

Preliminary and Incomplete- Please do not cite or circulate

**Premium or penalty?
Labor market returns to novice public sector teachers***

Juan Esteban Saavedra (University of Southern California and NBER)

Darío Maldonado (Universidad de los Andes)

Lucrecia Santibañez (Claremont Graduate University)

Luis Omar Herrera Prada (IADB)

November 7, 2016

Abstract

It is unclear whether public sector teachers are under or overpaid relative to other occupations because we generally lack knowledge on teachers' outside labor market options or other unobserved attributes related to compensation. We document causal labor market returns to public sector novice teachers in Colombia. In 2004, Colombia introduced a national, standardized, teacher-screening exam, scores on which determine eligibility for public sector teaching jobs relative to an exogenous cutoff point. We combine four nationwide administrative data sources in a regression discontinuity approach to document the labor market returns and outside options of recent education college graduates who take the screening test. Applicants who marginally pass the teacher screening test have annual earnings that are 15, 20 and 28 percent greater during the first three years of tenure, respectively, relative to those of applicants below the passing cutoff. Part of the earnings premium stems from the fact a substantial fraction of public sector teachers hold an outside, predominantly non-teaching job in the formal sector. The total earnings effect is the combination of greater labor supply during the first three years of tenure in the main teaching job and outside jobs and higher daily wages.

* Saavedra (corresponding author): juansaav@usc.edu. We acknowledge funding from the Inter American Development Bank (IADB). We thank Mariana Alfonso, Ivan Enriquez and Carlos Hipolito Garcia, Abhijit Singh and participants at the CESifo Economics of Education Conference 2016 for helpful comments and suggestions. Olga Victoria Dulce at Universidad de los Andes provided excellent research assistance in Colombia. We thank Dolly Ovalle and Luz Emilse Rincón at Colombia's Ministry of Social Protection for providing access to social security employment and earnings data. Colombia's Ministry of Education provided access to teacher rosters. All errors are our own.

1. Introduction

In an effort to raise educational quality, many countries grapple with the question of how to attract high quality teachers. Initiatives often target teacher compensation, arguing that public teachers are underpaid relative to similarly qualified professionals in other occupations (e.g. Auguste, Kim and Miller, 2010; Mizzala and Nopo, 2016). It is unclear, however, whether public sector teachers are under or overpaid because we often lack knowledge on teachers' outside labor market options, cognitive ability, and other unobserved attributes related to compensation.

In this paper we shed light on these questions. We investigate the labor market returns to novice public sector teachers in Colombia. In 2004, Colombia introduced a national, standardized, teacher-screening exam, scores on which determine eligibility for public sector teaching jobs relative to an exogenous cutoff point. We combine four nationwide administrative data sources in a fuzzy regression discontinuity approach to document the labor market returns and outside options of recent education college graduates. To our knowledge this is the first paper to document causal evidence on these questions.

If the test has screening value and is costly, it creates an entry-barrier effect that unambiguously results in higher *average* teacher wages (Angrist and Guryan, 2007). It is theoretically unclear, however, the extent to which the labor market returns to marginal applicants are positive or negative.

We find that novice public sector teachers earn a substantial earnings premium. Applicants who marginally pass the teacher screening test have annual earnings that

are 15, 20 and 28 percent greater, respectively during the first three years of tenure than earnings of applicants just below the test passing cutoff. Part of the premium stems from the fact that a substantial fraction of public sector teachers hold outside, predominantly non-teaching jobs in the formal sector. As a result, relative to applicants that fail, applicants who pass the teacher screening test work more days per year and have higher daily earnings, presumably because they work more hours (even though we do not observe hourly wages).

We take no stand as to whether this sizeable earnings premium for the marginal novice public sector teacher results in a better pool of public sector teachers. However, various pieces of evidence suggest that it may not. First, to the extent that the test is costly, some applicants will choose not to take it. Therefore, applicants on the margin between public teaching and an alternative occupation are the highest-quality teachers (Angrist and Guryan, 2007). Consistent with this theoretical prediction, we find that relative to recent education college graduates who take the screening test, those who choose not to take it have higher college admission exam scores and come from more socio-economically advantageous backgrounds. This finding is consistent with evidence from other public sector jobs that shows how raising salaries improves the applicant pool quality but does not affect overall performance (Dal Bo et al, 2013; Ferraz and Finnan, 2009).

Second, economic theory highlights how effort depends on marginal incentives (e.g. Holmstrom and Milgrom 1991; Holmstrom 1999). Despite having a probationary period, novice teachers in Colombia are rarely dismissed. Moreover, their public sector

teacher pay is untied to performance. Therefore, it seems plausible that, on the margin, novice public teachers will devote more effort to their outside, less secure, non-teaching job, possibly at the expense of teaching.

Third, evidence from other developing countries indicates considerable misallocation of pay and productivity among public sector teachers. In Pakistan, for example, reducing teacher pay by 35 percent had no adverse effect on teacher productivity, as measured by test-score value added (Bau and Das, 2016). In Indonesia, doubling teacher pay reduced the fraction of teachers holding outside jobs, but did not affect teacher effort or student performance (de Ree et al., 2015).

The rest of the paper proceeds as follows. Section 2 describes the structure of public teacher hiring in Colombia and the teacher screening test reform introduced in 2004. Sections 3 and 4 describe the data and research design. Section 5 presents our main findings. Section 6 concludes.

2. Public teacher hiring in Colombia

Evidence suggests that it is difficult to identify effective teachers at the point of hire: observable characteristics such as educational background are poor predictors of performance on the job (Rockoff, Jacob, Kane, and Staiger, 2011).

To improve the quality of the teaching force, countries around the world have introduced licensure requirements for prospective teachers. In England, for example, prospective teachers must pass the Teacher Training Agency skills test to become public sector teachers (Wang et al., 2003). In the US, more than 40 states require teachers to pass some sort of standardized certification test (Angrist and Guryan, 2007). Mexico

and Peru have similar teacher screening test as the one in Colombia that we study (Estrada, 2013).

In 2002, the Colombian government modified the conditions that determine entry into the public teaching profession. This reform was undertaken with the goal of raising entry standards for public sector teaching and making the selection process more transparent and meritocratic. The new selection rules went into effect in 2004.

The new recruitment process has three stages. In the first stage, aspiring teachers take a screening test; eligibility depends on the result of the test. Only applicants with scores at or above 60 points are eligible for public sector teaching jobs. The 60-point cutoff is fixed *ex-ante* by the Ministry of Education.

The teacher screening test assesses applicants' cognitive ability, content knowledge, and personality traits. The test has substantial power as a quality screen. In 2009 – the test application cohort we study – only 42 percent of applicants passed the test. Moreover, the probability of passing the test is correlated with applicants' outside option. Among recent education college graduates who take the 2009 screening test, scores on the test are positively correlated with earnings in the prior year.

In the second stage of the teacher selection process, officials from the Ministry of Education interview applicants with scores above the eligibility cutoff and verify applicant qualifications and degrees. At this second stage the government disqualifies applicants who lack minimum qualifications, such as having a Bachelor's degree. Qualified applicants move on to the third and final stage of the selection process. In the third stage, applicants choose the vacancy that he/she wants to fill. The order of choice

follows the score in the initial screening test: the applicant with the highest score chooses first and so on until the last vacancy is filled.

Since the reform was enacted there have been five teacher screening test administrations: 2004, 2005, 2006, 2009 and 2013. We focus on graduates who took the 2009, due to data restrictions, as explained in the data section below. The teacher screening test, as we show below, creates plausibly exogenous variation in the probability of obtaining a permanent public sector teaching position around the eligibility threshold.

3. Data and Methods

Data

Our dataset combines information from four administrative sources. These sources are merged using national identification information on names, dates of birth and adult identification (ID) numbers:

1. SPADIES data, 2004-2009. Contains individual-level information on all of the education graduates. Data include key baseline covariates such as gender, graduation and entry year, whether the individual received financial aid during college, time needed for graduation, and, the university and program from which each individual graduated which allows to have some measures of university quality. .
2. Teacher-screening test, 2004, 2005, 2006, 2009. These data contain individual-results on all individuals who took the teacher screening tests. Information

comes from Colombia's ICFES agency¹.

3. Public sector teacher payroll data, 2008-2013. Using these data we can identify which graduates hold permanent positions as public sector teachers, and what their monthly wages are.
4. National Social Security earnings records, 2008-2013. Includes data on all formal sector employees in the country, including monthly wages, employment intensity (days worked in the formal sector per year) and occupational codes. Importantly, the social security earnings records do not cover public sector teachers (or military officers) since these sectors have independent social security regimes. Therefore, to the extent that public sector teachers appear in the National Social Security files—and many do—it implies that they hold another formal sector job outside public teaching. This is something we investigate in detail in our empirical analyses.

Only about 60 percent of the 41,703 education graduates from the 2004-2009 graduating cohorts take the 2009 teacher-screening test. Because the cost of taking the test (monetary and in terms of effort) is likely common to all applicants, applicants on the margin of becoming public sector teachers or choosing a different occupation will be the highest-quality teachers and have the best outside options.

Table 1 shows mean comparisons of education graduates from graduating cohorts 2004-2009 who took the 2009 test and of those who did not.

[Table 1 HERE]

¹ ICFES (Instituto Colombiano para el Fomento de la Educación Superior) is the institution in charge of designing and administering standardized exams in Colombia.

Consistent with the “discouragement” hypothesis of a costly screening test (e.g. Angrist and Guryan 2007) non-test takers are potentially the best teachers to the extent that they are more likely to come from socio-economically advantaged households (as measured by income and the probability of having a mother with a college degree), have higher secondary graduation exam scores—which in Colombia are the main college admission criterion—and tend to graduate from better colleges (as measured by a higher fraction who graduate from a nationally certified college institution). Relative to non-test takers, test takers are also more likely to be female.

Among those who took the test, we have complete covariate and outcome information for 99.8 percent (24,791 graduates). Of these, 40 percent (9,874 graduates) also took the exam in at least one of the prior administrations (2004, 2005 or 2006). Exam repetition may raise concerns for test-score manipulation. While we show balance just above and just below the cutoff in covariates and density, we repeat all estimations without exam repeaters; our main findings remain unchanged.

Because earnings data begin in 2008, we limit our sample to education graduates who took the 2009 teacher- screening test. However, because our payroll and social security data contain information until 2013, we can observe labor market outcomes one, two, three and four years after an individual took the teacher-screening test.

[Table 2 HERE]

Formal labor force attachment is high in the study’s population (Table 2). In the two years of pre-exam labor market data we have (2008 and 2009) over 90 percent of 2009 test takers worked in the formal sector, as measured by being matched to social

security records. Over time, the fraction of test-takers with formal sector attachment diminishes as a higher fraction of eligible test-takers enter public sector teaching, which reaches about 26 percent in 2012-2013 (column 2, Table 2). Combining public teacher payroll data and social security data permits us to have labor market outcomes data during the 2008-2013 period for over 92 percent of test-takers (column 4, Table 2). A non-trivial percent of the sample—ranging from 7 to 10 percent in the years after the test (2010-2013)—appears in both teacher payroll and social security data, implying that, in addition to being public sector teachers, these teacher screening test takers simultaneously hold another public sector job (column 5, Table 2). As a fraction of public sector teachers, those who hold outside formal jobs range between 25 and 50 percent depending on the year.

Empirical Strategy

For the 2009 teacher screening test administration, only applicants with scores at or above 60 points were eligible for a public sector teaching position, as determined ex-ante by Ministry of Education officials. Figure 1 shows that the rule generates a highly non-linear relationship between score points and the probability of obtaining a public sector teacher position around the 60-point threshold.

[Figure 1 HERE]

Three aspects of Figure 1 are worth highlighting. First, virtually no applicant below the test-score obtains public sector teaching positions. This confirms that the cutoff rule was generally binding. Second, there is a discontinuity around zero (normalized score) in the probability of obtaining a public sector position. Third, the

discontinuity at the cutoff grows in magnitude over time.

The discontinuity grows over time because public sector teaching positions are filled in a descending priority order beginning with the highest scores. Many applicants who meet the admission cutoff do not obtain a public sector position immediately, because there are more eligible applicants than vacancies. As more vacancies open over time, these applicants are offered positions. This assignment rule also explains why the slope to the right of zero flattens over time: applicants with lower scores gain access to new vacancies. The first stage ranges from 23 percent in 2010 to 50 percent in 2012 and 2013.

The reduced-form regression equation we estimate is:

$$Y_i = \alpha_0 + \alpha_1 Eligibility_i + F(Score_i) + X_i' \delta + \varepsilon_i \quad (1)$$

where $Eligibility_i$ is a dummy variable that takes value of one if individual i had a score of 60 or more points in the screening test, $Score_i$ represents the score in the screening test normalized to be zero at the 60-point cutoff and X_i a vector of covariates that we include in some specifications as a robustness check. We impose $F(\cdot)$ to be a quadratic function of the normalized score.

As Figure 1 shows, not all individuals with a qualifying score necessarily end up working as public sector teachers. To estimate the effect of obtaining a public sector position on labor market outcomes, we instrument obtaining a position with the *Eligibility* variable, which as Table 2, shows, produces a very strong first stage. The regression equations that we estimate in this Fuzzy Regression Discontinuity research design are:

First Stage:

$$P(T_i) = \pi_0 + \tau_1 \text{Eligibility}_i + F(\text{Score}_i) + X_i' \delta + \varepsilon_i \quad (2)$$

Second stage:

$$Y_i = \tau_0 + \tau_1 \hat{P}(T_i) + F(\text{Score}_i) + X_i' \delta + \varepsilon_i \quad (3)$$

Without the vector X of controls, τ_1 is identical to the IV-Wald estimator, equivalent to the ratio of the reduced form in the outcome of interest at the cutoff to the first stage of the probability of obtaining a public sector position at the cutoff.

4. Validation of the Regression Discontinuity Research Design

The plausibility of the regression discontinuity approach rests on the assumption of continuity of potential outcomes at the eligibility cutoff. We show three pieces of evidence in support for this assumption: the density of scores in the screening test at the cutoff, covariate continuity at the cutoff and continuity of pre-test labor market outcomes.

The density of scores around the admission cutoff does not evidence any discontinuous jump that would indicate manipulation (Figure 2). Therefore, we feel confident that manipulation of applicants' position relative to the cutoff is not a concern in our setting, even despite the possibility of test retake.

[Figure 2 HERE]

Figures 3a and 3b show evidence of covariate continuity measured before or during college around the eligibility cutoff. On Figure 3a, we examine the smoothness around gender (female), proportion of graduates from a public university, proportion of graduates that received financial aid and proportion of graduates that graduate on-

time. Figure 3b shows graduation year, proportion of graduates from a certified institution and college exit test scores. As these figure show, there is no evidence of discontinuities at the eligibility cutoff in the characteristics of education graduates taking the 2009 teacher-screening test.

[Figures 3a and 3b HERE]

The most compelling evidence in support for the validity of the RD design strategy in this settings come from an examination of continuity in pre-test labor market outcomes. Figure 4a shows formal employment probabilities, annual earnings and labor supply (payroll days/year) in 2008. Figure 4b shows the same outcomes for 2009, the year in which applicant take the screening test. For both years, these figures show continuity of these different labor market outcomes, further substantiating the validity of the regression discontinuity research design.

[Figures 4a and 4b HERE]

5. Results

We report results on total annual earnings, labor supply and daily earnings. Monetary figures are in 2013 US dollars throughout). We then present evidence on the distribution of outside occupations and the relevance of the LATE treatment parameter in this context.

Earnings

We use two annual earnings measures: annual earnings from the main occupation and total annual earnings. In a given year, annual earnings from the main occupation are annual earnings (including zeros) from the teacher payroll from test-

takers who work as teachers and are annual earnings from the occupation with the highest number of payroll days/year for test-takers who are not teachers. Total earnings for all occupations are the sum of public teacher payroll earnings and social security earnings, including zeros. Recall that these two data sources measure earnings from mutually exclusive jobs.

Figure 5 shows total annual earnings around the screening test eligibility cutoff. The blue labels refer to earnings from the main occupation and the black labels refer to earnings from all occupations. Dots are conditional means with corresponding 95 percent confidence intervals. The fit is a non-parametric lowess fitted separately to each side of the cutoff. Earnings are an increasing function of scores on the teacher screening test.

The first result we highlight from Figure 5 is the jump at the cutoff in annual earnings. Based on the main occupation (blue), this jump ranges from about \$500/year in 2010 to about \$1,800/year in 2012. Relative to the mean just below the cutoff of around \$6,000, the jump represents an increase of 8 percent in 2010 to close to 30 percent in 2012, after three years of potential teacher tenure. Scaling up these differences at the cutoff by the corresponding probability of being a public sector teacher implies an annual earnings premium at the cutoff that ranges from \$2,248 in 2010 to \$3,607 in 2012 (Panel A, Table 3). Relative to earnings from main occupation just below the cutoff, the jump represents an increase of 37 percent in 2010 to 60 percent in 2012.

[Figure 5 HERE]

The second result from Figure 5 is that the annual earnings jump at the cutoff

may actually be bigger once we account for earnings from all occupations. This is particularly the case for test-takers above the cutoff, where the discrepancy between earnings from all occupations (black) and main occupations (blue) is largest. The premium at the cutoff accounting for earnings from all occupations is \$870/year (15 percent relative to base) in 2010, \$1,214 in 2011 (20 percent relative to base) and \$1,753 in 2012 (28 percent relative to base).

Scaling up the differences at the cutoff in earnings from all occupations by the corresponding probability of being a public sector teacher implies an annual earnings premium at the cutoff that ranges from \$3,710 in 2010 to \$3,476 in 2012 (Panel B, Table 3). Relative to earnings from main occupation just below the cutoff, the jump represents an increase of about 60 percent in the first three years of teacher tenure.

These numbers imply that in the first year as public sector teachers, 39 percent of the annual earnings premium is the result of earnings from outside jobs. In the second year 7 percent is the result of earnings from outside jobs and in the third year it is all driven by the main teaching job.

The third key finding we highlight from Figure 5 and the corresponding Fuzzy RD results in Table 3 is that by 2013, four years after the teacher screening test, the premium becomes a penalty. This is all due to a substantial increase in total earnings from test-takers below the cutoff. While part of this increase may be the result of a steeper earnings experience profile in the occupations of test-takers below the cutoff, most of this increase is explained by a tax reform passed between 2012 and 2013 that gave formal sector firms tax incentives to hire as permanent workers employees who

previously were on temporary contracts. Evidence suggests that the reform increased substantially the fraction of permanent workers in formal sector firms (Kugler et al 2016). By changing substantially formal sector attachment and labor supply, as we show below, this reform changed dramatically the counterfactual earnings of test-takers just below the cutoff. By 2013, relative to test-takers below the cutoff, public sector teachers have annual earnings from all occupations that are \$3,030/year less – or about 25 percent less relative to earnings from all occupations for test-takers just below the cutoff.

Labor supply

Labor supply is number of payroll days per year in formal employment. Labor supply is an increasing function of scores on the teacher screening test (Figure 6). Relative to graduates just below the eligibility cutoff, education graduates just above the screening test cutoff work more payroll days per year, both in the main occupation (blue) and in all occupations combined (black). The difference at the cutoff is larger, however, when we account for all occupations, suggesting that increased labor supply of marginal test-takers is the result of more workdays per year in the main occupation as well as additional work effort in other occupations.²

[Figure 6 HERE]

In the main occupation, education graduates just above the screening test cutoff work, on average, 14, 36 and 56 more days per year in 2010, 2011 and 2012, than graduates just below the cutoff. In 2013, however, graduates just above the cutoff work

² Our data do not have hours worked, only days worked per year in each occupation. We do not know the number of hours worked each day for each individual.

51 fewer days per year (Base is about 200 payroll days/year just below the cutoff). The result for 2013 is likely due to the tax incentive to formal sector firms to offer permanent positions to employees on temporary contracts. Between 2012 and 2013, labor supply for test takers below the cutoff increases from about 220 payroll days per year to about 350 payroll days per year.

[Table 4 HERE]

When we instrument being a public sector teacher with *Eligible*, we find that in the first year after taking the test, public school teachers work in the main occupation about 61 days more than individuals not working as public school teachers—those just below the cutoff. In 2011 this difference in main-occupation labor supply is 99 days, and in 2012 it is 111.7 days (Panel A, Table 4). In 2013 public sector teachers just above the cutoff work in the main occupation 103 fewer payroll days per year than non-public teachers just below the cutoff.

Combining the labor supply for all occupations, education graduates just above the screening test cutoff work, on average, 28, 40 and 57 more payroll days per year in 2010, 2011 and 2012, than graduates just below the cutoff (Base rate is about 220 payroll days/year just below the cutoff). In 2013, graduates just above the cutoff work 55 fewer days per year. When we instrument being a public sector teacher with *Eligible*, we find that in the first year after taking the test, public school teachers work in all occupations about 120 days more than individuals not working as public school teachers—those just below the cutoff. In 2011 this difference in main-occupation labor supply is 110 days, and in 2012 it is 114.5 days (Panel B, Table 4). In 2013 public sector teachers just above

the cutoff work in the main occupation 110.7 fewer payroll days per year than non-public teachers just below the cutoff.

Daily Earnings

Daily earnings per year are total annual earning divided by total payroll days per year. We compute daily earnings for the main occupation and for all occupations. Like total annual earnings and labor supply, daily earnings are also an increasing function of scores on the teacher screening test (Figure 7).

[Figure 7 HERE]

Relative to education graduates just below the cutoff, those about the cutoff have greater daily earnings in the main occupation (blue), as evidenced by the discontinuous jump at the cutoff. In 2010, one year after taking the screening test, graduates just above the cut-off earn about \$0.53/day more than graduates below the cutoff, although this difference is not statistically significant. In 2011 the jump at the cutoff is \$1.96/day, and it is \$2.79/day and \$2.63 in 2012 and 2013, respectively. In 2010, 2011 and 2012—one, two and three years after the screening test—daily earnings from main occupation overlap with daily earnings from all occupation. In 2013—four years after the test—daily earnings from the main occupation for eligible candidates above the cutoff exceed average daily earnings from all occupations.

[Table 5 HERE]

When we instrument being a public sector teacher with *Eligible*, we find that in the first year after taking the test, the public-school-teacher daily earnings premium is \$2.75/day, although not statistically significant. The daily earnings in the main

occupation increases with potential experience: it is \$4.74/day in 2011, \$5.99/day in 2012 and \$6.40/day in 2013 (Panel A, Table 5). Results are very similar—even though point estimates are slightly smaller—when we focus on the daily earnings premium from all occupations (Panel B, Table 5).

The daily earnings results suggest that the total annual earnings premium for public sector teachers in the main occupation is—at the margin—a combination of a labor supply and a wage effect. The total earnings premium from all occupations for public sector teachers at the margin is, however, a labor supply effect. For 2013, in particular, there is a large positive wage (i.e. daily earnings) effect for teachers, yet the negative labor supply effect yields a negative total earnings premium.

Outside Occupations

The social security database contains 4-digit ISIC occupation codes³. These codes are available for 2008, 2010, 2012 and 2013. In these years, however, there are changes to the ISIC revision (Rev 3 or Rev 4) codes reported in the data. For education-related occupations, the codes are mostly consistent across revisions, but not entirely.⁴

Due to this inconsistency, we code occupations as being education related two ways. One way only codes as education-related occupations only those for which the occupation classification is unambiguously related to education in both revisions. This

³ ISIC stands for International Standard Industrial Classification of All Economic.

⁴ Education codes in CIU Rev. 3 are numbers: 8011 (preschool), 8012 (primary), 8021 (middle school), 8022 (high school), 8030 (higher education) and 8090 (other types of education). Of this only 8030 is used in Rev. 4 for a different activity (detectives and private investigators). Education codes in CIU Rev. 4 are 8511 (early childhood), 8512 (preschool), 8513 (primary), 8521 (middle school), 8522 (academic high school), 8523 (technical high school and job training), 8530 (institutions that combine different levels of education), 8541 to 8544 (technical and professional), 8551 to 8553, 8559 (other types of education), 8560 (other activities related to education). Among these three codes are used in CIU Rev. 4 for different activities: 8511 (health services institutions), 8512 (medical practice activities), 8513 (dentistry activities). When assuming that all codes that refer to education in Rev. 3 or 4 correctly identify a worker of the education sector we will assume that codes 8030, 8511, 8512 and 8513 effectively identify workers from the education sector. In the second case we will assume that these four codes identify workers in other sectors.

is the majority of education related occupations. Since this classification may miss some employees in education related occupations to the extent that they use one or the other revision and we cannot tell for sure. This first approach provides a lower bound estimate of the probability of working in a private, education-related job.⁵

The second approach assumes that occupations in codes that in one revision are education-related and in the other revision are not, are all education-related. This approach produces an upper bound of the proportion of workers in education related occupations. The bounds are quite tight. In Figures 8a and 8b we show outside occupation distributions based on the lower bound codification.

The occupational distributions for test screening test takers below the cutoff and for those above the cutoff who do not work as public sector teachers are remarkably similar, both in 2010 (one year after the test, Figure 8a) and in 2013 (four years after the test, Figure 8b). About 35 percent work in an education related occupation. The next most important occupational categories for these two groups are real state, business, finance and other services. For public sector teachers, on the other hand, a lower proportion works in education related outside occupations about 27 percent, as well as in real state, business, finance and other services and in manufacturing. By contrast, the majority (over 40 percent) of public sector teachers who hold outside formal sector jobs work in health, public administration, administrative and security related positions. In other words, an education related occupation is the outside option for less than one third of the teachers that work outside jobs, which are between 25 and 50 percent of

⁵ Recall that the social security database does not cover public sector teaching.

public teachers on a given year. To the extent that the skills required in those outside occupations differ from those skills mostly needed and honed through teaching, it seems plausible to conclude that there may be little skill complementarity between public sector teaching and outside jobs. There is however, a scheduling complementarity that arises from the fact that the Colombian schooling context of double shifts facilitates lumpy time allocations to outside occupations.

LATE and the distribution of earnings for all college graduates

Most of the discussion about raising the quality of public sector teachers centers around attracting top college graduates into the teaching profession (e.g. Garcia et al 2014; Bruns and Luque 2014). Therefore, critics may argue that the earnings premia and labor supply effects documented in this paper are only local to the teacher screening test cutoff, in other words, for the potentially least desirable eligible candidate. Two pieces of evidence suggest that—at least in the Colombian context—this LATE causal parameter is relevant. First, over one third of public sector teachers in Colombia are not even college graduates.

[Figures 9a and 9b HERE]

Second, even compared to college graduates from all majors, those at the cutoff do not look disproportionately negatively selected in the overall earnings distribution (Figures 9a and 9b). One year after the screening test, for example, education college graduates who take the teacher screening test and score at the eligibility cutoff rank at about the 45th percentile in the earnings distribution of all graduates. Education college graduates who score at the 95th percentile in the screening test are only slightly above

those at the cutoff in 2010 (Figure 9a). Median graduates from health sciences, social sciences, engineering and those education graduates who do not take the screening test are well above.

By 2013, four years after the screening test, differences in the centiles for the different groups fan out. Education graduates who score around the public teacher eligibility cutoff are at about the 55th percentile while the top education graduate test taker is at about the 80th percentile. Medians for health sciences, social sciences and engineering are located somewhere in between the cut and the top test taker. The top college applicant is located above all others in terms of earnings four years after the test. In sum, these figures suggest that, even though public teaching in Colombia is not attracting the top college graduates, education graduates who score at the margin of the screening cut off earn near the middle of the earnings distributions for all college graduates, which is potentially desirable given that one third of public sector teachers in Colombia are not college graduates.

6. Conclusions

In an effort to improve teaching screening, in 2002 Colombia underwent a major education reform changing the way teachers were selected for public school teaching positions. The central piece of the reform was a national, centralized teacher-screening exam. At the heart of these policy efforts is a determination to make the teaching profession more attractive to highly qualified and motivated individuals.

We take advantage of the fact that the screening exam creates exogenous variation on the probability of obtaining a permanent public sector teaching position to

examine earnings and labor supply differentials or those graduates that work as public sector teachers against those that work in alternative occupations.

Applicants who marginally pass the teacher screening test have earnings during the first three years of tenure that substantially greater than earnings of applicants below the test passing cutoff. Part of the premium stems from the fact that 25-50 percent of public sector teachers hold a second non-teaching job in the formal sector. For less than one-third of teachers this outside job is education-related.

Relative to applicants below the cutoff, public teachers work harder, as measured by payroll days/year in the first three years of potential teacher tenure. In the fourth year, public sector teachers at the margin have a total earnings penalty, likely due to the fact that a tax reform passed between 2012 and 2013 incentivized many formal firms to offer permanent positions to workers on temporary contracts. As a result, the labor supply for those below the cutoff increased substantially in 2013 and, relative to these, public sector teachers at the margin worked less payroll days/year.

The total annual earnings premium in the main occupation in the first three years of potential tenure for public sector teachers in the main occupation is – at the margin – a combination of a labor supply and a wage effect. The total earnings premium from all occupations for public sector teachers at the margin is, however, a labor supply effect. For 2013, in particular, there is a large positive wage (i.e. daily earnings) effect for teachers, yet the negative labor supply effect yields a negative total earnings premium.

We cannot conclude whether the sizeable annual earnings premium in the first three years of potential tenure and greater daily wages for the marginal novice public

sector teacher results in a better pool of public sector teachers. But ancillary evidence suggests that may not. Applicants who have higher college admission exam scores and who come from more socio-economically advantageous backgrounds choose not to take the test and these are likely the highest-quality teachers.

Also, despite having a probationary period, novice teachers in Colombia are rarely dismissed and their public sector teacher pay is untied to performance. Therefore, it seems likely that, on the margin, novice public teachers are devoting more effort to their outside, less-secure non-teaching job, possibly at the expense of teaching.

Finally, evidence from other developing countries suggests that substantial teacher pay increases and reductions do not affect productivity. This evidence underscores the possibility of misallocation of pay and productivity among public sector teachers.

References

- Angrist, J. D., & Guryan, J. (2007). Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review*.
- Auguste, B. G., P. Kihn, and M. Miller (2010). Closing the talent gap: Attracting and retaining top-third graduates to careers in teaching: An international and market research based perspective.
- Das, J., and Bau, N. (2016). The misallocation of pay and productivity in the public sector: Evidence from the labor market for teachers. Unpublished manuscript.
- Bruns, B., and Luque J. 2014. Great teachers: How to raise student learning in Latin America and the Caribbean. Washington D.C: World Bank.
- Dal Bó, E., Finan, F., & Rossi, M. A. (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, 128(3), 1169-1218.
- De Ree, J., Muralidharan, K., Pradhan, M., & Rogers, H. (2015). Double for nothing? Experimental evidence on the impact of an unconditional teacher salary increase on student performance in Indonesia (No. w21806). National Bureau of Economic Research.
- Estrada, R. (2013). Rules Rather than Discretion. Teacher Hiring and Rent Extraction. Working Paper.

- Ferraz, C., & Finan, F. (2009). Motivating politicians: The impacts of monetary incentives on quality and performance (No. w14906). National Bureau of Economic Research.
- Garcia, S., Maldonado, D., Perry, G., Rodriguez, C and Saavedra, J.E. 2014. *Tras la excelencia docente: Como mejorar la calidad de la educación para todos los colombianos*. Bogotá, D.C.: Fundación Empresa Privada Compartir.
- Guarino, C. M., Santibanez, L., & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of educational research*, 76(2), 173-208.
- Hanushek, E. A., & Pace, R. R. (1995). Who chooses to teach (and why)?. *Economics of Education Review*, 14(2), 101-117.
- Hernani-Limarino, W.H. (2005). Are Teachers Well Paid in Latin America and the Caribbean?" Vegas, E. (ed.), *Incentives to Improve Teaching. Lessons from Latin America*, Directions in Development, World Bank. Ed., 3, 63: 102.
- Holmstrom, B. & Milgrom, P. (1991). Multi-task principal agent analyses: Incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organization*, January: 24-52.
- Holmstrom, B. (1999). Managerial incentive problems: A principal-agent perspective. *Review of Economic Studies*, 1: 169-182.
- Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. NBER Working Paper.

- Kugler, A., Kugler, M. and Herrera, L.O. 2016. The effects of the 2013 tax reform. Mimeo, Georgetown University.
- Liang, X. (1999). "Teacher Pay in 12 Latin American Countries: How Does Teacher Pay Compare to Other Professions, What Determines Teacher Pay, and Who Are the Teachers?" Latin America and the Caribbean Region Human Development Department Paper 49. World Bank, Washington, D.C.
- Mizala, A., & Ñopo, H. (2016). Evolution of Teachers' Salaries in Latin America at the Turn of the 20th Century: How Much Are They (Under or Over) Paid? *International Journal of Educational Development*.
- OECD (2009). "Evaluating and Rewarding the Quality of Teachers International Practices".
- Psacharopoulos, G., Valenzuela, J. and Arends, G. (1996). "Teacher Salaries in Latin America: A Review." *Economics of Education Review* 15(4), 401-6.
- Rivkin, S.G., E.A. Hanushek, and J.F. Kain. 2005. Teachers, schools, and academic achievement. *Econometrica* 73 (2):417-458.
- Rockoff, J. (2004). The impact of individual teachers on student achievement: evidence from panel data. *American Economic Review*, 94(2004): 247-252.
- Rockoff, J., Jacob, B., Kane, T. and Staiger, D. (2011). Can You Recognize an Effective Teacher When You Recruit One? *Education Finance and Policy*, 6 (1): 43-74.

Saavedra, JE., Maldonado, D., Santibanez, L. and Herrera-Prada LH. (2016). Labor markets for novice teachers in Latin America: Evidence from Colombia. Unpublished report.

Table 1. Descriptive statistics and “discouragement” effect of the test

	Never	Took Exam		F_{test}
	took the exam	once	times > once	
Female	0.62 (0.49)	0.66 (0.47)	0.68 (0.46)	53.36 0.000
Household income above median	0.53 (0.50)	0.43 (0.49)	0.43 (0.50)	147.95 0.000
Score in Secondary graduation test	65.68 (26.65)	64.39 (25.19)	60.09 (26.08)	112.14 0.000
Graduated from a Public HEI	0.65 (0.48)	0.75 (0.43)	0.75 (0.43)	215.28 0.000
Mother with college degree	0.12 (0.32)	0.08 (0.27)	0.07 (0.26)	61.30 0.000
Graduated received financial aid	0.12 (0.32)	0.13 (0.34)	0.12 (0.33)	7.89 0.000
Graduated on time	0.26 (0.44)	0.22 (0.41)	0.39 (0.49)	482.50 0.000
Graduated from a certified HEI	0.16 (0.37)	0.10 (0.30)	0.12 (0.33)	113.96 0.000
College Reputation	68.51 (12.02)	68.73 (11.29)	64.98 (11.39)	422.79 0.000
Year of Graduation	2007 (1.58)	2008 (1.46)	2006 (1.71)	2125.76 0.000
Observations	12,184	18,301	11,218	

Notes: Table shows mean characteristics (and standard deviation in parentheses) of education graduates 2004-2009 who did not take the 2009 teacher screening test and of those who took the test, whether once or more than once. The sample is 41,703 university education graduates 2004-2009. Graduate characteristics shown in the table come from the SPADIES dataset. Public University and certified HEIs are characteristics taken from a Ministry’s directory. Graduation on time is a dummy variable that takes the value of 1 if the graduate was 10 semesters or less to get the degree.

Table 2. Match rates of education graduates that presented the 2009 teacher-screening exam to teacher payroll and social security records by year

Year	Matched to Social Security Records	Matched to Public Payroll (Permanent)	Matched to Teacher Data	Matched to Public Teacher Payroll Data (Temporary)	Matched to Social Security or Teacher Payroll Data	Matched to Social Security and Teacher Payroll Data
2008	97,9%	3,9%		9,6%	98,8%	3,1%
2009	90,9%	5,1%		12,3%	93,7%	2,4%
2010	84,4%	17,2%		9,8%	91,4%	10,3%
2011	77,9%	22,0%		10,0%	91,7%	8,3%
2012	75,2%	26,4%		8,9%	93,1%	8,5%
2013	73,3%	26,2%		10,9%	92,1%	7,4%

Notes: Sample of applicants includes 24,836 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test. Rows do not sum up to 100 percent.

Table 3. Total annual earnings, main and all occupations

	Annual Earnings (USD '000)			
	2010	2011	2012	2013
	<i>A. Main occupation</i>			
Public sector teacher	2.248 (0.511)	3.098 (0.312)	3.607 (0.231)	-2.299 (0.484)
	<i>B. All occupations</i>			
Public sector teacher	3.710 (0.582)	3.321 (0.346)	3.476 (0.260)	-3.030 (0.542)
Observations	24,792	24,792	24,792	24,792
Controls	Yes	Yes	Yes	Yes

Notes: Table shows fuzzy RD results for total annual earnings (in USD 2013 thousands) as a function of applicant's score in the 2009 teacher screening test. Sample of applicants includes 24,791 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test and for whom we have complete covariate information. Scores are normalized to be zero at the 60-point cutoff. Being a public sector teacher is instrumented with a dummy for being above the cutoff point. See Figure 1 for first-stage results. Covariates in the regressions not shown in the table are normalized score, normalized score squared, an interaction of normalized score and normalized score squared with above cutoff dummy, female, whether graduated from a public institution, whether received financial aid in college, whether graduated on time, whether graduated from certified institution, college reputation and year of graduation and a constant. Robust standard errors in parentheses.

Table 4. Labor supply, main and all occupations

	Payroll Days/Year			
	2010	2011	2012	2013
	<i>A. Main occupation</i>			
Public sector teacher	61.0 (15.8)	99.0 (9.6)	111.7 (6.9)	-103.1 (13.6)
	<i>B. All occupations</i>			
Public sector teacher	120.5 (18.8)	110.3 (11.1)	114.5 (7.9)	-110.7 (15.8)
Observations	24,792	24,792	24,792	24,792
Controls	Yes	Yes	Yes	Yes

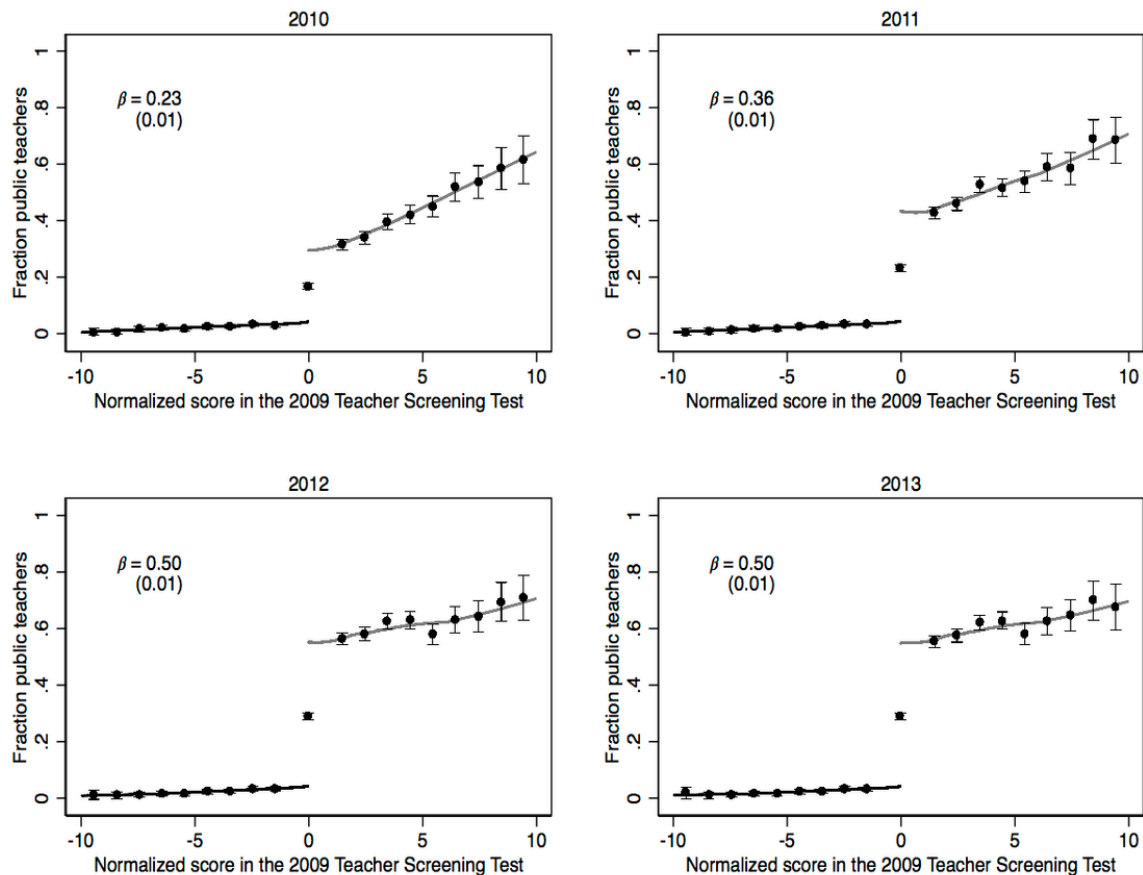
Notes: Table shows fuzzy RD results for labor supply (payroll days per year) as a function of applicant's score in the 2009 teacher screening test. Sample of applicants includes 24,791 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test and for whom we have complete covariate information. Scores are normalized to be zero at the 60-point cutoff. Being a public sector teacher is instrumented with a dummy for being above the cutoff point. See Figure 1 for first-stage results. Covariates in the regressions not shown in the table are normalized score, normalized score squared, an interaction of normalized score and normalized score squared with above cutoff dummy, female, whether graduated from a public institution, whether received financial aid in college, whether graduated on time, whether graduated from certified institution, college reputation and year of graduation and a constant. Robust standard errors in parentheses.

Table 5. Daily earnings, main and all occupations

	Daily Earnings (USD)			
	2010	2011	2012	2013
	<i>A. Main occupation</i>			
Public sector teacher	2.753 (1.455)	4.738 (0.987)	5.990 (0.698)	6.404 (0.752)
	<i>B. All occupations</i>			
Public sector teacher	2.268 (1.430)	4.360 (0.938)	5.528 (0.684)	5.255 (0.738)
Observations	24,792	24,792	24,792	24,792
Controls	Yes	Yes	Yes	Yes

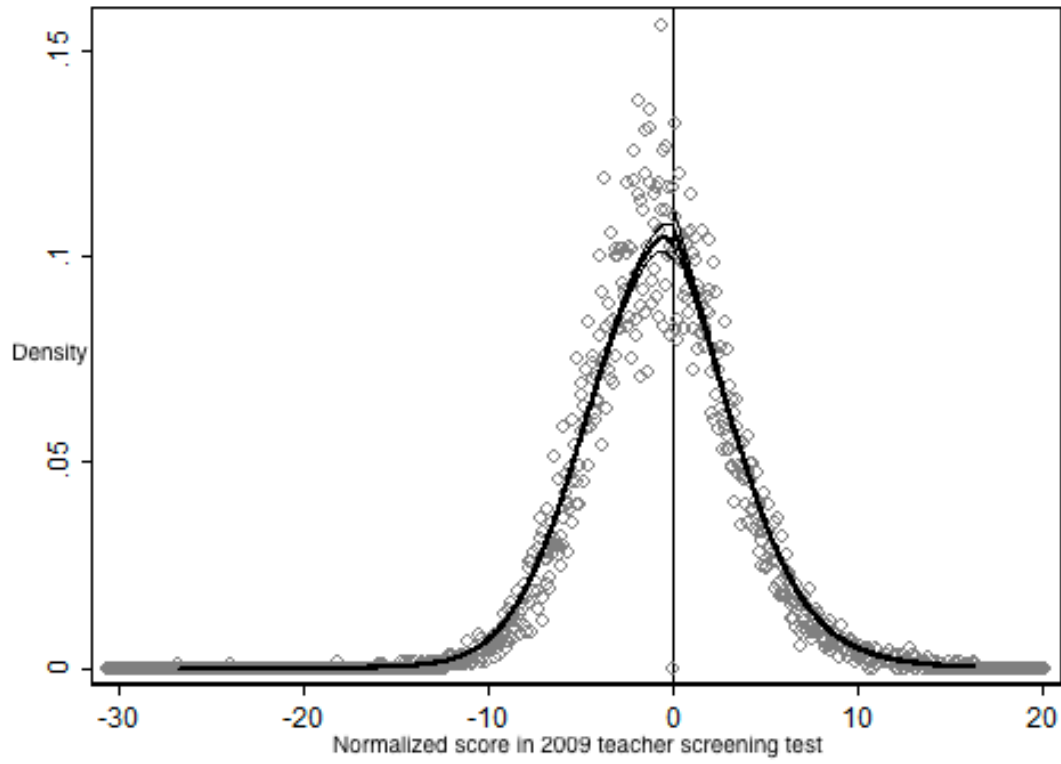
Notes: Table shows fuzzy RD results for daily earnings (in USD 2013) as a function of applicant's score in the 2009 teacher screening test. Daily earnings are annual earnings divided by payroll days per year. Sample of applicants includes 24,791 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test and for whom we have complete covariate information. Scores are normalized to be zero at the 60-point cutoff. Being a public sector teacher is instrumented with a dummy for being above the cutoff point. See Figure 1 for first-stage results. Covariates in the regressions not shown in the table are normalized score, normalized score squared, an interaction of normalized score and normalized score squared with above cutoff dummy, female, whether graduated from a public institution, whether received financial aid in college, whether graduated on time, whether graduated from certified institution, college reputation and year of graduation and a constant. Robust standard errors in parentheses.

Figure 1. First stage: probability of working as a public sector teacher as a function of 2009 teacher screening test scores



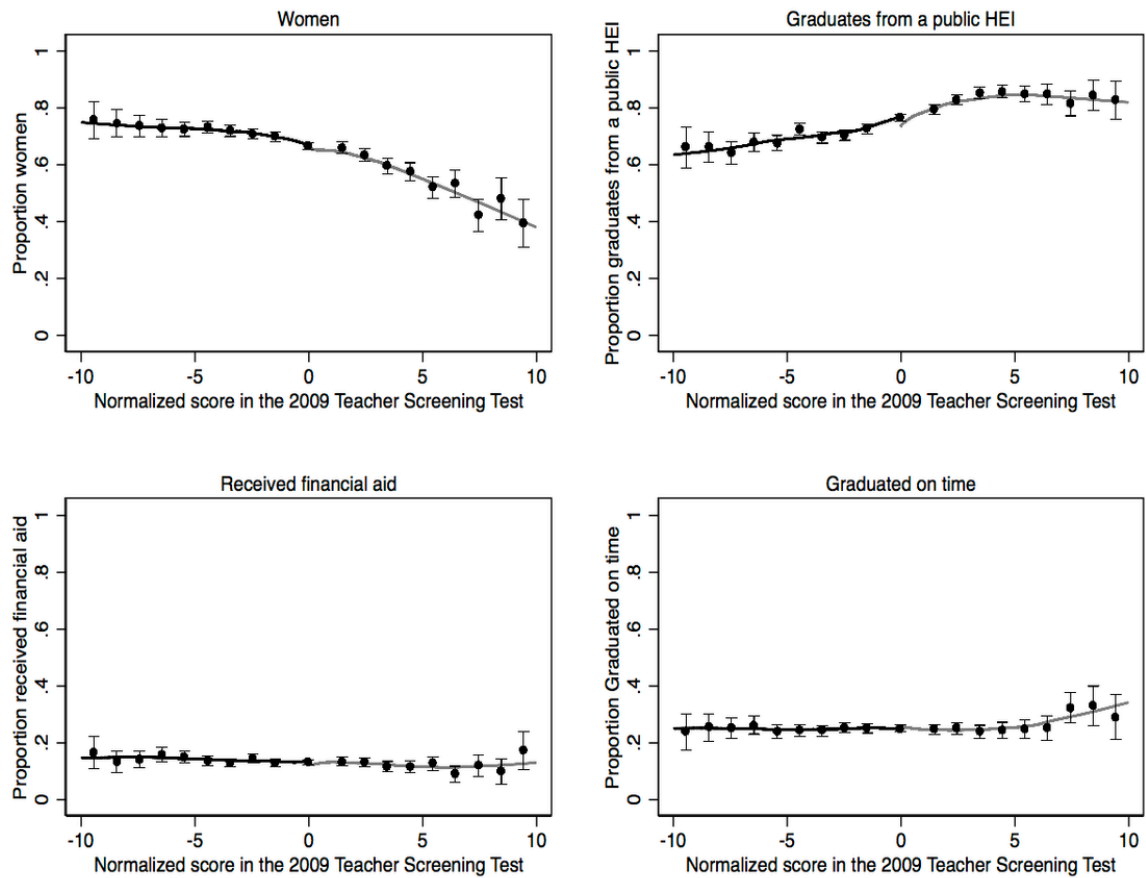
Notes: Figure shows the probability of working as a public sector teacher in 2010, 2011, 2012 and 2013 as a function of applicant's score in the 2009 teacher screening test. Sample of applicants includes 24,836 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test. Scores are normalized to be zero at the 60-point cutoff. Each dot is the conditional probability of public sector employment for a given integer score. Whiskers are 95% confidence intervals. Separate fits at each side of the cutoff are non-parametric (lowess) regressions. Coefficient and standard error in each panel is from a linear regression of the probability of being a public sector teacher on normalized score, normalized score squared, an interaction of normalized score and normalized score squared with above cutoff dummy and control variables in Table 1 with robust standard errors.

Figure 2. Density of scores in the 2009 teacher-screening test



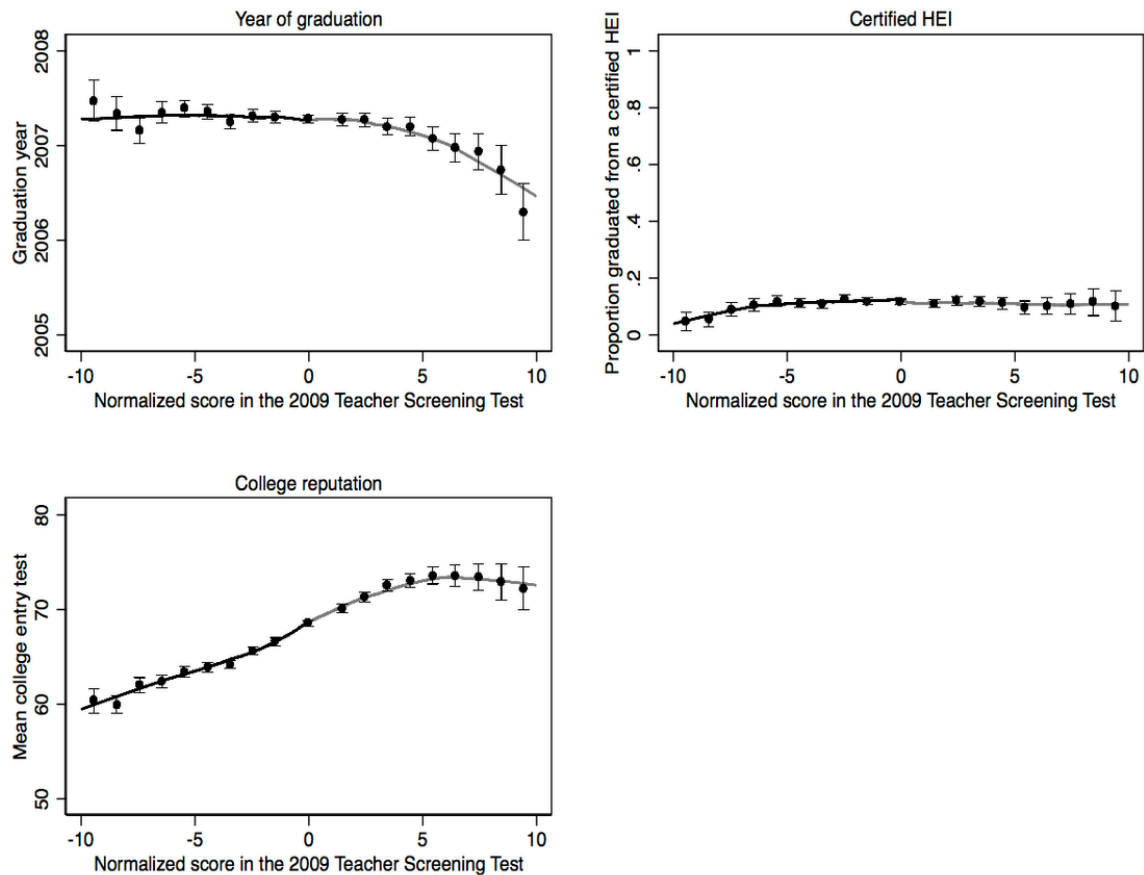
Notes: Sample includes 24,836 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test. Scores are normalized to be zero at the 60-point cutoff.

Figure 3a. Covariate continuity (2009 teacher screening test)



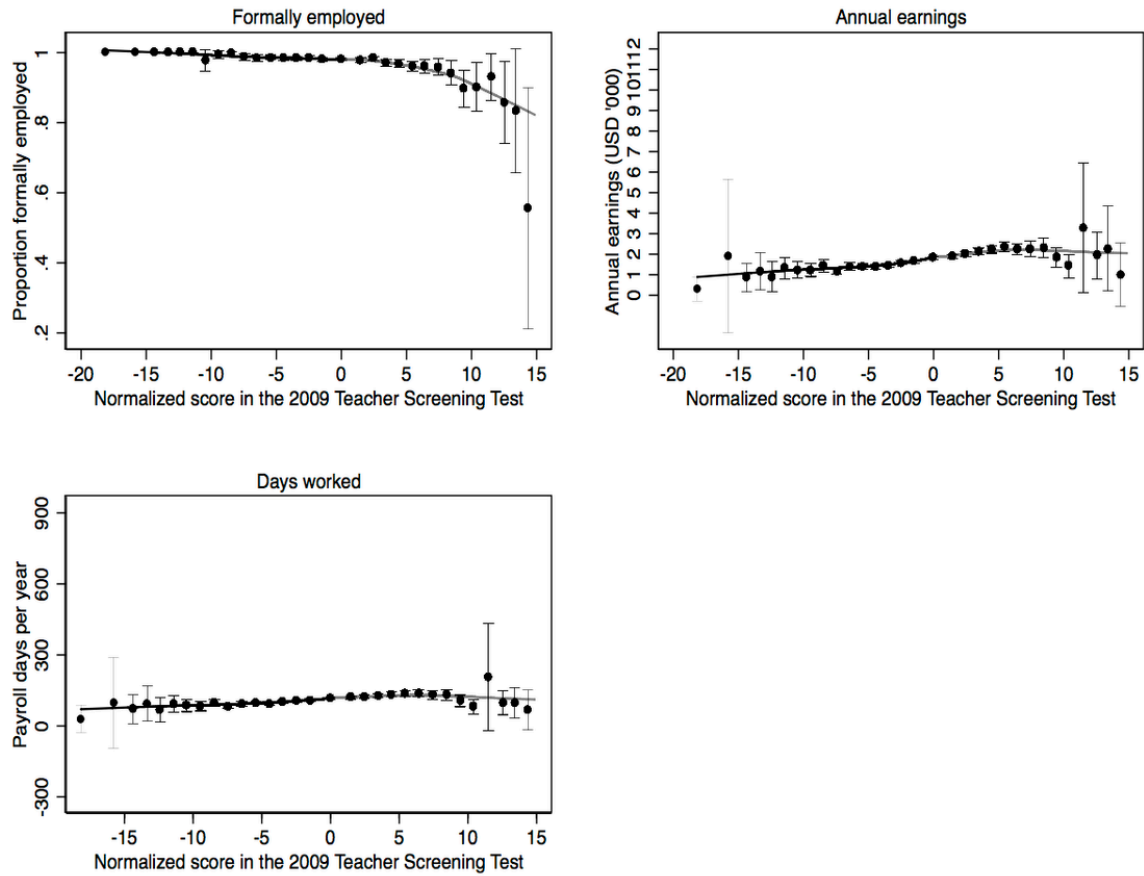
Notes: Figure shows covariates as a function of applicant's score in the 2009 teacher screening test. Covariates are from SPADIES data. Sample of applicants includes 24,791 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test. Scores are normalized to be zero at the 60-point cutoff. Each dot is the conditional probability of public sector employment for a given integer score. Whiskers are 95% confidence intervals. Separate fits at each side of the cutoff are non-parametric (lowess) regressions.

Figure 3b. Covariate continuity (2009 teacher screening test)



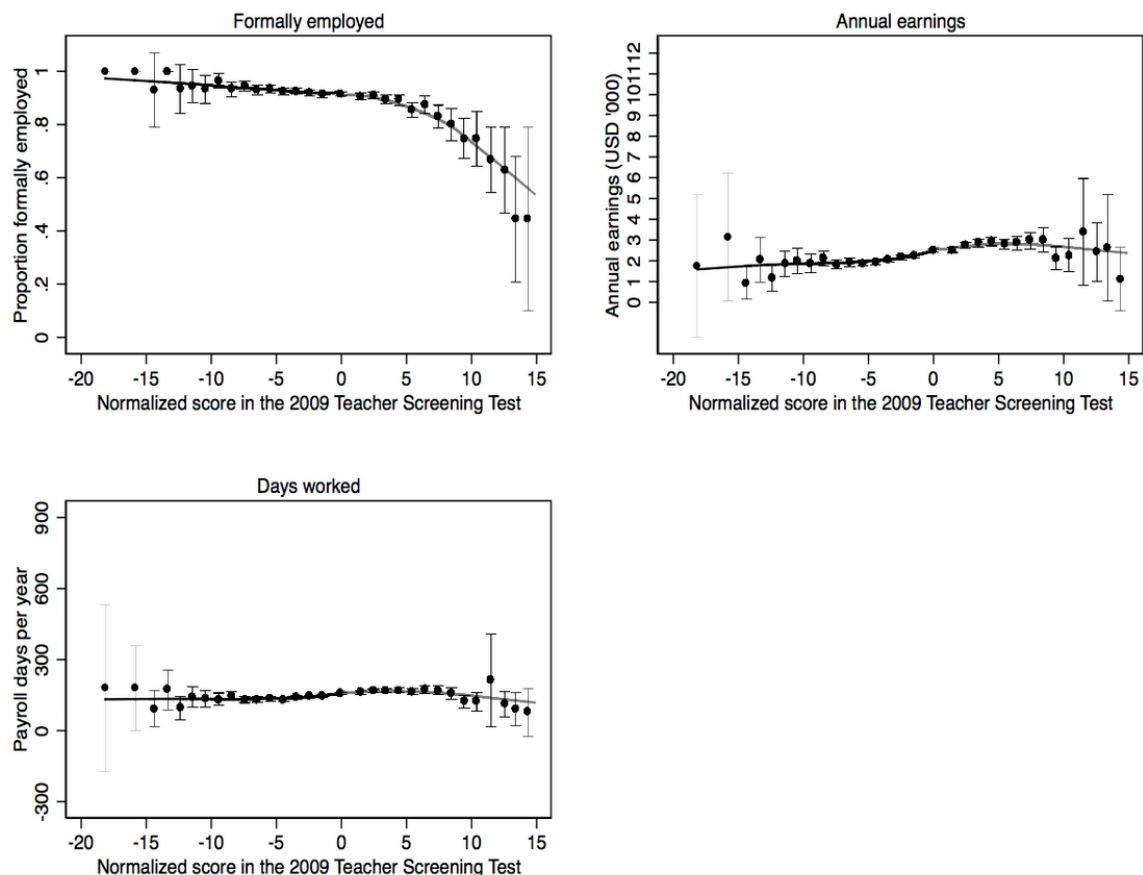
Notes: Figure shows covariates as a function of applicant's score in the 2009 teacher screening test. Covariates are from SPADIES data. Sample of applicants includes 24,791 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test. Scores are normalized to be zero at the 60-point cutoff. Each dot is the conditional probability of public sector employment for a given integer score. Whiskers are 95% confidence intervals. Separate fits at each side of the cutoff are non-parametric (lowess) regressions.

Figure 4a. Placebo test: labor market outcomes 2008 (pre-2009 screening test)



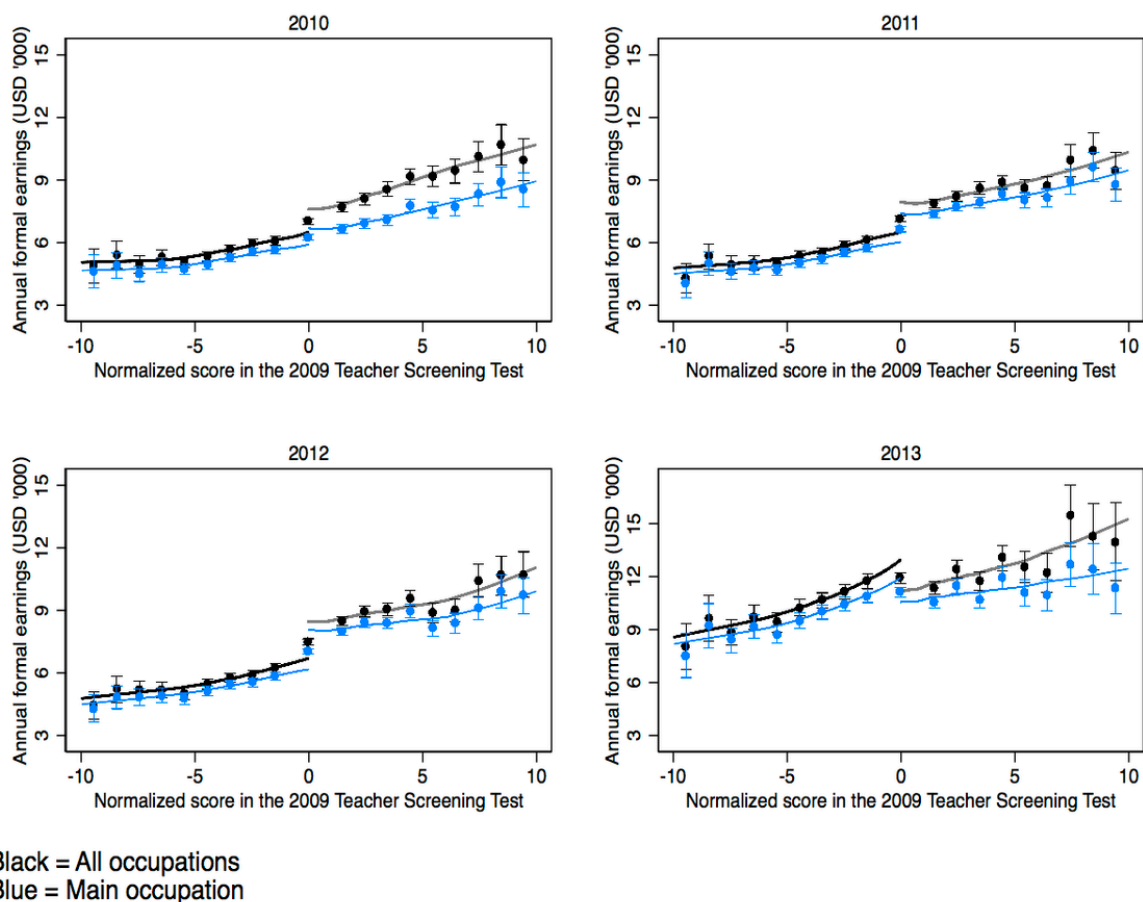
Notes: Figure shows labor market outcomes (probability of formal employment, annual formal earnings and payroll days per year) in 2008 as a function of applicant's score in the 2009 teacher-screening test. Sample of applicants includes 24,791 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test. Scores are normalized to be zero at the 60-point cutoff. Dots are conditional means. Whiskers are 95% confidence intervals. Separate fits at each side of the cutoff are non-parametric (lowess) regressions.

Figure 4b. Placebo test: labor market outcomes 2009 (year of 2009 screening test)



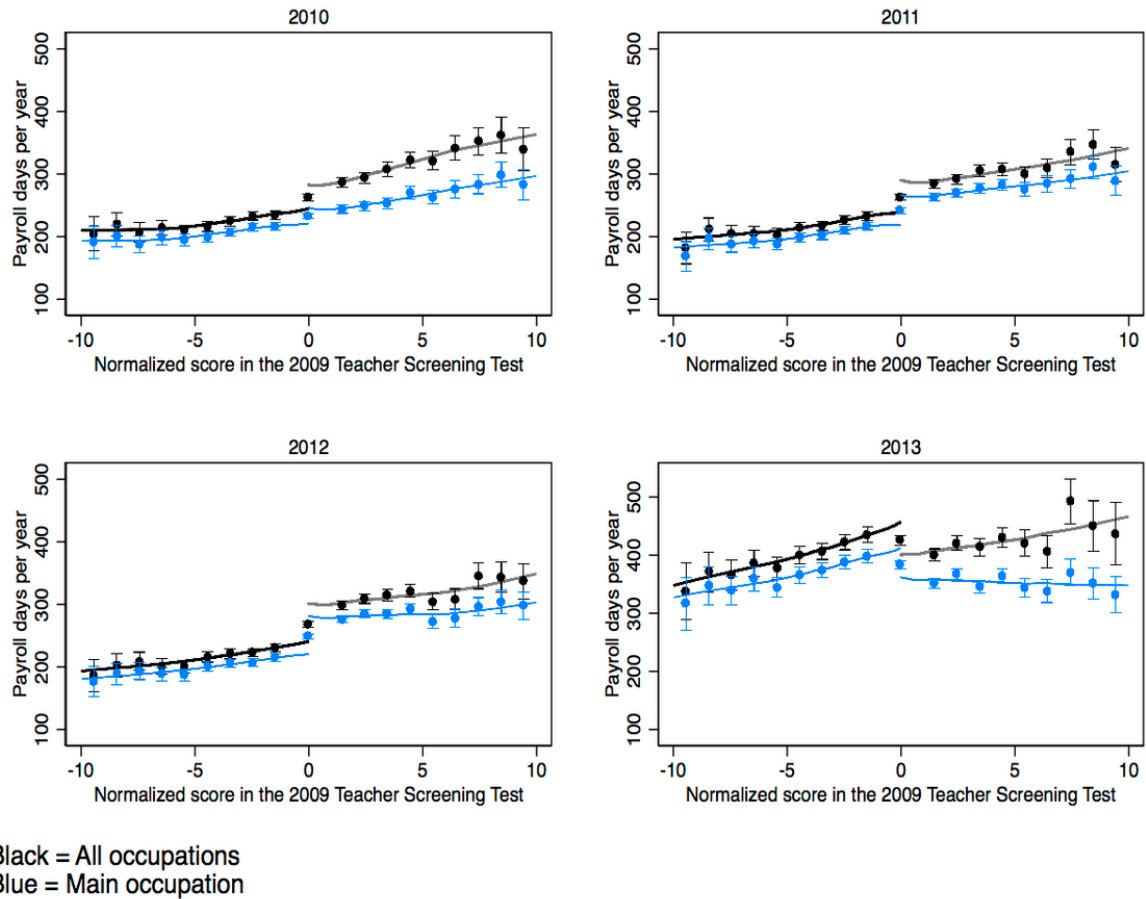
Notes: Figure shows labor market outcomes (probability of formal employment, annual formal earnings and payroll days per year) in 2009 as a function of applicant’s score in the 2009 teacher-screening test. Sample of applicants includes 24,791 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test. Scores are normalized to be zero at the 60-point cutoff. Dots are conditional means. Whiskers are 95% confidence intervals. Separate fits at each side of the cutoff are non-parametric (lowess) regressions.

Figure 5. Annual earnings (USD '000)



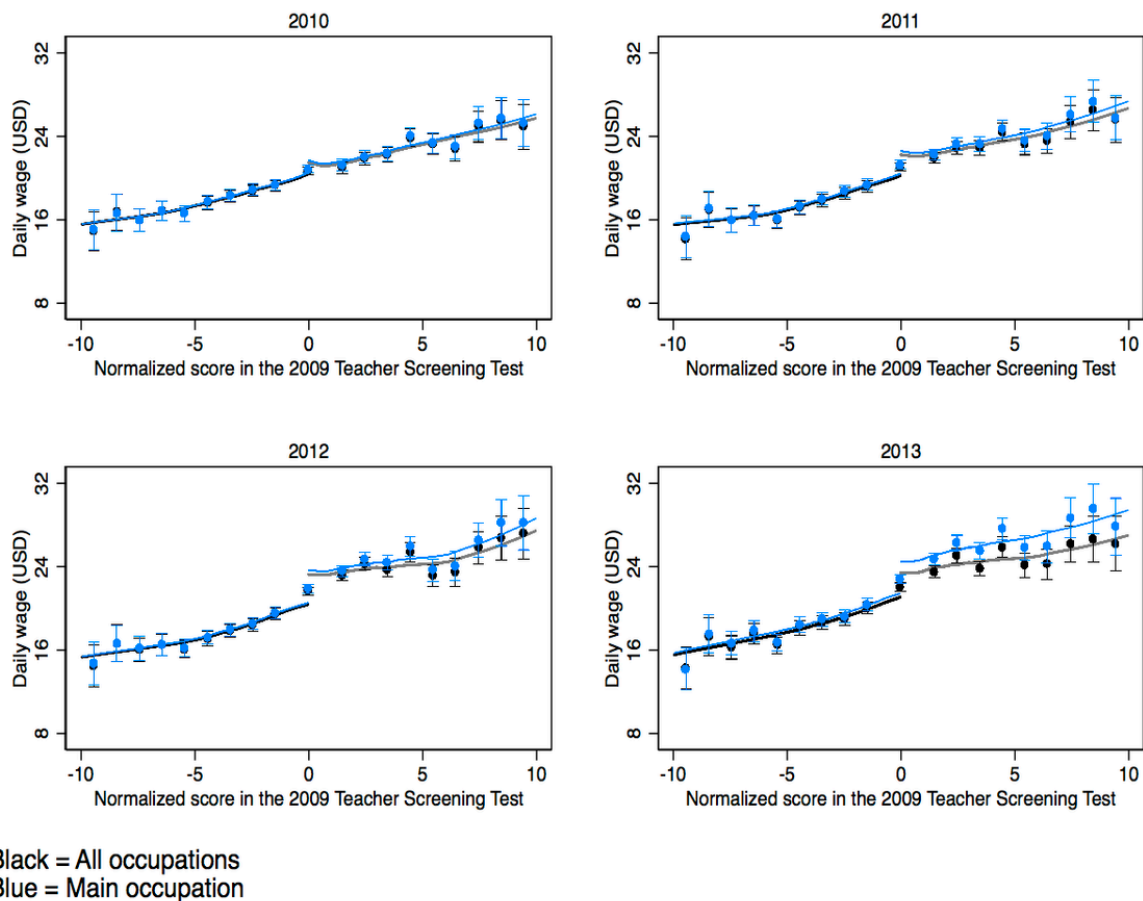
Notes: Figure shows total annual earnings (in 2013 US dollars) as a function of applicant's score in the 2009 teacher-screening test. Black labels are total annual earnings from all occupations, including teacher payroll and other formal occupations. Blue labels are annual earnings from the main occupation, which is public sector teaching for those in the public teacher payroll and the occupation with the greatest number of payroll days per year for those who are not in the teacher payroll, including zeros. Sample of applicants includes 24,791 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test. Scores are normalized to be zero at the 60-point cutoff. Each dot is the conditional mean of total earnings for a given integer score. Whiskers are 95% confidence intervals. Separate fits at each side of the cutoff are non-parametric (lowess) regressions.

Figure 6. Labor supply



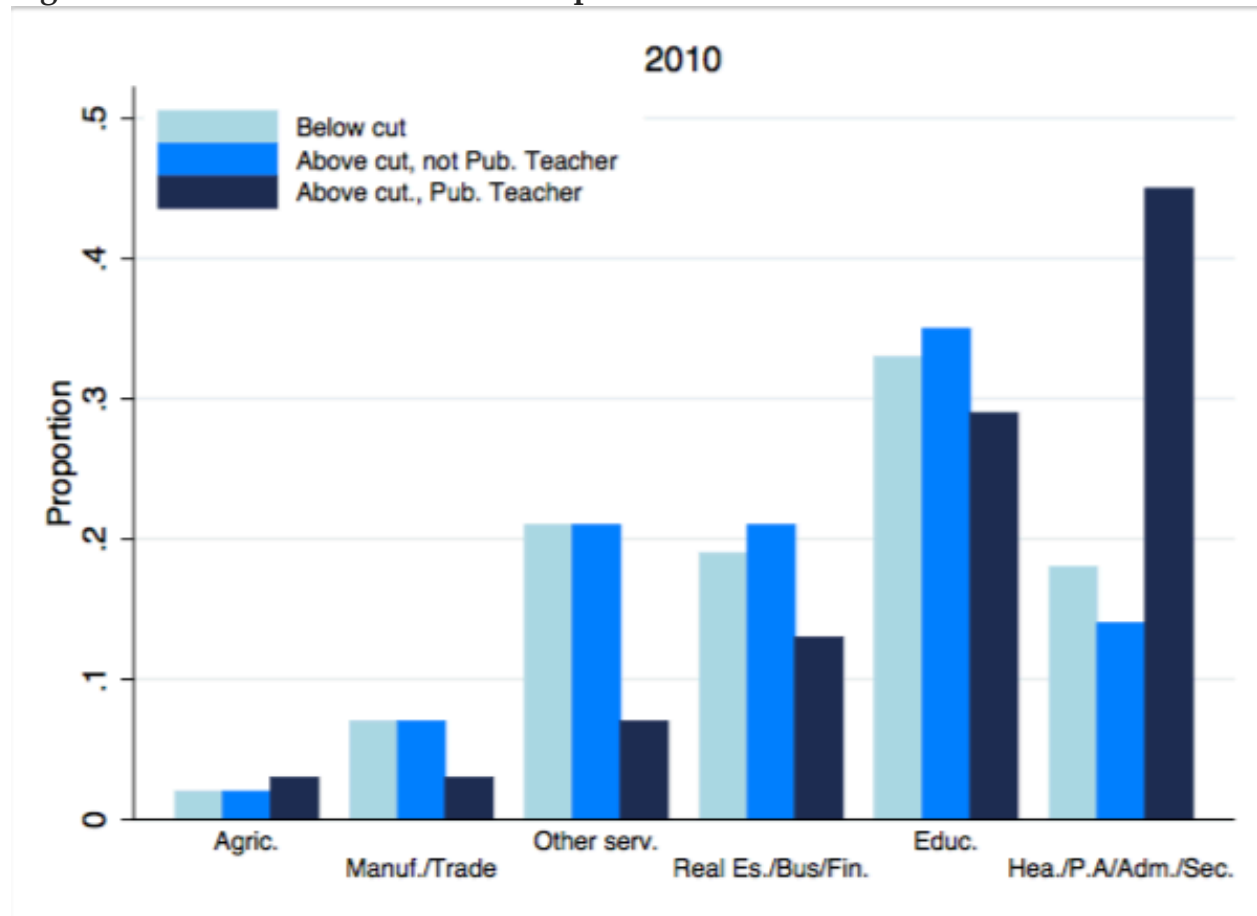
Notes: Figure shows labor supply, as measured by number of payroll days per year, for all occupations (black) and main occupation (blue). The main occupation is public sector teaching for those in the public teacher payroll and the occupation with the greatest number of payroll days per year for those who are not in the teacher payroll, including zeros. Sample of applicants includes 24,791 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test. Scores are normalized to be zero at the 60-point cutoff. Each dot is the conditional mean of earnings for a given integer score. Whiskers are 95% confidence intervals. Separate fits at each side of the cutoff are non-parametric (lowess) regressions.

Figure 7. Daily wages (USD)



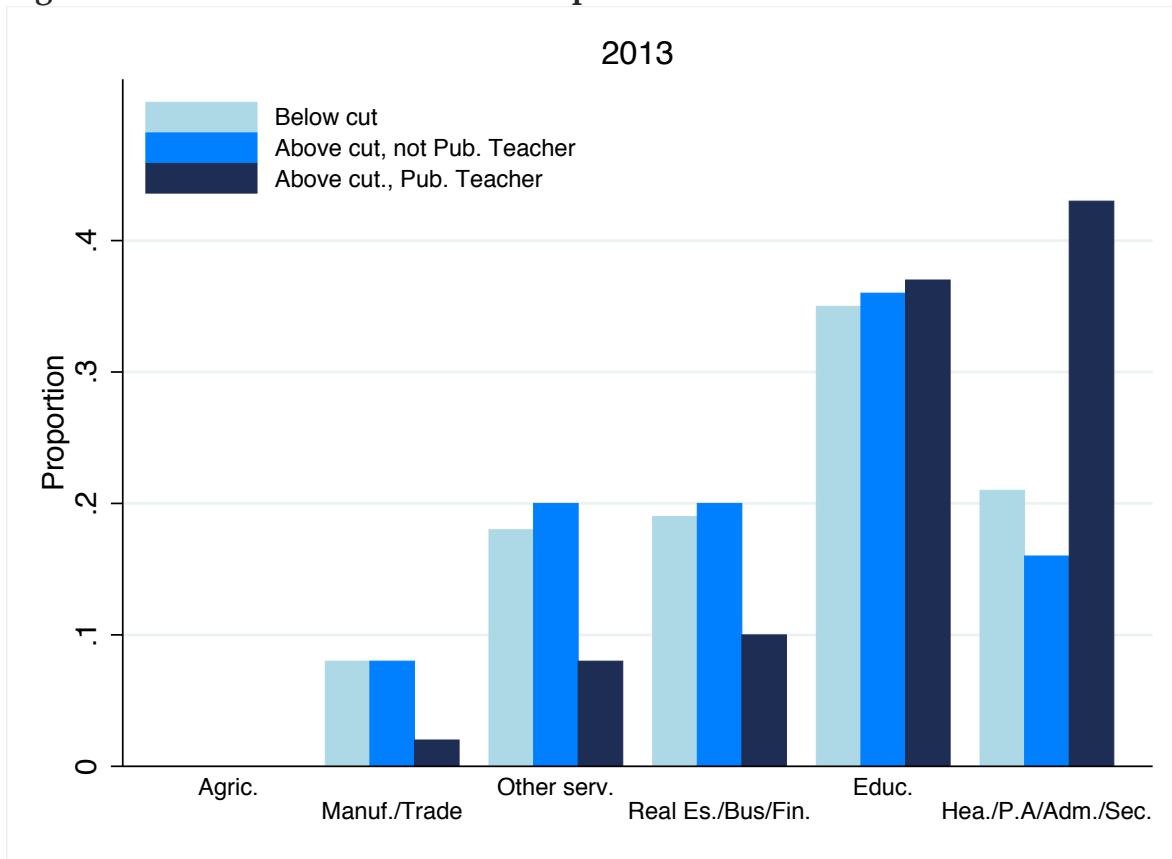
Notes: Figure shows average daily earnings in the year as a function of applicant's score in the 2009 teacher-screening test. Daily earnings for all occupations (black) are total annual earnings divided by total days worked from any source, including zeros. Daily earnings for main occupation (blue) are total earnings from main occupation divided by total payroll days in the main occupation. Sample of applicants includes 24,791 education graduates from graduating cohorts 2004-2009 who take the 2009 screening test. Scores are normalized to be zero at the 60-point cutoff. Each dot is the conditional mean of number of days worked for a given integer score. Whiskers are 95% confidence intervals. Separate fits at each side of the cutoff are non-parametric (lowess) regressions.

Figure 8a. Distribution of outside occupations in 2010



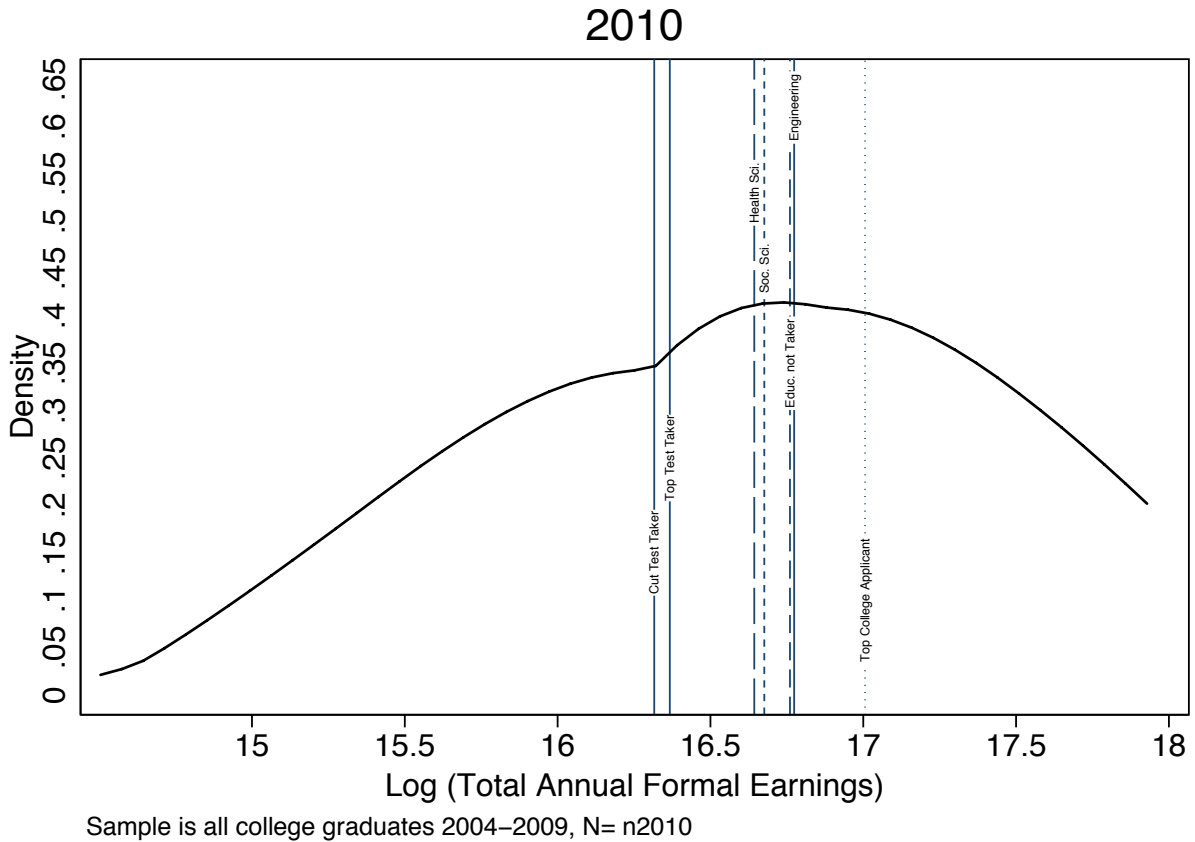
Notes: Figure shows the occupational distribution of outside jobs in 2010 for various groups of 2009 teacher screening test-takers. Occupational codes are aggregated to one digit, with the exception of education related codes.

Figure 8b. Distribution of outside occupations in 2013



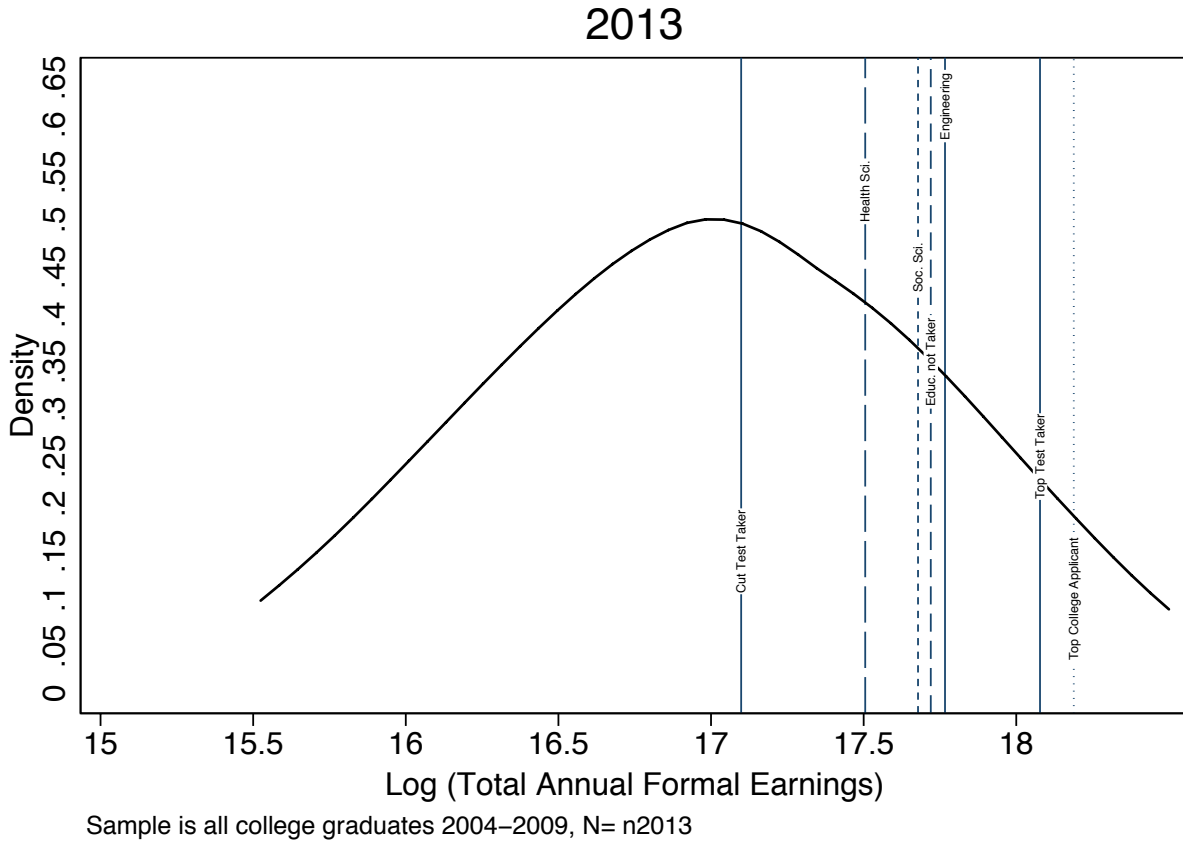
Notes: Figure shows the occupational distribution of outside jobs in 2013 for various groups of 2009 teacher screening test-takers. Occupational codes are aggregated to one digit, with the exception of education related codes.

Figure 9a. LATE and the earnings distributions of all college graduates in 2010



Notes: Figure shows the (log) distribution of total annual earnings in 2010 of all college graduates 2004-2009, and different relevant centiles, including those at the 60 point cutoff for eligibility for a public position, the 95th percentile of the 2009 teacher screening test score distribution, median for education graduates who do not take the screening test, medians for health science, social science and engineer graduates and the 95th percentile of college entry exam scores.

Figure 9b. LATE and the earnings distributions of all college graduates in 2013



Notes: Figure shows the (log) distribution of total annual earnings in 2013 of all college graduates 2004-2009, and different relevant centiles, including those at the 60 point cutoff for eligibility for a public position, the 95th percentile of the 2009 teacher screening test score distribution, median for education graduates who do not take the screening test, medians for health science, social science and engineer graduates and the 95th percentile of college entry exam scores.