in ETS' exam design.[13]

II.C. Matched Employer-Employee Data

To estimate the pay gap, I construct a novel administrative panel dataset covering the near-universe of workers in KY. The panel allows me to track individuals who take the Praxis through to their future employment and to observe worker flows in and out of public schools.

The main variables of interest are individual earnings, industry, Praxis scores, and em-

[10] See state department of education and ETS websites. States vary in which Praxis exams they require and in what constitutes a passing score. 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.

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

[12] Some programs accept students on the condition that they pass the CASE during the program. However, even if conditional admittance is allowed, students must pass all three tests in order to graduate.

[13] Prior to 2012, most EPPs in KY 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. 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.

ployment as a teacher. Earnings and industry come from the KY Unemployment Insurance (UI) database, which contains quarterly earnings and 6-digit industry NAICS codes associated with each source of income, for workers between the ages of 18 to 60.[14] Individual Praxis scores come from the Educator Professional Standards Board (EPSB): for each person, these data include the date and scores of all Praxis attempts and annual indicators of the type (e.g. full, emergency, temporary) and validity of their teacher certification status. Indicators for employment in the public school system come from the KY Longitudinal Data System, a database maintained by KYSTATS that contains person-year records on all staff in the public school system provided by the KY Department of Education (KDE) and MUNIS. These data include information on each individuals' district and school of employment, occupation (e.g. teacher or principal), and links to classrooms and student outcomes, allowing me to observe when individuals move between schools or districts or change occupations within public schools. I also construct indicators for postsecondary enrollment, completion, or Pell grant receipt in-state by merging in data from the Council on Postsecondary Education (CPE), and indicators for whether an individual is listed as a parent on a birth certificate for a newborn child from the Vital Statistics records.

The final sample includes the union of all individuals who ever attempted the Praxis CASE in KY between 2008 and 2022 and all individuals who ever appear as employed in the KY UI data system between 2009 until 2022. Demographic information such as gender, race, and ethnicity exists for nearly all individuals.

There are two notable limitations of the UI data that are typical across state UI systems: the lack of information on hours and out-of-state employment, and on occupations outside of the public school system. I refer to individuals who are not observed in KY employment records as "non-employed." To learn about teachers' next-best occupations, I supplement the administrative data with the matched sample in the Current Population Survey (CPS), which contains a nationally representative, two-period panel with observations spaced one year apart. From 2009 to 2023, nearly 5,000 individuals in the matched CPS report being a teacher in the first round but holding a different occupation in the second round.

Table I shows sample means describing all individuals who attempted the Praxis CASE for the first time between 2013 and 2017 in KY.[15] Panel A shows key demographic characteristics. First-time test-takers tend to be in their third year of college and are mostly white women, reflecting the demographic composition of teachers in the state.[16] One-fifth hold a

[14]I create a crosswalk between firm names under the education NAICS code and the official list of private schools on the KY Department of Education (KDE) website to identify private schools where possible.

[15]I focus on individuals who took the exam between 2013 and 2017 because the Praxis CASE was last revised then, and because the full panel ends in 2022, meaning I can only observe employment outcomes four years out for individuals who took the exam before 2018.

[16]Although the mean age is 24, the median age is 21, consistent with our understanding that individuals

Pell grant, and two-thirds are working in the year before the exam.

Panel B summarizes statistics related to Praxis performance and retaking behavior. The 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, Panel C 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, while 87% of individuals who get certified work as a teacher shortly thereafter.

II.D. Empirical Strategy I: Regression Kink Design

My main empirical strategy for estimating the pay gap is a fuzzy RK design. The RK leverages variation in who becomes a teacher, generated by the cutoff on the Praxis exam, to identify the treatment effect of becoming a teacher on earnings. The key insight is that, in this environment, the treatment effect is by definition the difference in potential earnings between teaching and their next-best job for interested teachers—or, in other words, the pay gap of interest.

Why an RK and not an RD? Figure I shows the first-stage relationship between first-attempt test scores and whether test-takers eventually pass, earn teacher certification, or become teachers. Visually, there is a strong kink and no discontinuity in the likelihood of becoming certified or a teacher, motivating the use of a fuzzy RK design rather than the fuzzy RD designs used in related work (e.g., Jepsen, Mueser, and Troske, 2016).

The kink arises because retaking is allowed and commonplace: among those who initially fail, those who score closer to the cutoff are more likely to retake the exam than those who score further away (Figure II), smoothing out any discontinuity in the first stage. The slope in retaking probability is non-trivial: those who score a 146 or 144 are around 5 p.p. less likely to retake than those who score a 148, despite the scores differing by only one or two correct multiple-choice answers. This pattern is consistent with an "encouragement" effect—receiving a higher score signals a higher chance of passing, inducing retaking—rather

are most likely to take the Praxis CASE in the third year of college.

than selection, as the relationship persists after controlling for baseline characteristics including gender, race, previous earnings, age, and Pell grant status.[17]

The lack of discontinuity rules out a fuzzy RD design. However, in cases where there is a kink and no discontinuity, a fuzzy RK design can identify the treatment effect of interest when certain conditions hold (Dong, 2018).[18]

Identification. The fuzzy RK design requires four assumptions. First, test-takers cannot precisely manipulate their test scores, so that there is as-good-as-random assignment around the cutoff. The distribution of first-attempt math scores does not exhibit a jump or kink at the cutoff (Figure III(a,b)).[19,20] Furthermore, individuals on either side of the cutoff are balanced on a large set of pre-determined characteristics.[21] I verify balance visually, using a covariates index following Card et al. (2015) (Figure III(c,d)), and via paired t-tests within 10–15 points of the cutoff. All three tests suggest balance.

Second, there is virtually no discontinuity in the first-stage relationship between first-attempt test scores and becoming a teacher. As shown in Dong (2018), if there is a kink but no discontinuity, a fuzzy RK (but not a fuzzy RD) will identify a weighted average treatment effect. The first-stage relationships in Figure I support this assumption.

Third, the instrument is relevant: Figure I shows that the slope of the likelihood of becoming a teacher sharply decreases at the cutoff, and first-stage estimates confirm this is statistically significant.

Fourth, the exclusion restriction requires that the change in slope at the cutoff affects earnings only through its effect on the likelihood of becoming a teacher. The change in slope arises from three features: (1) quasi-random assignment above or below the cutoff, which varies the cost of getting certification, (2) the option to retake below the cutoff, and (3) the encouragement effect inducing higher retaking rates closer to the cutoff. Quasi-random assignment is excluded by construction. The main question is whether retaking itself affects earnings through channels other than becoming a teacher—for instance, if studying for a

[17] This pattern is not unique to Kentucky teacher licensing. In another study leveraging discontinuities generated by cutoffs in teacher licensing exams in Texas, Deneault, Riehl, and Zou (2026) similarly find an encouragement effect—what they call "only a modest discouragement effect"—resulting in a small labor supply response at the cutoff margin.

[18] I provide a detailed walkthrough of the reasoning behind why the first stage exhibits a kink rather than a discontinuity, along with simulations, in Supplementary Materials I.B. Following Dong (2018), I also show that if there is both a discontinuity and a kink, a fuzzy RD (but not a fuzzy RK) will identify a weighted average treatment effect. See Supplementary Materials I.A for formal proofs.

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

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

[21] The full list of baseline characteristics used in balance tests is reported in the notes to Figure III.

retake built human capital or motivation that raised earnings potential independently. This is unlikely given that the minimum wait time for retaking is 28 days and the Praxis tests middle-school math and reading skills.

If these assumptions hold, the interaction of whether one passes on their first attempt and how far they score from the cutoff serves as a valid instrument for becoming a teacher, which is precisely what the fuzzy RK leverages via 2SLS.

Estimation. 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:

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

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, particularly as a teacher.

Estimand interpretation. Like the RD estimand, the RK estimand can be interpreted as a weighted average treatment effect across the distribution of individual types (DiNardo and Lee, 2011). I show that in settings with retaking, the RK estimand is a weighted average treatment effect for what I call "re-taking compliers": individuals whose choice to retake is responsive to their first-attempt score.[22] The weighted average excludes individuals who would never retake or who would always retake regardless of their score. This interpretation is intuitive: because the kink in the first stage is driven by the retaking process, the estimand is identified off individuals whose retaking decision responds to their quasi-random assignment near the cutoff.

II.E. Regression Kink Estimates of the Pay Gap

First-stage and reduced form plots. Figure I shows the first-stage relationships between the first-attempt test scores and the likelihood of reaching any point of the teacher

[22] In Supplementary Materials I.C, I re-cast the classical complier/always-taker/never-taker/defier framework to account for retaking behavior, and show formally that the RK estimand identifies a weighted average treatment effect for re-taking compliers. See also Supplementary Materials I.A for identification proofs.

pipeline: ever passing the Praxis CASE, enrolling in an EPP, getting certified, and becoming a teacher.

At the earliest stage of the 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.[23]

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 IV. There appears to be no discontinuity or kink at the test score cutoff in the likelihood that individuals are employed, suggesting no effect of engaging in the teacher pipeline on being in the workforce. However, there is a weak kink at the cutoff in short-run earnings.

Treatment effect estimates. 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, Cattaneo, and Farrell (2020) in Table II and for the full range of bandwidths in Figure V.

I begin with 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 II, Panel A and Supplementary 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 these null effects, in the following analysis of treatment effects on earnings, I restrict the sample to those who are employed (i.e., those with non-zero earnings).

Figure V shows the 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, Cattaneo, and Farrell (2020), indicated by a vertical dotted line at 14.[24] 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 of 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 driving the overall effect on earnings.

[23] Due to our small sample sizes, the corresponding F-stats are small (Figure A.1).

[24] Consequently, the estimates shown in Table II, Panels B and C are for a fixed bandwidth of fourteen, at which the estimates begin to stabilize.

Interpretation. The RK results indicate that for retaking compliers, there is a large positive premium of getting certified: around \$18,000–\$21,000 per year. Because 87% of those who become certified also become teachers, becoming a teacher appears to be the key driver of this effect—consistent with the expectation that the skills developed along the pipeline pay off most for those who complete their training. Implicit in this interpretation is the monotonicity assumption that "retaking never-takers" do not exist—that is, no one would retake only if they had already passed. The RK estimate also excludes "retaking always-takers," who would retake regardless of their score. This is reassuring: because always-takers may be those with the largest expected return to teaching, their exclusion means the RK estimate is unlikely to be inflated by these high-return individuals.

Robustness checks. Because of small sample sizes and the resulting weak first-stages in the RK, I present several weak-instrument-robust confidence intervals in addition to the classic robust 2SLS standard errors: Anderson-Rubin CIs, tF CIs (Lee et al., 2022), and VtF CIs (Lee et al., 2024). The tF intervals are the widest, encompassing zero for bandwidths less than fourteen, but my preferred VtF intervals closely correspond with the regular CI estimates.[25] 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.

II.F. Empirical Strategy II: Event Study

The second strategy I use to estimate the pay gap is an event study of earnings changes around teacher exits, using the KY administrative data and the matched CPS. While the RK identifies the pay gap for marginal entrants, the event study identifies it for experienced teachers who leave the profession— a different population, providing a complementary estimate. One caveat is that exit 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.

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. Aligning their dates of exit at $t = 0$, I estimate the following specification via

[25] The results are also not sensitive to omitting Northern KY University from the sample, which is particularly known to send its graduates across KY's border to work in Cincinnati, Ohio, potentially affecting the earnings outcomes I observe in KY UI data.

OLS:

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

where the coefficients α_τ capture the adjusted earnings difference between year $t = -1$ and year τ, and $X_{i,t=-1}$ are controls for 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). The CPS serves two roles: it allows me to check whether the administrative pay gap reflects data-specific features such as the coverage of only within-state employment, and it contains data on occupations and hours, allowing me to examine the types of jobs teachers leave for.

II.G. Event Study Estimates of the Pay Gap

Figure VII illustrates the raw earnings trajectories for teachers who leave in 2010, alongside those who stay through 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. Those who leave to non-employment mostly do not return to teaching or other employment and thus earn the least.[26]

Figure VI shows regression-adjusted estimates, controlling for gender, race, age, years of experience at exit, 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. To assess whether this gap is driven by selection, I exploit the fact that teacher exits vary in their degree of voluntariness. Tenured teachers— whose moves are largely voluntary given their job protections— experience larger earnings drops than untenured teachers, but even untenured teachers see drops of \$15,000 to \$25,000.[27] Inspired by the mass layoffs literature, I also examine exits that coincide with school-level shocks— mass turnover events or principal changes— which are less likely to reflect individual-level selection.[28] The earnings drops are similarly large.[29]

[26] I code non-employment as zero earnings to maintain the same sample throughout.

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

[28] I define a "mass turnover" as a school turnover rate that falls above the 75th percentile in a given year.

[29] These estimates are comparable to the earnings effects of mass layoffs and worker displacement in the broader labor literature: Jacobson, LaLonde, and Sullivan (1993), Couch and Placzek (2010), Davis and Von Wachter (2011), and Lachowska, Mas, and Woodbury (2020) find first-year earnings losses of 33–66%, while I estimate first-year losses of 33–40% among leaving teachers whose exit coincides with a school-level shock.

The matched CPS corroborates these findings: teachers leave to jobs that pay a median of \$13,000 less per year. The most common post-teaching sectors include education (teacher assistants, administrators), healthcare (nurses, social workers), administrative support (secretaries, office clerks), and retail. Nearly half of leaving teachers exit to full-time jobs while one-fifth exits to part-time work.

Despite targeting different populations— marginal entrants versus experienced exiters— the event study and RK estimates converge on similar pay gap magnitudes. These magnitudes are plausible given the CPS occupations analysis: part-time or full-time work as a cashier or secretary pays much less than teaching. The pay gap also persists up to ten years after exit.

II.H. Contextualizing the Pay Gap Estimates

Using two empirical strategies that leverage different sources of variation in who is a teacher in KY, 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 USD). Both estimates are local: the RK identifies the pay gap for marginal test-passers, while the event study identifies the pay gap for experienced teachers who exit.

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, who include 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, which includes a sizable share of individuals who obtain the highest possible score of 200. If performance on the Praxis is a proxy for earnings potential, marginal passers may have lower outside options than the average teacher, implying a larger pay gap. The reduced-form analysis is consistent with this: short-term future income is weakly increasing in initial test scores. That said, because the Praxis CASE tests middle-school level mathematics, it is unclear how strongly scores predict earnings potential outside of teaching. Similarly, teachers who exit may be negatively selected on earnings potential, which would also push the event study pay gap upward.[30] As such, both designs likely deliver the pay gap for individuals with relatively low earnings potential, and the estimates plausibly serve as an upper bound on the pay gap for the average teacher. This is advantageous in our context: an upper bound on the pay gap implies an upper bound on the rent, provided the working conditions gap is similar across teacher populations. If even with the largest plausible pay gap, the rent is moderate for experienced teachers and zero for inexperienced teachers, the

[30] It remains an open question whether teacher value-added or other measures of teacher effectiveness correlate with general ability (Chingos and West, 2012).

rent for the average teacher is likely no larger.[31]

Despite their local nature, the RK estimates are informative for several reasons. First, marginal passers are a policy-relevant population: KY lowered the Praxis cutoff in 2023, directly affecting who enters the teacher pipeline at the margin. Second, the convergence of the RK and event study estimates is reassuring about external validity. The two designs leverage different sources of variation and cover different populations—marginal entrants versus experienced exiters—yet produce similar pay gap magnitudes. Third, marginal passers represent a substantial share of test-takers: nearly 40% of the sample within the optimal bandwidth scored near the cutoff. That said, the pay gap may differ for teachers who passed easily on their first attempt, and it is possible that high-ability teachers face a smaller or even negative pay gap relative to their outside options.

Finally, I use a back-of-the-envelope calculation to put my estimates in context. Table III shows the results of dividing the estimates of the pay gap by benchmark values of KY 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.[32,33] 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.[34] 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 KY teacher's annual earnings, depending on the benchmarks used.[35]

[31] It is also worth mentioning that although these estimates provide insight into the pay gap at different points in the teaching pipeline, one may be interested instead in the gap in lifetime earnings between teaching and teachers' next-best jobs. Because earnings growth is lower in teaching than in other jobs and industries, it is plausible that even with a large early-career pay gap, there may be a small or even negative lifetime earnings gap between teaching and other occupations.

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

[33] There 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.

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

[35] My 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 V. 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 V shows, this point estimate shrinks and appears to stabilize at lower values as the bandwidth and first-stage F-stat grows.

III. Estimating the Working Conditions Gap

III.A. Identification Challenge

In the second half of this paper, the goal is to estimate how much more or less teachers value the non-wage amenities in teaching relative to the amenities in their next-best job options. Doing this requires not only estimating teachers' willingness to pay for various amenities, but also knowing what teachers' next-best jobs are and how the amenities differ between teaching and teachers' next-best jobs. Both are challenging to pin down using observational data on workers' job transitions, and data on amenities are generally scarce.

To learn about the amenities in teachers' next-best jobs, I leverage survey data from the American Working Conditions Survey (AWCS) (Maestas et al., 2023) and the results of the earlier quasi-experimental analysis. The RK design and event studies show that teachers' next-best jobs are low-paying options in the education, healthcare, administration, and retail industries. Reasonable proxies for the amenities in teachers' next-best job can therefore be obtained using the survey responses of workers in the AWCS in the education and healthcare sectors earning below-median wages.

To estimate teachers' preferences for a wide range of job attributes, I administer a set of choice experiments to a representative sample of teachers in KY. Teachers are presented with a series of job pairs with randomly varied job attributes and asked which job they most prefer. The resulting choice data allows me to observe the trade-offs teachers face and the choices they make, and back out teachers' willingness to pay for each attribute included in the experiment.

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

III.B. New Data on Working Conditions

Survey of KY 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 KY. The final sample include responses from over 1200 teachers, and a total of 3,700 choices made in the choice experiments.[36] The email response rate was 6%, exactly the average email survey response rate according to Qualtrics.

The two goals of the survey are to elicit information about teachers' working conditions, and to run a choice experiment that allows me to estimate teachers' willingness-to-pay for var-

[36] Data come from two survey rounds run in May 2024 and another two rounds in May 2025. Each respondent was asked to respond to a minimum of three job choices, with the option to evaluate one or two additional choices. The majority chose to evaluate the maximum of five job choices.

ious working conditions. To achieve these goals, I randomize school districts into two groups and give school staff in each group a survey that closely replicates the American Working Conditions Survey (AWCS). The first group received a survey with the exact same questions and experimental design as the AWCS (Maestas et al., 2023).[37] The second group received a survey that asked about different, "teaching-specific" working conditions, but retained the same structure and experimental design. The teaching-specific working conditions included: job hours that follow the school calendar (including "summers off"), hours that allow one to leave the workplace mid-day, working with children, working with parents ("clients"), job security, support from management in situations of conflict with clients, facing hostility at work (e.g., verbal aggression), and the amount of respect the job receives from the public. I selected attributes that were either reported reasons for dissatisfaction in teaching (e.g., school safety, student discipline, administration) in the Teacher Follow-up Survey (NCES), or that were salient characteristics specific to teaching (e.g., school calendar, working with children).[38,39]

There are distinct advantages to surveying KY teachers in this way. Targeting KY teachers ensures internal consistency, meaning the willingness-to-pay estimates and earlier pay gap estimates are derived from the same population. Asking respondents about the same core attributes studied in the AWCS ensures I am assessing job attributes that workers find salient when choosing between jobs. Finally, given my survey provides the first set of estimates of teachers' willingness-to-pay for a wide range of general job attributes, it is useful that the AWCS data provides several natural benchmark estimates: namely, willingness-to-pay estimates for women, college-educated workers, and workers in the education and health sector.[40]

[37]The AWCS was the first survey designed to elicit "detailed information about a broad range of working conditions in the American workplace," and was first fielded to a nationally representative sample of U.S. workers in 2015 by Maestas et al. (2023) through the RAND American Life Panel (ALP). The survey contained two parts: a questionnaire asking respondents about their current job attributes, and a stated-preference module that presents respondents with ten pairs of hypothetical jobs and asks which job they prefer in each case.

[38]The inclusion of "public respect" as a job attribute was inspired by Dube, Naidu, and Reich (2025). The Teacher Follow-Up Survey is part of the Schools and Staffing Survey run by the National Center for Education Statistics (NCES). The teaching-specific questions were formulated by adapting existing questions on similar topics from the AWCS, the Centers for Disease Prevention (CDC) National Institute for Occupational Safety and Health (NIOSH) Quality of Worklife Module, and the Pew Research Center's American Trends Panel.

[39]Notably, this list excludes retirement plans. This is intentional. In KY, teachers are offered a defined benefit plan, provided by the Teachers' Retirement System of Kentucky, but no social security benefits —and teachers' WTP for retirement plans has been previously explored, (Johnston, 2025). Furthermore, with our data, it is difficult to tell whether teachers' next-best jobs offer retirement benefits or not, and whether the provision of these benefits is a relevant job attribute to consider for most workers: young teachers, who have the highest turnover rates in the profession, likely have relatively low willingness to pay for retirement plans, as found by Johnston (2025).

[40]The AWCS data also gives an imprecise benchmark for teachers—"imprecise" because since the AWCS

Table IV shows summary statistics on the demographic composition of the survey respondents, with comparisons to the public school teaching population in KY. The survey sample closely resembles the teaching 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.

III.C. Incidence of Working Conditions in Teaching

I begin by examining the incidence of the nine core non-wage job attributes among KY teachers. Figure VIII presents a summary of the incidence of working conditions for four groups of workers: KY teachers, all workers in the U.S., college-educated workers, and workers in the education and health industries with below median pay.[41] Means are weighted using survey weights.

Compared to the job of the average college-educated worker, teaching is significantly less likely to offer the valued working conditions included in the AWCS, like setting one's own schedule, but is much more likely to offer "teaching-specific" attributes, like working with children. 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 more likely to report more physically demanding work (2% report mostly sitting, compared to 58% of college-educated workers). Predictably however, teaching is highly likely to offer "teaching-specific" job attributes. All teachers report having hours that follow the school calendar and working with children, and the majority report having high job security.

These differences may not be surprising given the traditional model of classroom teaching and the structure of the school year: classroom teachers are required to be physically present at school five days a week, and "summers off" do not technically count as paid time off. However, teaching also differs from other jobs along other less obvious dimensions. Perhaps most starkly, teachers are significantly more likely than all of the other groups of workers

does not collect occupation for the majority of respondents, only a small sample of 46 (out of the total 1,748 respondents) report a string similar to "teacher" as their occupation (e.g., "teach," "teac"), and these individuals reside in various places across the country.

[41] All statistics in the figure, along with additional summary statistics on the job attributes in teaching compared to the jobs held by other workers in the AWCS, are provided in Supplementary Tables A.1, A.2, and A.3.

to experience a stressful "fast-pace" at work (87% vs. 64% for college-educated, 67% for women, and 70% for all workers). Relatedly, teaching is significantly more likely to involve dealing with hostility in the workplace compared to other jobs. 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 from the AWCS 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. Using the AWCS data, I construct similar comparisons between other workers and teachers, where I identify teacher respondents by merging in occupation information from the RAND American Life Panel.[42] Reassuringly, I reach the same qualitative conclusions using the AWCS data: that is, that teaching is less likely to offer valued working conditions compared to the jobs of observationally similar workers. Both surveys also show that teaching pays less than the jobs of demographically similar workers, in line with the findings of previous studies (Allegretto and Mishel, 2018).[43]

Finally, I focus on the comparison most relevant to the working conditions gap: how the working conditions in teaching compare to those of teachers' "next-best" jobs. The RK and event studies suggest that the average teachers' next-best job is a low-paying option in the education or healthcare sector.[44] I therefore compare the attributes in teaching to jobs in the AWCS in the education and healthcare sectors with below-median pay, represented by the triangles in Figure VIII.[45] The differences in attributes all go in the same direction as

[42]Supplementary Table A.4 shows summary statistics on the differences between the AWCS teachers and other workers. Statistical tests in Supplementary Tables A.5 and A.6.

[43]Supplementary Table A.1 shows that KY 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.

[44]The next-best jobs identified using the quasi-experimental designs align with teachers' beliefs, as revealed in the survey. Figure A.6 shows the the distribution of responses to the question: what industry they would work in if not teaching? Over a third of teachers indicated 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.

[45]Supplementary Table A.2 summarizes the job attributes in teaching versus in the average job in the next-best industries, while Supplementary Table A.3 summarizes the attributes of the jobs that pay below the industry median wage. The distribution of attributes is similar between the average paying and the below-median paying jobs within industry. In terms of pay, the average job in all industries, aside from leisure and hospitality, pays more on average than teaching does. 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 teaching does not have, including scheduling flexibility, telecommuting options, relaxed pace of work, and more than 15 days paid time off.

the differences between teaching and college-educated workers' jobs: even when compared to their lower-paying next-best options, teaching is significantly less likely to offer most desirable attributes, aside from the "teaching-specific" conditions.

III.D. Estimating Teachers' Willingness to Pay for Working Conditions

To estimate teachers' willingness to pay for each job attribute, I conducted a series of stated-preference experiments with each survey respondent. I include attention-check questions to screen inattentive respondents; results are robust to excluding them.[46]

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. Respondents are randomly assigned to one of two attribute sets: the nine core job attributes from the AWCS (e.g., schedule flexibility, physical demands, paid time off)[47] or nine teaching-specific attributes not in the AWCS (e.g., having "summers off," working with children, dealing with parents).

In the second part of the survey, respondents are presented with three choice experiments.[48] In each experiment, respondents are asked to select between two hypothetical job profiles that partially vary in terms of job attributes, hours, and wages, exactly following the technical and visual design described in Maestas et al. (2023).

I construct the hypothetical job profiles in two stages. First, I define each respondent's "baseline job" as the pay, hours, and attributes of their current or former teaching job.[49] Second, I create Jobs A and B by randomly varying two of the nine non-wage attributes (without replacement), plus the wage and hours, so that each job profile differs on four dimensions. To maximize estimate precision, I redraw wages and/or attribute values in cases where one job dominates the other on all dimensions.[50] Respondents are presented

[46]See Supplementary Materials I.D for details on the experimental design, attention checks, and randomization.

[47]See Supplementary Table 2 of Maestas et al. (2023) for a list of the job attributes and their potential values.

[48]To maximize the data collected without inducing survey fatigue, I also gave respondents the option to respond to at most two "bonus" experiments, so that respondents could complete up to five experiments. No additional incentives were provided; survey respondents were simply told that completing the additional experiments would "be a huge help to the study." While the majority actually did respond to at least one bonus experiment, my main results use the first three experiments to avoid selection bias from additional responses.

[49]To address concerns that respondents may prefer familiar job profiles, I also use a "common" baseline job— essentially the average teaching job— in one of the three experiments. The inclusion of common baseline jobs does not affect the willingness-to-pay estimates. See Supplementary Table 2 of Maestas et al. (2023) for the common baseline values.

[50]See Supplementary Materials I.D for details on wage elicitation, randomization, and robustness checks.

with a table showing a side-by-side comparison of the two jobs' attributes, and are 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 Supplementary Figure A.5, is virtually identical to the presentation in Supplementary Figure 3 of Maestas et al. (2023).

Estimation. I estimate KY teachers' willingness-to-pay for each job attribute using the choice data. Assume that the respondents have the following indirect utility function:

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

where A_{ij} is the vector of job attributes offered by job j and w_{ij} is the corresponding wage. 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

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

In theory, the β_i and δ_i subscripts allow for heterogeneity in preferences over amenities and pay. However, because I focus on workers in a specific occupation, and thus 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. The willingness-to-pay estimates are robust to a battery of specification checks detailed in Supplementary Materials I.D.

I transform the estimated parameters from the discrete choice model into measures 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 offers w_i without r and a job that offers r but a lower wage, $w_i - \text{WTP}_i^r$:

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

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

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

For ease of interpretation, I report the 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.

Comparison between jobs. Finally, I use the preference estimates to estimate how much teachers are willing to pay for one job over another. I study 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.[51]

I introduce notation for the three measures. Let S denote the set of all attributes, $\beta^{s,1}$ the preference parameter for the highest attribute value of s, and $\beta^{s,2}$ the parameter for the second highest value.[52,53]

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

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

The sum in this expression only puts positive weight on the estimated preferences for the most preferred value of each attribute.[54] This measure is equivalent to the best-to-worst measure used in Maestas et al. (2023).

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

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

where $\mathbb{I}_{\text{teach}}^{s,j}$ is an indicator variable for whether the average KY teaching job offers the jth highest attribute value of s or not. The summation only gives weight to the attributes that teaching does *not* offer. Thus, 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.[55]

[51] I continue using low-paying jobs in the education and health sectors in the AWCS as a proxy for teachers' next-best jobs.

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

[53] For 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.

[54] Note that this baseline measure does not "double count" attribute values, which would inflate the measure. 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.

[55] The expression also makes clear why it is necessary to include multiple attribute values, not only the highest valued ones, when valuing a job relative to teaching. Take for example the paid time off attribute, which has three values—no time, 10 days, and 20 days—that workers value in increasing order. If teaching offers 10 days paid time off (meaning $\mathbb{I}_{\text{teach}}^{s,2} = 1$), then when evaluating how much teachers would value having 10 additional days paid time off, we would want to include teachers' valuations of 20 days paid time off *netting out* their valuation of the 10 days paid time off that they already get as teachers, which is given

Finally, the third measure 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:

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

where $\mathbb{I}_{\text{nxbest}}^{s,1}$ is an indicator variable for whether the average next-best job has the highest value of attribute s or not. 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.

Identification requires that respondents treated the two jobs as identical aside from the randomly varied characteristics, as they were explicitly instructed to do. 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 teachers are somewhat homogeneous (e.g., mostly women) and engage in similar work.

III.E. *Main Estimates of Teachers' Willingness to Pay for Working Conditions*

I present my main estimates of KY teachers' willingness-to-pay for each job attribute in Figure IX, alongside willingness-to-pay estimates for other workers from Maestas et al. (2023) and my own survey where available.[56] 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, relative to some baseline attribute. For example, the estimate for teachers in the first row indicates that the average KY 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.

I highlight four key findings. First, like other workers, KY teachers are willing to pay large amounts for certain 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. Teachers also highly value switching to a job with more

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

[56] Estimates are also shown in Supplementary Table A.7.

boss/management support (equivalent to a 21 percent wage increase), with high job security (equivalent to a 24.5 percent wage increase), and with a schedule that follows the school calendar (equivalent to a 22.4 percent wage increase).

Second, I find that the average teacher and the average worker share similar preferences for most of the core AWCS attributes. The only exceptional difference is having a relaxed pace at work: teachers value switching from a fast-paced job to a relaxed pace job at a 10 percent wage increase, nearly twice as much as the average worker does.[57] One explanation for this difference is the prevalence of teacher "burnout": workers who experience fast-paced, stressful jobs at the time of the survey may place a higher value on a relaxed pace of work. Indeed, surveys show that stress and burnout are common among teachers and contribute to high turnover rates, especially among starting teachers (e.g., Herman, Hickmon-Rosa, and Reinke (2018)).

However, my third finding is that teachers' preferences deviate from other workers' preferences when it comes to teaching-specific attributes. Compared to other workers, teachers are more willing to pay for a job that follows the school calendar, involves working with children, and offers a high level of support from their boss/management (i.e., principal or school administrators) in conflicts with 'clients' (i.e., parents).[58] Furthermore, teachers are not willing to pay as much as other workers to work with 'clients' (i.e., parents) or job security.

Finally, I show that teachers are willing to pay a substantial amount of their salary to switch to typical their next-best job for its better amenities. Figure X shows the estimates of how much teachers are willing to pay to switch jobs. Switching from the least desirable job to the most desirable job, in terms of AWCS amenities only, is equivalent to a 62.2 percent wage increase for KY teachers. Reassuringly, this 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). However, once teaching-specific attributes are included as well, switching from the "worst" job to the "best" job is equivalent to a much larger 93 percent wage increase. Switching from teaching to the "best" job, in terms of amenity value, is equivalent to a 43 percent wage increase. Finally, switching from teaching to the average teachers' next-best job, in terms of amenities, is equivalent to a 17 percent wage increase.

[57]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 VIII, 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.

[58]In fact, working with children is the only attribute included in the survey that other workers find to be a *disamenity*, while the average teacher has a positive willingness to pay to work with children.

Importantly, there is significant heterogeneity underlying the average willingness to pay estimates. The average total willingness-to-pay to switch from teaching to the next-best job of 17% is entirely driven by teachers with more than five years of experience. In contrast, less experienced teachers are willing to pay significantly more—as much as 51 percent of their teaching salary—to switch to their next-best job. This heterogeneity is mostly driven by differences in working conditions. Compared to experienced teachers, inexperienced teachers are significantly more likely to report facing hostility, less likely to receive support from management in situations of conflict with parents/children, and less likely to report high job security. The difference in job security in particular reflects the institutional setting: in KY, teachers are on one-year contracts until they obtain "tenure" (high job security), which can only be obtained after working in the same district for four consecutive years.

III.F. Qualitative Evidence from Open-Ended Responses

The quantitative estimates above are internally consistent with KY teachers' views expressed in the open-ended portion of the survey.[59] When asked to share anything more about their experience choosing between teaching and other jobs, the most commonly raised issues were student behavior and parent involvement rather than pay.

IV. Discussion

Returning to the rent equation defined in Section II and combining my most conservative estimates, I find that experienced teachers in KY earn a rent of around 16% of their salary and a compensating differential of around 17% of their salary. In contrast, I find that inexperienced teachers in KY earn no rent, or even a "negative rent," given their large working conditions gap.

My results suggest that, although teaching pays more than teachers' next-best jobs, much of the pay premium is a compensating differential rather than a rent. In particular, for inexperienced teachers, the entire pay premium functions as a compensating differential due to the more challenging working conditions they tend to face. Through the lens of the compensating differentials framework in Rosen (1986), these estimates can be interpreted as showing that experienced teachers are inframarginal while inexperienced teachers are "marginal" to teaching. This is a sensible conclusion, as workers may derive rents from experience, and turnover rates are highest among less experienced teachers.

This conclusion is in line with increasing reports of teacher shortages and difficult work-

[59] Appendix Figure XII visualizes the most common themes in teachers' open-ended responses.

ing conditions in teaching, especially since COVID-19.[60], and has significant implications for education policy. Given that over \$500 billion are spent each year on teacher salaries, under the simple approximation that 70% of the wage bill is spent on experienced teachers' salaries, my estimates imply that over \$120 billion are being spent each year to compensate teachers for challenging working conditions. The magnitude of this estimate raises the question of whether there may be more cost-effective solutions that could address the gap in working conditions between teaching and other jobs.

Comparison to private schools. My estimates are large, but a comparison of private and public school teacher pay suggests they are plausible. Private schools can select their students, leading to smaller class sizes and fewer behavioral issues.[61] Consistent with compensating differentials, in 2020–21, the average private school teacher in the US earned around \$19,000 *less* than the average public school teacher in 2018 USD (National Center for Education Statistics, 2024)—around 42% of an early-career teaching salary in KY, a very similar number to my estimated working conditions gap of 51% for inexperienced teachers.

Comparison to prior work. The central insight of this paper— that much of the teaching pay premium is a compensating differential— helps reconcile three findings in the literature that have previously been treated as separate puzzles.

First, and most directly, CPS-based policy reports show teachers earn less than other college graduates (Allegretto and Mishel, 2018), while event studies consistently find that exiting teachers leave to lower-paying jobs (Stinebrickner, 2002; Scafidi, Sjoquist, and Stinebrickner, 2006; Goldhaber et al., 2022). Both are true: teachers earn less than college graduates on average, and their actual next-best options pay less than teaching. My findings show that the pay premium exists *because* it is a compensating differential, not because teachers are overcompensated.

Second, this lens offers a new explanation for the increasing negative selection of teachers on academic performance (Bacolod, 2007; Corcoran, Evans, and Schwab, 2004), previously attributed to wage compression (Hoxby and Leigh, 2004) and expanding opportunities for women. Worsening working conditions provide a third channel: as teaching conditions deteriorate, the occupation increasingly selects workers with lower-paying outside options who accept the pay premium as a compensating differential.

Third, while prior work attributes school districts' monopsony power to locational prefer-

[60] For example, see Sejla Rizvic, "Teachers, Facing Increased Level of Stress, Are Burned Out," *The New York Times*, March 13, 2023.

[61] For example, see Ben Orlin, "Why Are Private-School Teachers Paid Less Than Public-School Teachers?", *The Atlantic, October 24, 2013* .

ences and pension lock-in (Ransom and Sims, 2010; Boyd et al., 2013), teaching's uncommon yet desirable amenities— job security, summers off— may provide an additional source of market power, allowing schools to reduce wages *and* sustain challenging conditions.

Turning to the magnitude of the rent itself, the 16% rent I find for experienced teachers is in a similar range to recent estimates of worker rents in the broader labor market— 13–18% in Lamadon, Mogstad, and Setzler (2022) and 10–14% in Jäger et al. (2024)— where most workers are not unionized. This may seem surprising for an occupation typically associated with strong unions, but KY has low union coverage: less than 30% of teachers were covered by collective bargaining in 2011.[62] The moderate rent I find is therefore consistent with the low-union setting, and my findings are most applicable to states with similarly low union involvement.

Finally, this is also the first study to incorporate empirical estimates of non-wage amenity value when evaluating rents. Prior measures based on staff expenditures miss amenities that entail no explicit cost (e.g., working in a team versus working alone). My results are consistent with a robust education literature that documents that poor working conditions are a key predictor of teacher turnover (Loeb, Darling-Hammond, and Luczak, 2005; Borman and Dowling, 2008; Ladd, 2011; Boyd et al., 2011; Johnson, Kraft, and Papay, 2012).

Limitations. The main results come from one state, which limits generalizability. The pay and working conditions gaps likely vary by state, as suggested by the fact that the event study estimates from the matched CPS data—which covers the entire U.S., as opposed to one state—give a smaller, but still positive, pay gap between teaching and teachers' next-best jobs. That said, existing studies consistently find that exiting teachers leave to lower-paying jobs, suggesting the pay premium is not unique to KY.

V. Conclusion

Applying independent quasi-experimental designs to novel administrative data from KY, I find that teaching in KY offers a large pay premium relative to teachers' next-best options. I estimate a pay gap of approximately $20,000 per year, between 33–40% of the average teachers' annual earnings. I also find that KY teachers are willing to pay 17% of their salary to switch from the bundle of attributes in teaching to the overall more desirable bundle in their next-best jobs. The working conditions gap estimates vary by teaching experience, with inexperienced teachers willing to pay much more (51% of their salary) for the attributes in their next-best job than experienced teachers. More broadly, this paper demonstrates a

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

general approach for evaluating occupation-specific rents that accounts for both pay and non-wage amenities, which could be applied to other occupations where compensating differentials are suspected to play a role.

The idea that disamenities may account 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 does teaching lack that can be introduced or 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? Investigating these questions may be a critical step in addressing the hiring challenges that schools are facing today.

Microsoft Research

References

Allegretto, Sylvia and Lawrence Mishel (2018). "The Teacher Pay Penalty Has Hit a New High: Trends in the Teacher Wage and Compensation Gaps through 2017." *Economic Policy Institute.*

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

Bates, Michael, Michael Dinerstein, Andrew C Johnston, and Isaac Sorkin (2025). "Teacher labor market policy and the theory of the second best," *The Quarterly Journal of Economics*, 140, 1417–1469.

Biasi, Barbara, Chao Fu and John Stromme (2025). "Equilibrium in the market for public school teachers: District wage strategies and teacher comparative advantage," *Journal of Political Economy (Revise and Resubmit).*

Borman, Geoffrey D and N Maritza Dowling (2008). "Teacher attrition and retention: A meta-analytic and narrative review of the research," *Review of Educational Research*, 78, 367–409.

Boyd, Donald, Pam Grossman, Marsha Ing, Hamilton Lankford, Susanna Loeb, and James

Wyckoff (2011). "The influence of school administrators on teacher retention decisions," *American Educational Research Journal*, 48, 303–333.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff (2005). "Explaining the short careers of high-achieving teachers in schools with low-performing students," *American Economic Review*, 95, 166–171.

——— (2013). "Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers," *Journal of Labor Economics*, 31, 83–117.

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

Card, David, David S Lee, Zhuan Pei, and Andrea Weber (2015). "Inference on causal effects in a generalized regression kink design," *Econometrica*, 83, 2453–2483.

Card, David, Jesse Rothstein and Moises Yi (2024). "Industry wage differentials: A firm-based approach," *Journal of Labor Economics*, 42, S11–S59.

Chingos, Matthew M and Martin R West (2012). "Do more effective teachers earn more outside the classroom?" *Education Finance and Policy*, 7, 8–43.

Corcoran, Sean P, William N Evans and Robert M Schwab (2004). "Changing labor-market opportunities for women and the quality of teachers, 1957–2000," *American Economic Review*, 94, 230–235.

Couch, Kenneth A and Dana W Placzek (2010). "Earnings losses of displaced workers revisited," *American Economic Review*, 100, 572–589.

Davis, Steven J and Till M Von Wachter (2011). "Recessions and the cost of job loss," *Brookings Papers on Economic Activity*.

Deneault, Christa, Evan Riehl and Jian Zou (2026). "Disparate Impacts of Teacher Certification Exams," Tech. rep., National Bureau of Economic Research.

DiNardo, John and David S Lee (2011). "Program evaluation and research designs," in *Handbook of Labor Economics*, Elsevier, vol. 4, 463–536.

Dong, Yingying (2018). "Jump or kink? Regression probability jump and kink design for treatment effect evaluation," *Working Paper*.

Drake, Marshall, Neil Thakral and Linh Tô (2022). "Wage differentials and the price of workplace flexibility," *Working Paper*.

Dube, Arindrajit, Suresh Naidu and Adam D. Reich (2025). "Power and Dignity in the Low-Wage Labor Market: Theory and Evidence from Wal-Mart Workers," *Working Paper*.

Goldhaber, Dan, John Krieg, Stephanie Liddle, and Roddy Theobald (2022). "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*, 1–40.

Herman, Keith C, Jal'et Hickmon-Rosa and Wendy M Reinke (2018). "Empirically derived profiles of teacher stress, burnout, self-efficacy, and coping and associated student outcomes," *Journal of Positive Behavior Interventions*, 20, 90–100.

Hoxby, Caroline Minter (1996). "How teachers' unions affect education production," *The Quarterly Journal of Economics*, 111, 671–718.

Hoxby, Caroline M and Andrew Leigh (2004). "Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States," *American Economic Review*, 94, 236–240.

Jacobson, Louis S, Robert J LaLonde and Daniel G Sullivan (1993). "Earnings losses of displaced workers," *The American Economic Review*, 685–709.

Jäger, Simon, Christopher Roth, Nina Roussille, and Benjamin Schoefer (2024). "Worker beliefs about outside options," *The Quarterly Journal of Economics*, 139, 1505–1556.

Jepsen, Christopher, Peter Mueser and Kenneth Troske (2016). "Labor market returns to the GED using regression discontinuity analysis," *Journal of Political Economy*, 124, 621–649.

Johnson, Susan M, Matthew A Kraft and John P Papay (2012). "How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement," *Teachers College Record*, 114, 1–39.

Johnston, Andrew C. (2025). "Preferences, Selection, and the Structure of Teacher Pay," *American Economic Journal: Applied Economics*, 17, 310–346.

Kentucky Legislative Research Commission (2022). "Teachers in Kentucky: Changes in Turnover and Workforce Composition," Accessed: 2024-11-02.

Kroft, Kory, Yao Luo, Magne Mogstad, and Bradley Setzler (2025). "Imperfect competition and rents in labor and product markets: The case of the construction industry," *American Economic Review*, 115, 2926–2969.

Lachowska, Marta, Alexandre Mas and Stephen A Woodbury (2020). "Sources of displaced workers' long-term earnings losses," *American Economic Review*, 110, 3231–3266.

Ladd, Helen F (2011). "Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement?" *Educational Evaluation and Policy Analysis*, 33, 235–261.

Lamadon, Thibaut, Magne Mogstad and Bradley Setzler (2022). "Imperfect competition, compensating differentials, and rent sharing in the US labor market," *American Economic Review*, 112, 169–212.

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack Porter (2022). "Valid t-ratio Inference for IV," *American Economic Review*, 112, 3260–3290.

Lee, David S, Jack Porter, Justin McCrary, Luther Yap, and Marcelo J Moreira (2024). "Quantile Treatment Effects and Economic Policy Evaluation," *Working Paper.*

Loeb, Susanna, Linda Darling-Hammond and John Luczak (2005). "How teaching conditions predict teacher turnover in California schools," *Peabody Journal of Education*, 80, 44–70.

Lyon, Melissa Arnold, Matthew A Kraft and Matthew P Steinberg (2024). "The Effects of Teacher Strikes on Compensation, Working Conditions, and Productivity," *Working Paper.*

Maestas, Nicole, Kathleen J Mullen, David Powell, Till Von Wachter, and Jeffrey B Wenger (2023). "The value of working conditions in the United States and implications for the structure of wages," *American Economic Review*, 113, 2007–2047.

Manning, Alan (2011). "Imperfect competition in the labor market," in *Handbook of Labor Economics*, Elsevier, vol. 4, 973–1041.

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

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

National Center for Education Statistics (2024). “Public School Expenditures,” *Condition of Education*. U.S. Department of Education, Institute of Education Sciences, retrieved Oct 16 2024, from https://nces.ed.gov/programs/coe/indicator/cmb.

Ransom, Michael R and David P Sims (2010). “Estimating the firm’s labor supply curve in a “new monopsony” framework: Schoolteachers in Missouri,” *Journal of Labor Economics*, 28, 331–355.

Rosen, Sherwin (1986). “The theory of equalizing differences,” *Handbook of Labor Economics*, 1, 641–692.

Scafidi, Benjamin, David L Sjoquist and Todd R Stinebrickner (2006). “Do teachers really leave for higher paying jobs in alternative occupations?” *Advances in Economic Analysis & Policy*, 6.

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

Sutcher, Leib, Linda Darling-Hammond and Desiree Carver-Thomas (2016). “A coming crisis in teaching? Teacher supply, demand, and shortages in the US,” *Learning Policy Institute*.

Wiswall, Matthew and Basit Zafar (2018). “Preference for the workplace, investment in human capital, and gender,” *The Quarterly Journal of Economics*, 133, 457–507.

VI. Tables

Table I: Summary Statistics on First-Time Test-Takers

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

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

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

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

Table III: Benchmarking the Pay Gap Estimates to Teacher Salaries

		Pay Gap / Salary	
	Salary	Gap = \$18,000	Gap = \$21,000
	(1)	(2)	(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%

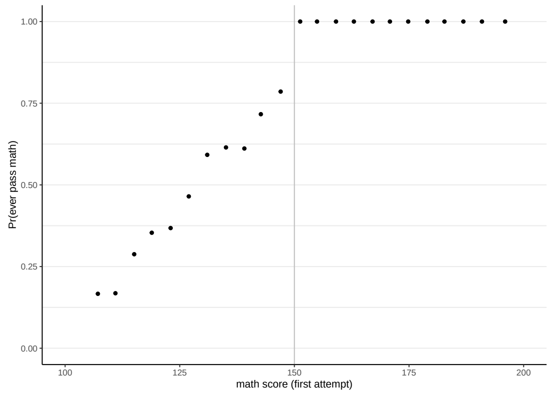
Note: Column (1) shows salaries from the 2019 salary schedules for teachers, available online from the KY Department of Education. Column (2) shows the ratio of the lower bound estimate of the pay gap to the salary in Column (1). Column (3) shows the ratio of the upper bound estimate of the pay gap to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

Table IV: Summary Statistics on Teachers in Survey versus Administrative data

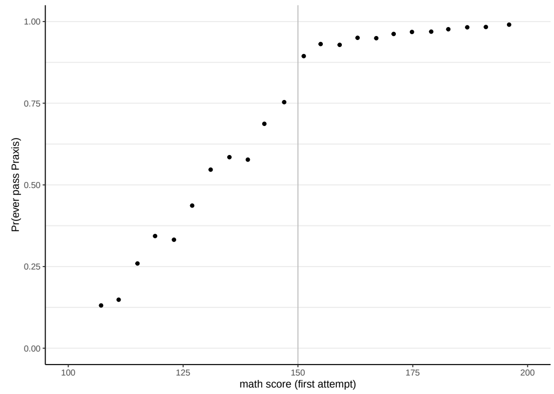
	Survey	Administrative data
	(1)	(2)
Panel A: Demographics		
White	0.90	0.95
	(0.30)	(0.21)
Woman	0.76	0.78
	(0.43)	(0.41)
Resides in rural area	0.57	0.27
	(0.49)	(0.45)
Age: 26-29	0.07	0.11
	(0.25)	(0.31)
Age: 30-34	0.08	0.15
	(0.27)	(0.35)
Age: 35-39	0.14	0.16
	(0.34)	(0.37)
Age: 40-44	0.17	0.15
	(0.38)	(0.36)
Age: 45-49	0.16	0.15
	(0.37)	(0.36)
Age: 50 plus	0.33	0.23
	(0.47)	(0.42)
Experience: 3-5 years	0.08	0.14
	(0.28)	(0.35)
Experience: 5-10 years	0.14	0.19
	(0.35)	(0.40)
Experience: 10+ years	0.72	0.57
	(0.45)	(0.50)
Teaches at high school	0.35	0.26
	(0.48)	(0.44)
Teaches at elementary school	0.63	0.48
	(0.48)	(0.50)
Teaches at Title I school	0.68	0.67
	(0.47)	(0.47)
N	1282	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.

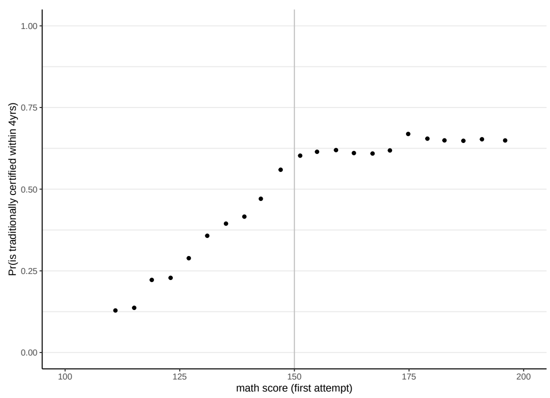
VII. Figures



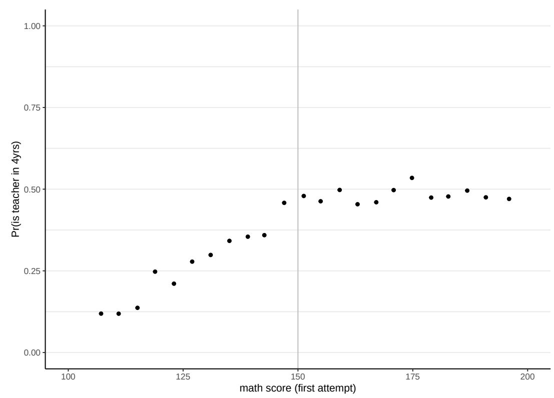
(a) Ever pass math.



(b) Ever pass whole Praxis CASE.



(c) Certified within 4yrs.



(d) Teacher in 4yrs.

Figure I: First Stage Relationships between First-Attempt Scores and Certification Outcomes

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 KY public schools in the fourth year after their first attempt.

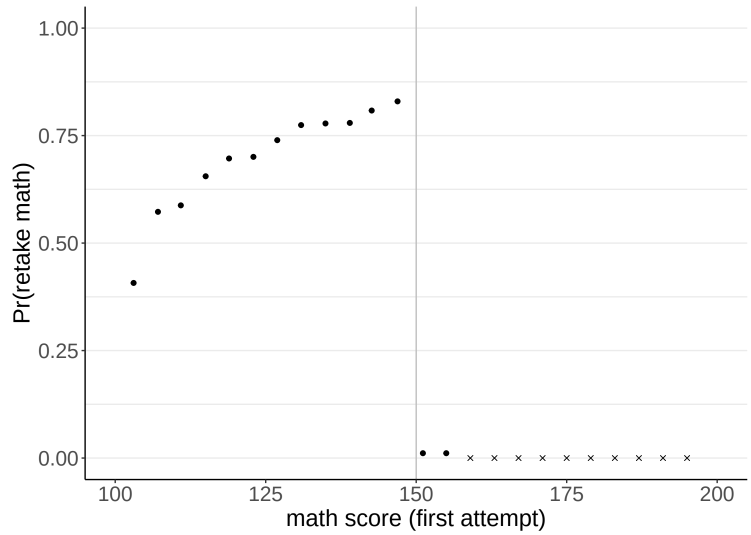
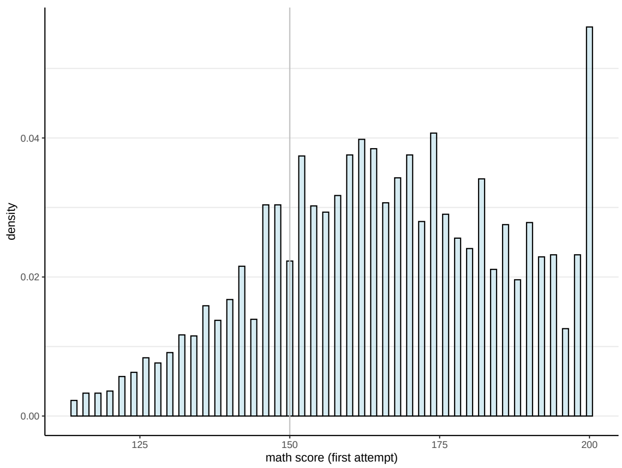
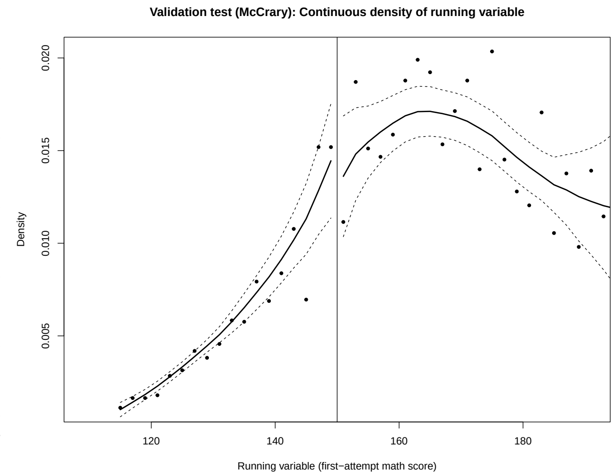


Figure II: Retaking Behavior

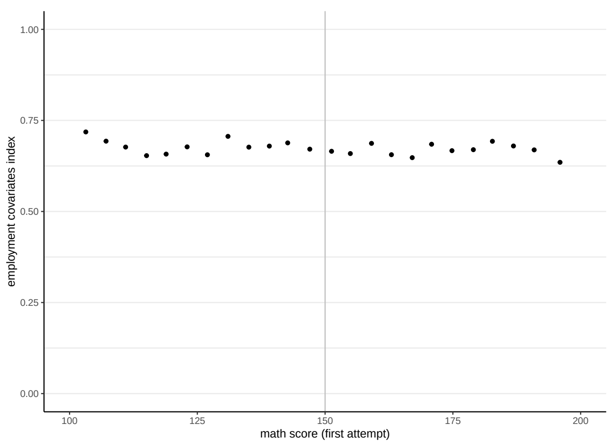
Note: This figure is a binscatter of the share of individuals who retake the exam at least once, plotted over first-attempt math scores. X's represent cells with too few observations to show. The increasing rate of retaking approaching the cutoff from the left, and the discontinuity in retaking after the cutoff, drives the kink.



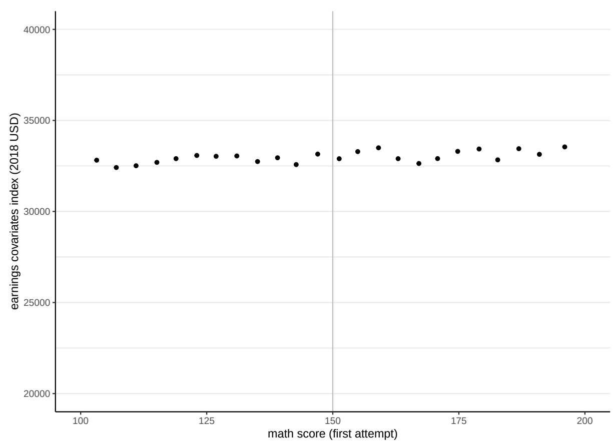
(a) Distribution of first-attempt math scores



(b) Manipulation test



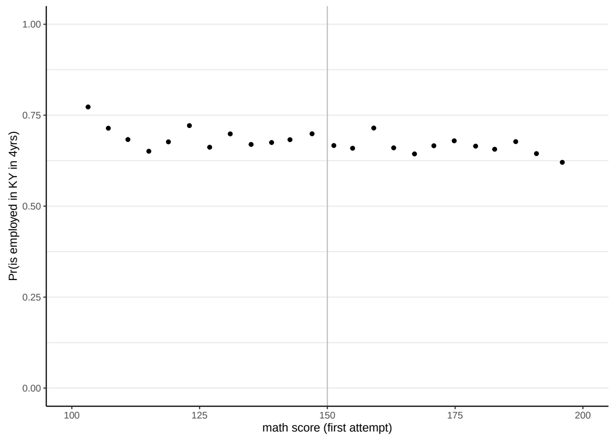
(c) Predicted Pr(employed in 4yrs)



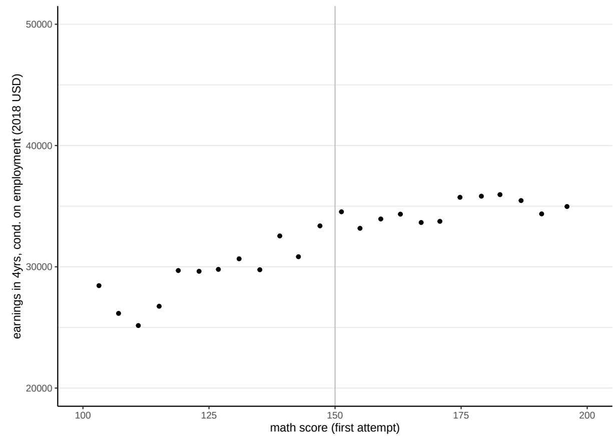
(d) Predicted earnings in 4yrs

Figure III: Tests for Manipulation of First-Attempt Math Scores AND Tests of Selection Around the Cutoff, Using Covariates Index

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). AND Each panel presents binscatters for an "index" variable that is predicted using only covariates at "baseline," 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.



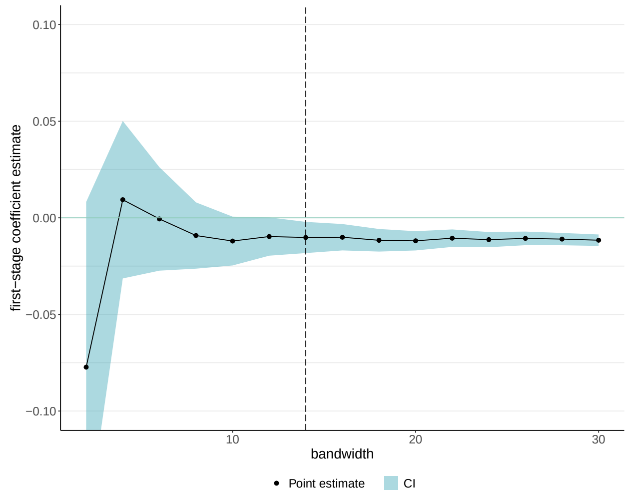
(a) Pr(employed in 4yrs in KY).



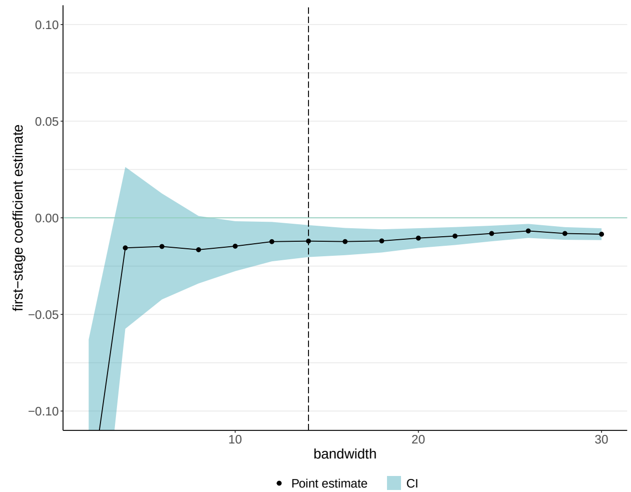
(b) Earnings in 4yrs (2018 USD).

Figure IV: Reduced Form Relationships underlying Regression Kink Design

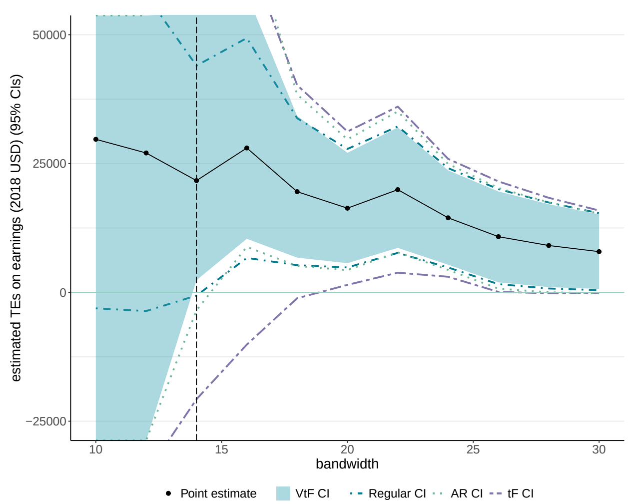
Note: Binscatters of the likelihood of being employed and earnings 4 years after one's first attempt on the Praxis CASE math, plotted over bins of first-attempt math scores.



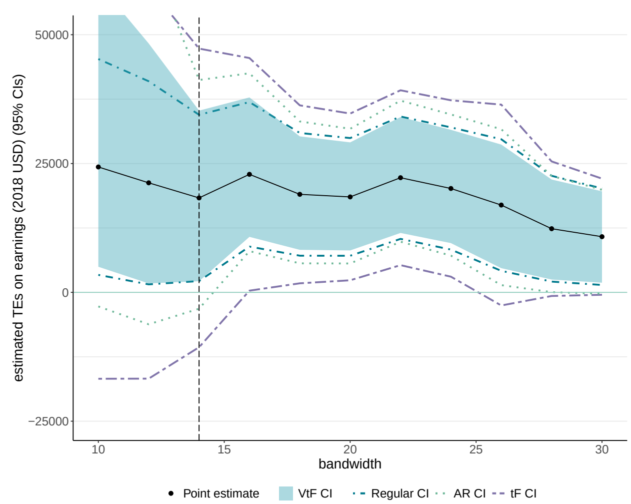
(a) First stage: Becoming certified.



(b) First stage: Becoming a teacher.



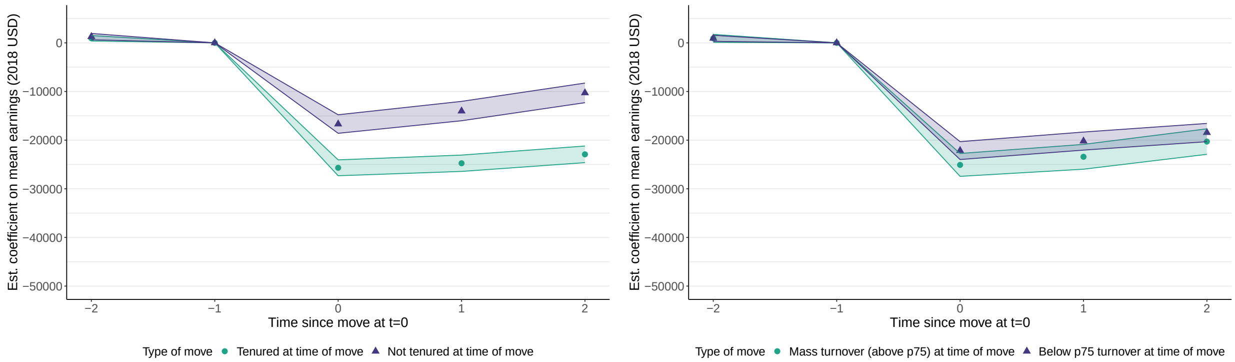
(c) Effect of becoming certified on earnings.



(d) Effect of becoming a teacher on earnings.

Figure V: Estimates from the Regression Kink Design

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(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).

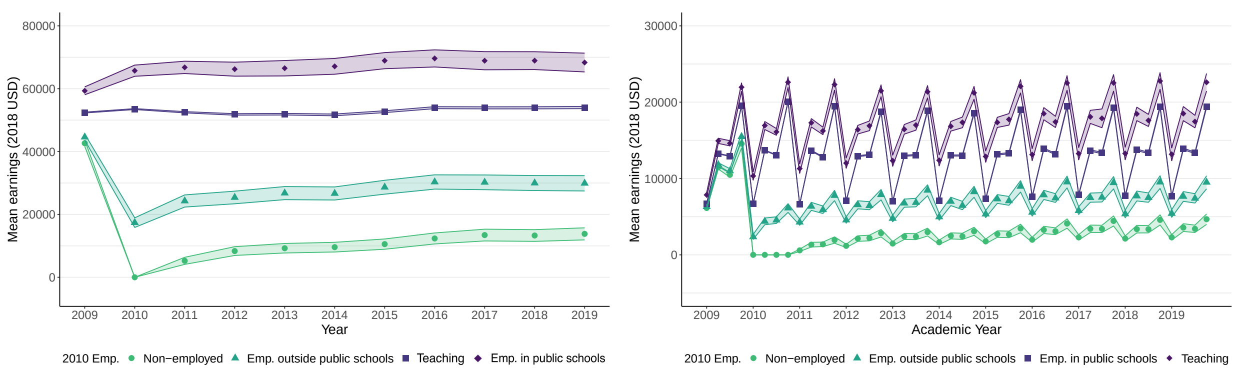


(a) By movers' tenure status

(b) By degree of turnover at school

Figure VI: Event Study Estimates of Earnings Changes around Teacher Exits

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.

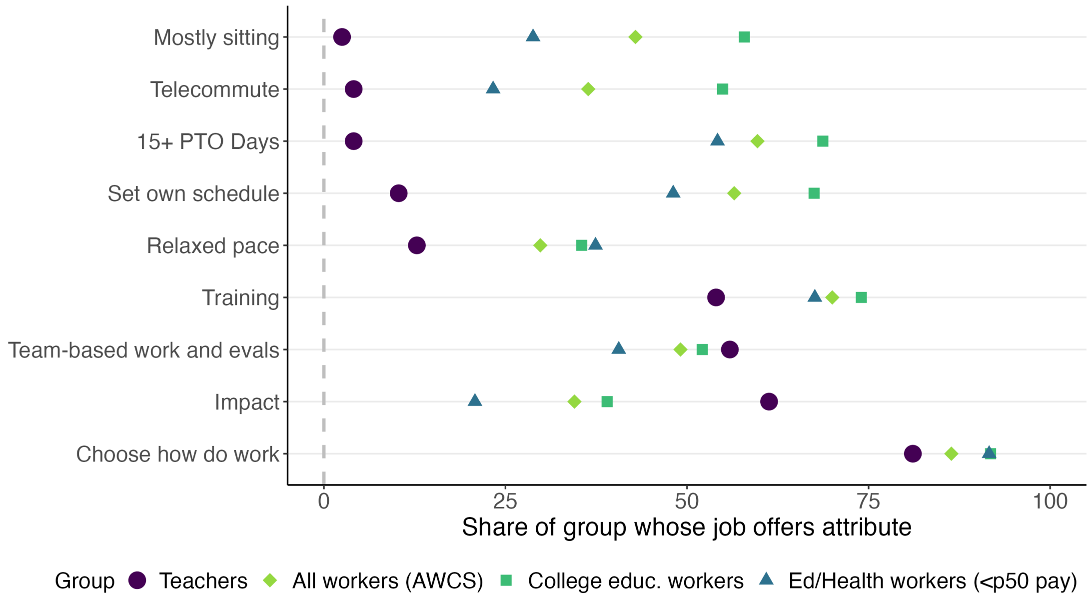


(a) Annual earnings.

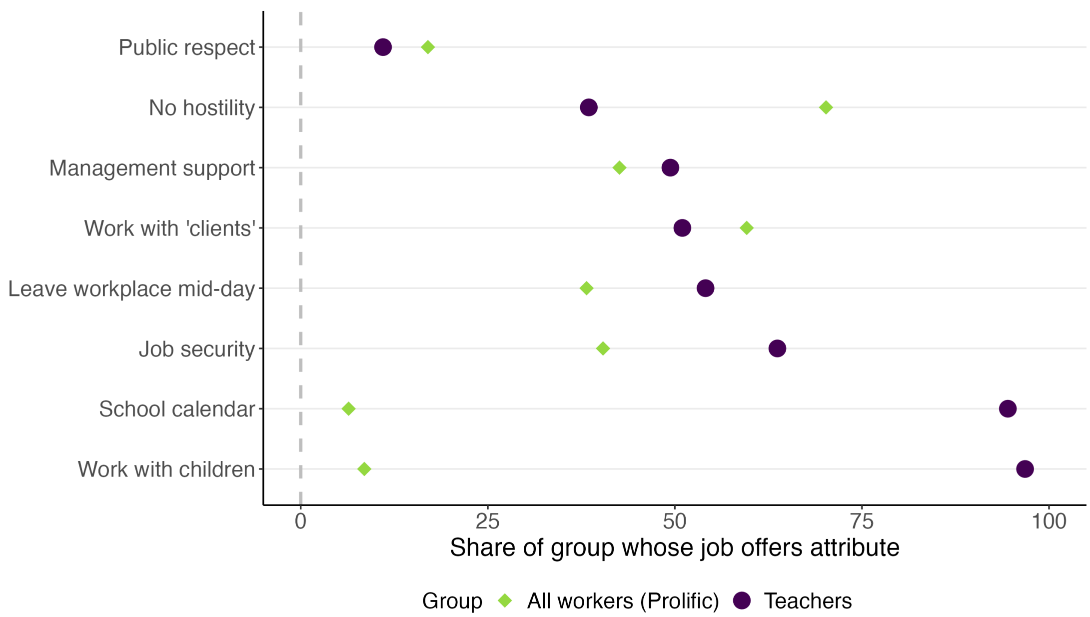
(b) Quarterly earnings.

Figure VII: Exiting Teachers' Earnings Trajectories

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.



(a) AWCS attributes



(b) Teaching-specific attributes

Figure VIII: The Incidence of Working Conditions Across Jobs

Note: This figure shows the percentage of workers who report having each of 17 desirable attributes. Figure (a) shows 4 groups: KY teachers (circles, from my survey), all workers in the US (diamonds, from the AWCS), college educated workers in the US (squares, from the AWCS), and workers in the education and health sector with below-median earnings in their sectors (triangles, from the AWCS). Figure (b) shows 2 groups: KY teachers (circles, from my survey) and all workers (diamonds, from my survey). The attributes are sorted from least common in teaching (top) to most common in teaching (bottom).

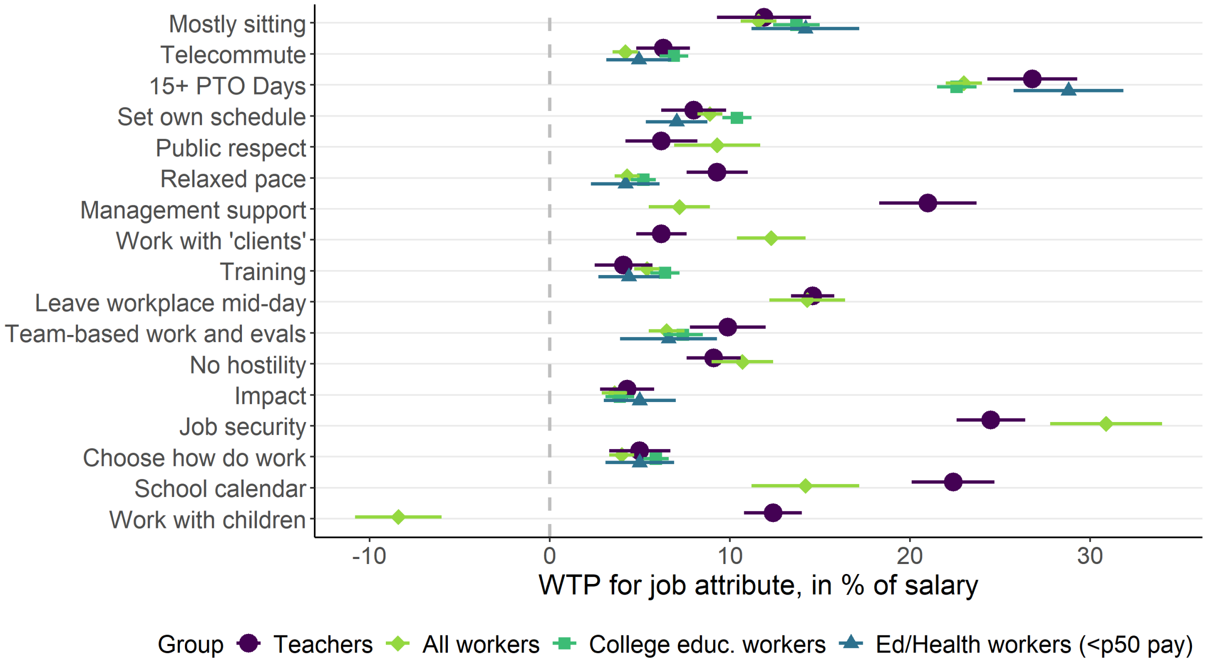


Figure IX: Willingness-to-pay Estimates

Note: This figure shows willingness-to-pay estimates for each of the nine job attributes and for 4 different groups: KY teachers (circles, from my survey), all workers in the US (diamonds, from the AWCS and my survey), 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). Note that for the attributes originally from the AWCS, the "all workers" estimates use the AWCS estimates, while for the teaching-specific attributes from my survey, the "all workers" estimates use my survey data. The attributes are sorted from least common in teaching (top) to most common in teaching (bottom).

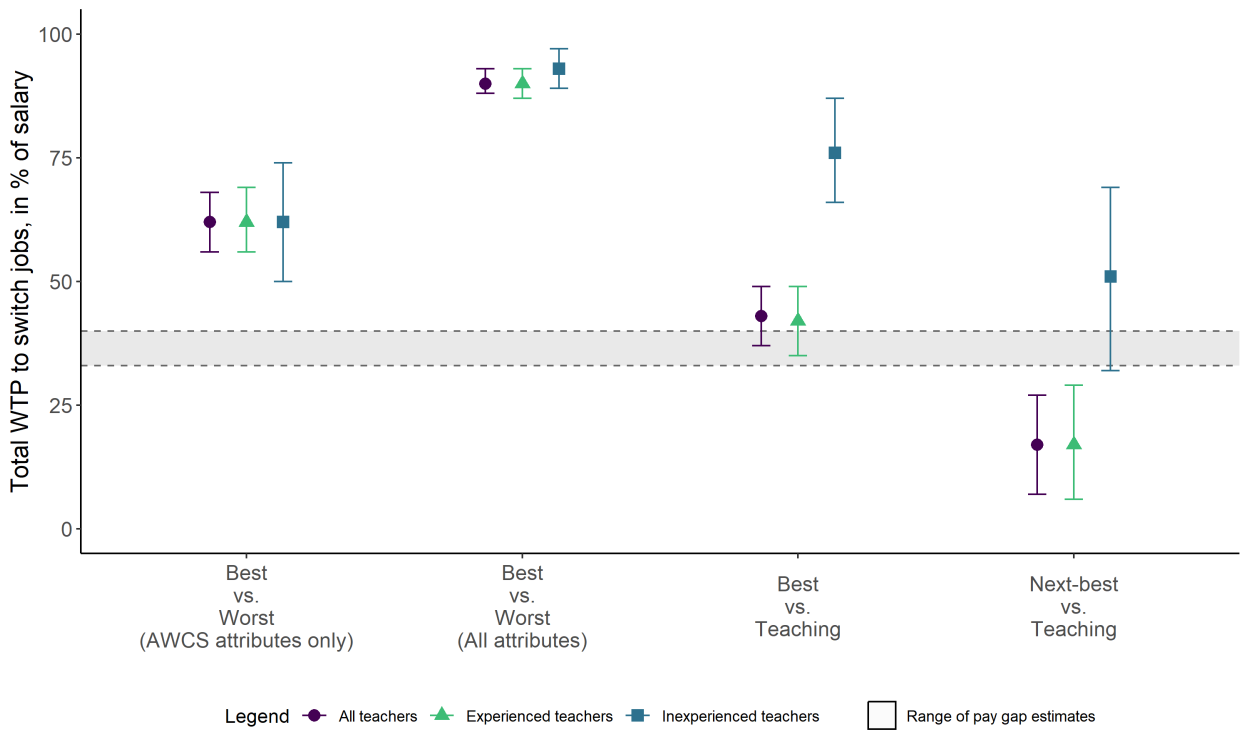


Figure X: Estimates of Total Willingness-to-pay to Switch Jobs, By Experience

Note: This figure shows estimates of KY teachers' willingness to pay for one bundle of attributes over another, using data from the choice experiments. 95% confidence intervals are shown, with standard errors estimated using the delta method and clustered at the respondent level. The shaded region shows the range of pay gap estimates from the quasi-experimental designs in the first half of the paper, ranging from 33-40% of the teachers' salary.

Supplementary Material

I.A. Identification Proofs

Here I 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.

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 $Y_{NC,i} = y(0, U)$.

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

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

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

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

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

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 $s \geq 0$:

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

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

Similarly, taking the limit from the left yields

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

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

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

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

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

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

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

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

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

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

From (17), we also see that depending on the setting, the discontinuity can be positive or negative. Although $p_c(u)$ is smooth through the cutoff, the retaking function $p_r(s,u)$ is not; those randomly assigned to the right of the cutoff retake with probability zero, while those randomly assigned to the left may retake with a positive probability. The discontinuity could be positive if those to the left retake with some constant probability that is independent of u. But, the discontinuity could be negative if for those with $s < 0$, $p_r(s,u)$ is only positive for people with high $p_c(u)$, leading to

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

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

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

where the weights are given by

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

We can interpret the $\phi_{RD}(u)$ terms as weights as they are positive for all u. The RD estimand is most heavily weighted towards "compliers" as defined in Supplementary Materials I.C: 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.[63]

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.

However, in this paper's 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, meaning the RD design cannot be applied. Instead, the same random assignment induces 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.

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

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

is given by

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

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

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

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

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

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

and from the left hand side, we have

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

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

$$
\begin{aligned}
&\lim_{s \to 0^-} \frac{d\mathbb{E}[y(C,U) \mid S_1 = s]}{ds} - \lim_{s \to 0^-} \frac{d\mathbb{E}[y(C,U) \mid S_1 = s]}{ds} \\
&= \int [y(1,u) - y(0,u)]\, p_c(u) \times \\
&\qquad \left\{ \lim_{s \to 0^+} \Lambda(s) - \Pr(S_2 \geq 0 \mid u, R = 1) \lim_{s \to 0^-} \left[\frac{\partial p_r(s,u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} + p_r(s,u)\Lambda(s) \right] \right\} dF_U(u) \\
&= \int [y(1,u) - y(0,u)] \times \Bigg(\lim_{s \to 0^+} p_c(u)\Lambda(s) \\
&\qquad\qquad - \lim_{s \to 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) p_r(s,u)\Lambda(s) \qquad (24) \\
&\qquad\qquad - \lim_{s \to 0^-} \Pr(S_2 \geq 0 \mid u, R = 1) \frac{\partial p_r(s,u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} \Bigg) dF_U(u) \\
&= \underbrace{\int [y(1,u) - y(0,u)] \cdot \Lambda(0) \cdot \left(\lim_{s \to 0^+} p_c(u) - \lim_{s \to 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) p_r(s,u) \right) dF_U(u)}_{\text{(a)}} \\
&\qquad - \underbrace{\int [y(1,u) - y(0,u)] \left(\lim_{s \to 0^-} p_c(u) \Pr(S_2 \geq 0 \mid u, R = 1) \frac{\partial p_r(s,u)}{\partial s} \frac{f_{S_1|u}(s)}{f_{S_1}(s)} \right) dF_U(u)}_{\text{(b)}}.
\end{aligned}
\tag{25}
$$

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

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

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

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

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

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

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

where the last line follows from equation (17).

With unrestricted heterogeneity in u, it is possible that the first term is positive or negative. However, recall that there is virtually no discontinuity in $\Pr[C = 1 \mid S_1 = s]$ at the cutoff, which is only possible if $\frac{\partial p_r(s,u)}{\partial s} > 0$ for $s < 0$; or in other words, scoring higher on the exam makes one more likely to retake (the "encouragement" effect). Thus, under the assumption there is no discontinuity in $\Pr[C = 1 \mid S_1 = s]$, the entire (a) term is zero.

As a result, the numerator is given by only (b),

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

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

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

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

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

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

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

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

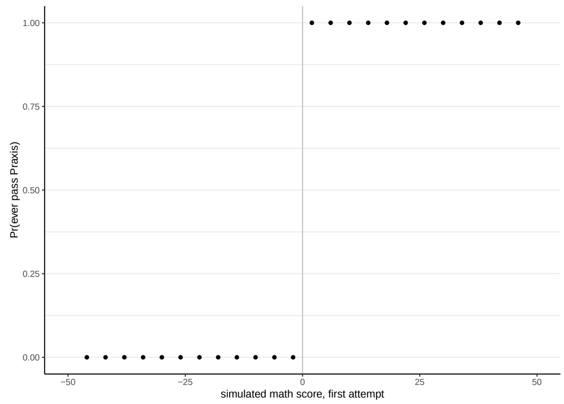
The weights are given by

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

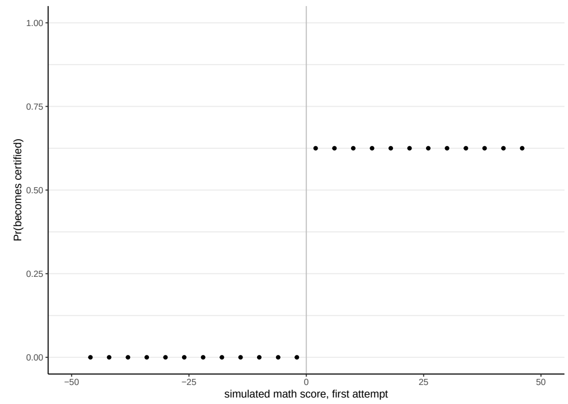
We can interpret $\psi_{RKD}(u)$ as weights because $\psi_{RKD}(u)$ is non-negative for all u. The RK estimand gives no weight to individuals whose retaking behavior does not respond to their score, thus excluding individuals with such low intrinsic motivation that they would never retake the test, as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

I.B. Model simulations: From sharp RD to fuzzy RK

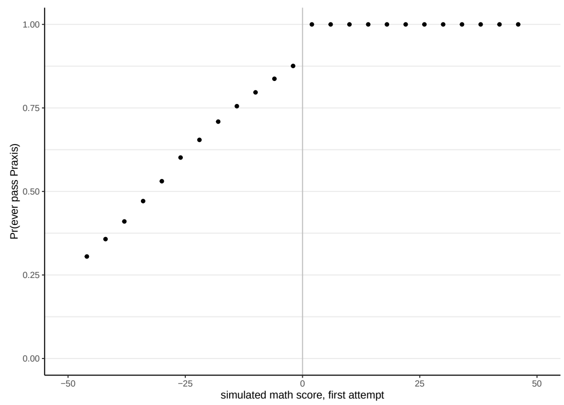
This appendix provides a series of simulations that illustrate when and why a kink might arise in place of a discontinuity under dynamic regression discontinuities. First, we show how a RD design approximates a randomized experiment if test-takers were not allowed to retake the test. Second and third,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



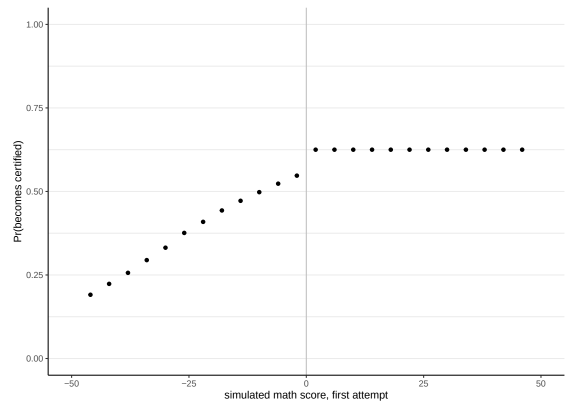
(a) Simulated Pr(pass), one-shot



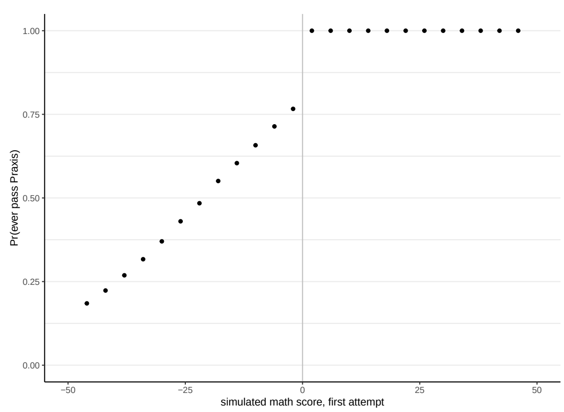
(b) Simulated Pr(certified), one-shot



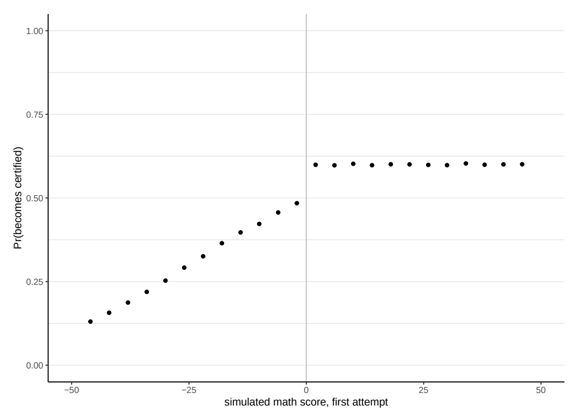
(c) Simulated Pr(pass), with automatic retaking



(d) Simulated Pr(certified), with automatic retaking



(e) Simulated Pr(pass), with behavioral retaking



(f) Simulated Pr(certified), with behavioral retaking

Figure XI: 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).

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: $C = C(\{S_1, S_2\}, R, \xi)$,[64] 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 XI 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.[65] Those who do not pass can never pass 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

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

[65] We assume that among those who pass, high-scorers are no more likely than barely-passers to enroll in and complete an EPP. However, as we know from the descriptives in Section ??, many who pass the Praxis do not continue through an EPP. We therefore choose parameters in our simulation to replicate this observation, with passers having a 2/3 probability of becoming certified.

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.

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 XI 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.

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 the "encouragement" effect: Getting a score closer to the cutoff increases the likelihood one retakes the exam. Under this assumption, 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 XI 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.

Reconciling the model and data. The last row of Figure XI can be contrasted with the actual first-stages in the data shown in Figure I. 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.

I.C. 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 Supplementary Materials I.A 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.

I.D. 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 KY 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.

Wage elicitation. To obtain each respondent's hourly wage, I ask respondents to report their teaching earnings at the hourly, weekly, biweekly, twice-monthly, monthly, or annual level, along with their hours worked per week and weeks worked per year.

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.

Wage randomization. Hypothetical wages for Jobs A and B are generated by scaling the respondent's actual hourly wage w_i by $\theta_A w_i$ and $\theta_B w_i$, where $\theta_A, \theta_B \sim N(1, 0.1^2)$, truncated at 0.75 and 1.25 to keep wages within a plausible range. Hours are randomly varied in 5-hour intervals between 15 and 60 hours per week.

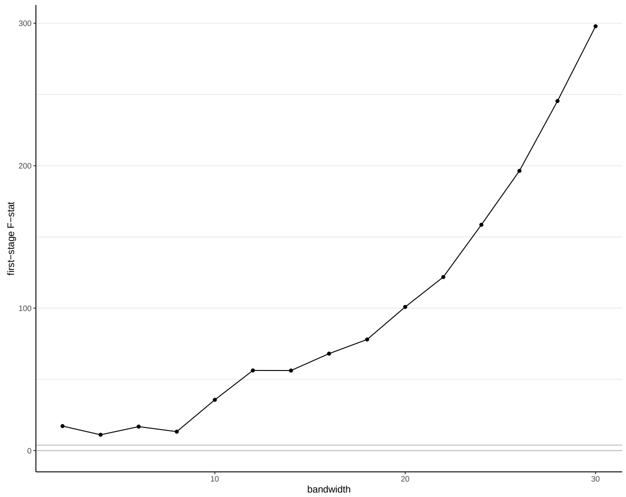
Limiting cases where one jobs dominates another. We used the willingness-to-pay estimates from the college-educated workers in Maestas et al. (2023) as a benchmark with which to rank job attribute values, but not the attributes themselves. When one of the randomly generated jobs dominated the other on 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.

Robustness checks. I conduct several specification checks on the willingness-to-pay estimates: (i) excluding "inattentive" respondents who incorrectly answered both trick questions, (ii) estimating a probit model instead of a logit to evaluate sensitivity to the error term distributional assumption, (iii) excluding experiments that used the "common" baseline job instead of the respondent's own baseline job, and (iv) estimating the model without survey weights. The estimates are robust to all of these variations.

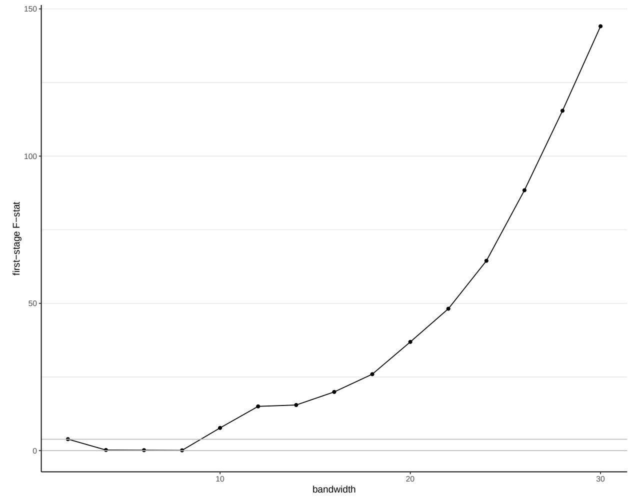
Figure XII: "Word Cloud" of Factors Affecting Teachers' Job Choice

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?"

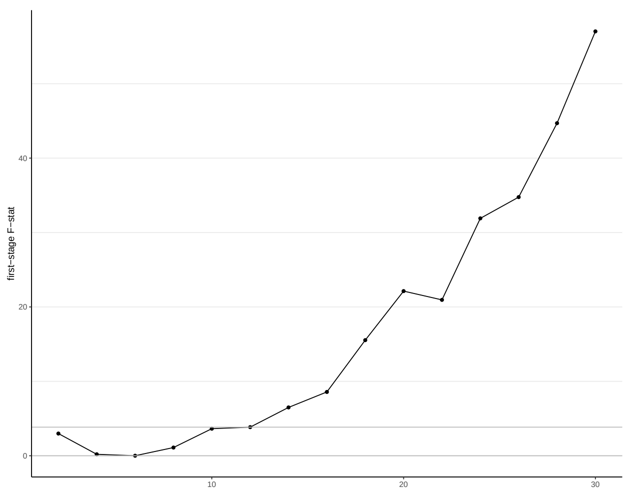
I.E. Additional Figures



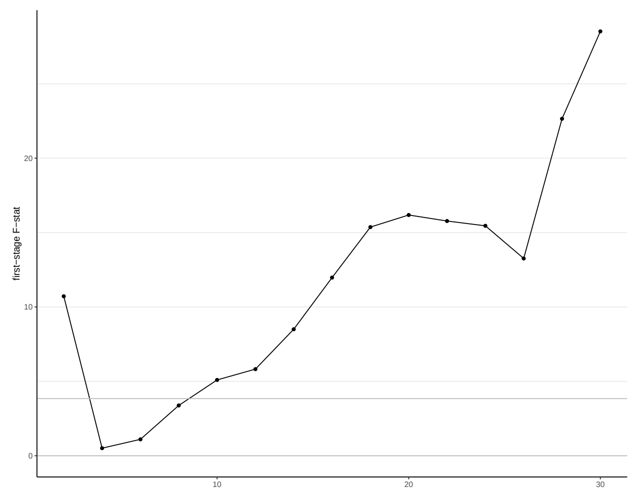
(a) Ever-passing the math test.



(b) Ever-passing all 3 tests.

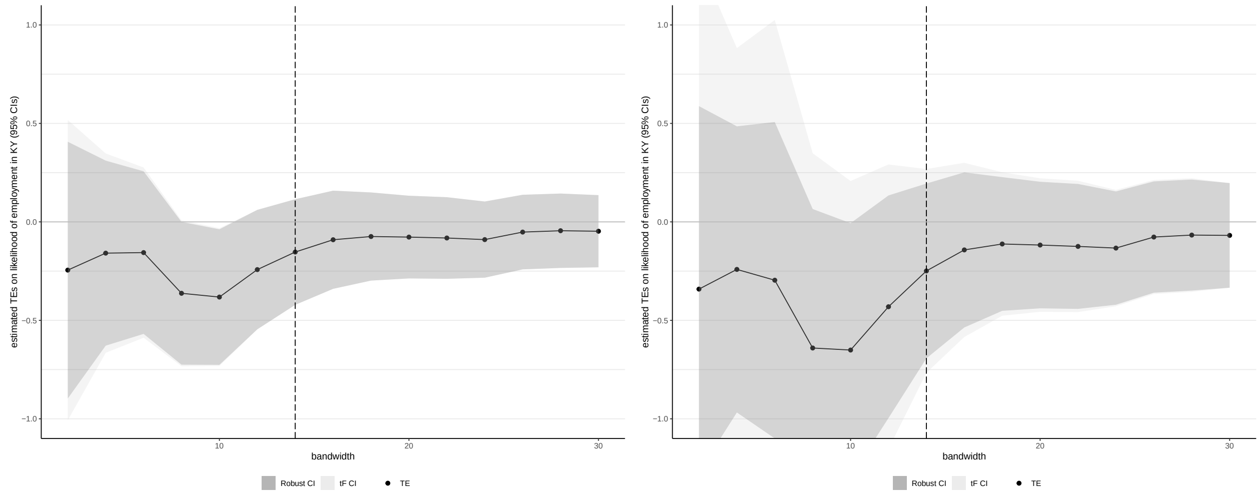


(c) Becoming certified within 4yrs.



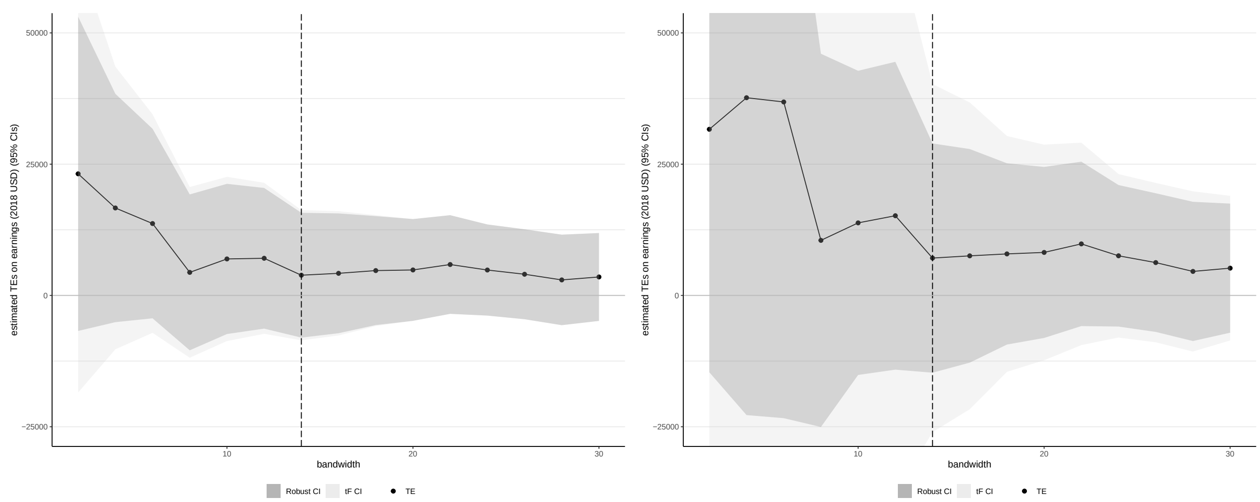
(d) Being a teacher in 4yrs.

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



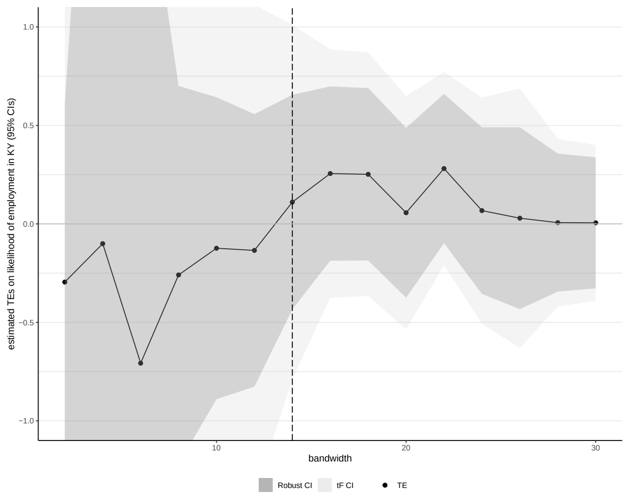
(a) Effect of ever-passing math on Pr(employed). (b) Effect of ever-passing all 3 on Pr(employed).

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(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?	○	○	○	○

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

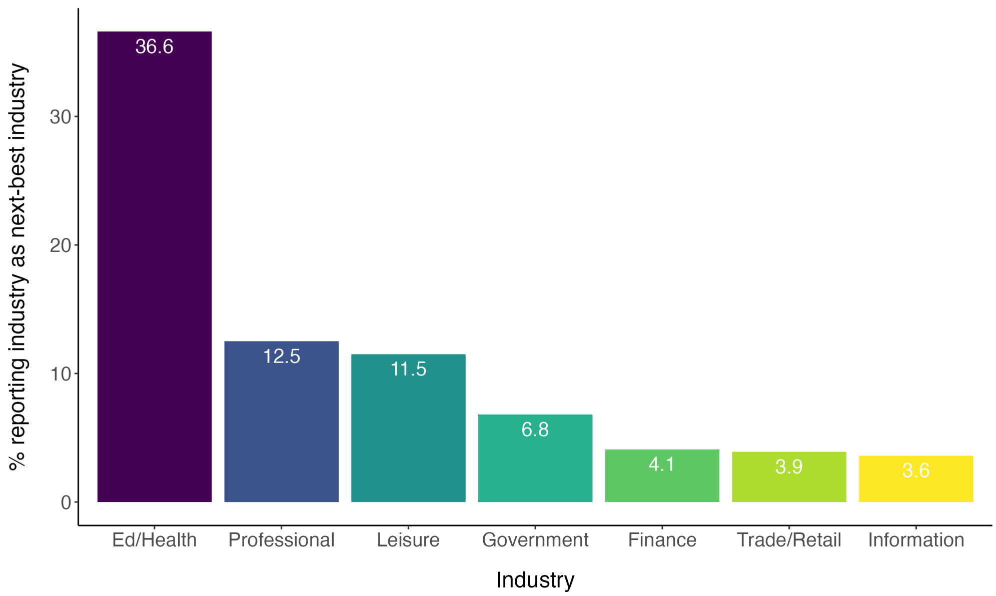


Figure A.6: Most Popular Next-Best Industries Reported by Kentucky Teachers

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?"

I.F. Additional Tables

Table A.1: Incidence of Working Conditions in Teaching and the AWCS

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

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

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

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 A.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 A.3: Comparison of Working Conditions in Teaching to Below-Average Paying Jobs in Other Industries in 2015 (AWCS)

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

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

Table A.5: Regressions of Working Conditions on Demographic Variables (part 1)

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

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

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

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

		AWCS		
	Teachers (1)	All workers (2)	College-educ. (3)	Ed/Health, < median wage (4)
Sets own schedule	0.080	0.089	0.104	0.071
vs. Schedule set by manager	(0.018)	(0.007)	(0.008)	(0.017)
Telecommute	0.063	0.042	0.069	0.050
vs. No telecommuting	(0.015)	(0.007)	(0.008)	(0.018)
Moderate physical activity	0.159	0.145	0.168	0.197
vs. Heavy physical activity	(0.021)	(0.010)	(0.013)	(0.026)
Sitting	0.119	0.116	0.137	0.142
vs. Heavy physical activity	(0.026)	(0.010)	(0.013)	(0.030)
Relaxed	0.093	0.043	0.052	0.042
vs. Fast pace	(0.017)	(0.007)	(0.007)	(0.019)
Choose how do work	0.050	0.040	0.059	0.050
vs. Tasks well defined	(0.017)	(0.007)	(0.007)	(0.019)
10 days PTO	0.196	0.164	0.158	0.197
vs. No days PTO	(0.024)	(0.009)	(0.010)	(0.025)
20 days PTO	0.268	0.230	0.226	0.289
vs. No days PTO	(0.025)	(0.010)	(0.011)	(0.031)
Team-based, evaluate own	0.099	0.065	0.074	0.067
vs. Team-based, evaluate team	(0.021)	(0.010)	(0.011)	(0.027)
Work by self	0.048	0.086	0.073	0.097
vs. Team-based, evaluate team	(0.022)	(0.010)	(0.012)	(0.026)
Training opportunities	0.041	0.054	0.064	0.044
vs. Already have skills	(0.016)	(0.007)	(0.008)	(0.017)
Frequent opp. to serve	0.043	0.036	0.039	0.050
vs. Occasional opp. to serve	(0.015)	(0.007)	(0.008)	(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 A.8: OLS estimates of cross-sectional differences in earnings between teachers and non-teachers among KY college graduates

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

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