Recruiting Better Teachers? Evidence from a Higher Education Reform in Chile

Adriano De Falco
University of Bologna

Benjamin Hattemer
European University Institute

Sofía Sierra Vásquez
University of St. Gallen

November 2025

Abstract

This paper analyzes the impact of a recruitment policy in Chile designed to improve the quality of new teachers by incentivizing high-achieving and restricting low-achieving high school graduates from entering the teaching profession. We document that the reform effectively improved the average test scores of new teachers. Using a teacher value-added (TVA) model, we find that the reform increased TVA for mathematics but not for Spanish teachers. Finally, we show that most of the effect cannot be explained by new teachers' higher average test scores, but rather can be attributed to beneficial but unintended effects of the reform.

Keywords: *educational policy, scholarship, teacher recruitment, teacher value-added*

JEL: I21, I22, J24, H52

We are grateful for the valuable comments from Sule Alan, Caterina Calsamiglia, Thomas Crossley, Andrea Ichino, Alexander Monge-Naranjo, Steve Pischke, Camille Terrier, and Michela Tincani as well as participants of the 2023 CESifo/ifo Junior Workshop on the Economics of Education, the 8th LEER, EALE 2023, EAYE 2024 and SAEe 2023 conferences, AIEL - Padua Workshop on the Economics of Human Capital 2024, 6th QMUL Economics and Finance Workshop for PhD & Post-doctoral Students, the BSE Summer Forum Workshop on Children's Health, Well-being, and Human Capital Formation, the XVI Labour Economics Meeting, the 39th AIEL conference, University of Bologna, the SEW Quantitative Seminar (University of St. Gallen), and the SKILS (Laax) Workshop. We thank the Departamento de Evaluación, Medición y Registro Educacional (DEMRE) for providing the databases of the Higher Education Admission System for the development of this research. We thank the Agencia de Calidad de la Educación for the primary school student-level data. All results of this study are the sole responsibility of the authors and do not represent either institution.

1 Introduction

Teacher quality plays a critical role in improving student learning and long-term outcomes (Bau and Das, 2020; Chetty et al., 2014b; Rockoff, 2004). Thus, how to design policies to improve teacher quality is a long-standing question. Among the set of interventions that could be implemented to improve the effectiveness of teachers, recruitment policies designed to attract more talented and motivated individuals into the teaching profession have become increasingly relevant. This is especially true in the context of low- and middle-income countries where teachers often lack the skills or motivation to teach effectively (World Bank, 2017).

There are several potential challenges in implementing recruitment policies for any occupation. First, there is the problem of designing effective incentives to attract people into a particular field. The second challenge lies in the availability of an observable and reliable measure to *ex-ante* determine a person's potential to be good at their job or, in other words, how talented they will be at it. Finally, there is the concern about individuals being selected on other unobservable qualities. Depending on the incentives in place, selection on unobservables could either hinder or reinforce the effectiveness of the recruitment policy. For example, those attracted by the policy may show a lower level of intrinsic motivation, generating a possible trade-off with talent. This trade-off has been widely studied in the literature and is particularly salient for civil servants, such as educators (Ashraf et al., 2020; Deserranno, 2019; Leaver et al., 2021). Conversely, the policy may attract individuals who would not have become teachers because their unobserved qualities are highly rewarded in other labor markets, even though those same qualities also make them good teachers (Bacolod, 2007).

In this paper, we analyze a policy designed to improve the average quality of new teachers by targeting the first stage of their recruitment process. The reform, implemented in Chile in 2011, was designed to reshape the academic ability composition of high-school graduates enrolling in undergraduate teacher training programs. To do so, it incentivized the enrollment of high-achieving prospective college students by creating a scholarship called *Beca Vocación de Profesor* (BVP). The scholarship imposed requirements on both college students and higher education institutions: on the one hand, individuals who received the scholarship were required to work as teachers in public schools for three of the seven years following college graduation; and on the other hand, universities were required to limit the number of low-achieving college students admitted to their programs so that high-achieving ones would be eligible for the BVP.[1] The screening device used was the national standardized university entrance exam, known as *Prueba de Selección Universitaria* (PSU). It defined low-achieving students as those who scored below the mean, and high-achieving students as those who scored one standard deviation above the mean. The implicit assumption behind this reform was that high PSU scorers would be better teachers than low scorers.

Our empirical assessment of this policy answers three main questions related to the challenges outlined above: First, was the policy effective in (permanently) attracting high-achieving high-

[1] Note that this was not a complete ban on low-achieving college students, as they could still enroll in these restricted slots at BVP-eligible universities, or enroll at non-BVP-eligible institutions.

school graduates to the teaching profession? Second, did this translate into an increase in average teaching quality? And finally, did the reform have any effect beyond selection based on academic ability? The analysis is divided into three parts.

In the first part, we study the effectiveness of the reform in recruiting higher PSU scorers. We leverage administrative data that covers the entire population of individuals who registered for the PSU between 2009 and 2014. We construct a unique panel dataset that follows them during their higher education years and into their subsequent teaching careers. The sharp cutoffs used by the reform to define high and low-achieving college students allow us to estimate the local causal effects of its two features in a Regression Discontinuity (RD) design framework.

We find that the reform was effective in reshaping the composition of enrollees in teacher training programs. The scholarship encouraged high-achieving college students to pursue a teaching career, increasing their likelihood of enrolling in a teacher training program by 27% at the cutoff, and it discouraged low-achieving ones from enrolling, resulting in a 36% decrease in the probability of enrollment at the cutoff. Tracking individuals through higher education and into the labor market, we show that these effects translated into a smaller yet economically significant increase in the probability of graduating and working as a teacher up to five years after graduation. Finally, and consistent with the requirements and incentives introduced by the BVP, we document that incentivized high-achieving teachers are more likely to work in publicly funded schools and that these effects persist even beyond the scholarship requirements. We furthermore provide suggestive evidence that these effects are likely not local. By comparing the enrollment decisions of PSU cohorts before and after the reform, we estimate effects that are similar to those obtained in our local analysis.

In the second part of the paper, we rely on a Teacher Value-Added (TVA) model, which we estimate using primary school students' standardized test scores, to provide evidence of the effectiveness of the policy in terms of teacher quality. In this empirical framework, we analyze the impact of the reform on overall teacher quality and the validity of PSU scores as a predictor of pre-reform teacher quality. Our population of interest for this part of the analysis is the universe of math and Spanish elementary school teachers who began their training after 2004 and for whom we can reliably estimate TVA measures. Given these sample restrictions, we are left with nearly 9,500 teachers whom we link to the standardized test scores of the students they teach (around 430,000) as well as to characteristics of the jobs they hold and the schools in which they work.

We estimate the standard deviation of teacher value-added, controlling for a wide range of covariates, including lagged test scores. We then proceed to evaluate the effectiveness of the reform, comparing the value-added of teachers who started their training before and after 2011, while controlling for the effect of years of experience. Our results show that the reform was successful in increasing the average TVA for mathematics teachers by 0.038, which corresponds to 38% of the standard deviation of TVA. For Spanish teachers, however, we find a precise null effect. Turning to the role of PSU in explaining the results, we show that its contribution is marginal. We estimate that before the reform, a one standard deviation increase in PSU was associated with a statistically significant 0.02 increase in TVA for mathematics teachers. We find, moreover, that this correlation does not exist for Spanish teachers. Our estimates of the pre-reform correlation between PSU scores and TVA for math teachers suggest that around 10% of the overall effects can be explained by changes to the academic ability composition of teachers.

In the third part of the paper, we analyze channels other than the increase in teachers' academic ability that might explain the observed increase in teacher quality for mathematics teachers. To guide our analysis, we develop a simple framework à la Roy (1951) of occupational choice in which individuals decide whether to enter teaching based on two forces: intrinsic motivation for the profession and the attractiveness of outside career options. High value-added teachers tend to be both more motivated and more productive outside of teaching, so the balance of these forces determines which individuals are drawn in or pushed out by the policy. The model predicts that scholarships can raise average teacher quality if they succeed in attracting high-ability individuals who would otherwise be pulled into outside occupations, whereas disincentives can raise quality if they discourage the least motivated candidates from entering. For mathematics teachers, this is exactly what we find empirically. We observe a higher value-added (0.065) for those teachers with scores below the mean, i.e., subject to restricted access, as well as for those eligible for the BVP, relative to comparable teachers from previous cohorts. Consistent with the theoretical framework, we also find that the restriction raised the average intrinsic motivation of teachers with scores below the mean, as measured through teacher survey data collected by the Ministry of Education. At the same time, the average motivation of incentivized teachers who studied after the reform did not decrease. Taken together, these results show that the policy improved teacher quality through both channels and allow us to reject the concern that incentives for high-academic-ability individuals came at the cost of lower intrinsic motivation.

Related literature. We mainly contribute to two different strands of the literature. First and foremost, we contribute to the literature on teacher recruitment policies. These studies have primarily examined the short-term effects of two types of interventions: the introduction of entry barriers and financial incentives for prospective teachers. Entry barriers, often in the form of tests that assess candidates' attitudes and content knowledge, have been shown to be successful in improving teaching quality.[2] Nevertheless, they can have negative unintended effects on student learning if not properly designed (Busso et al., 2024). Regarding the role of financial incentives, recent studies show that schools that financially reward teacher performance are more likely to recruit high-performing teachers (Biasi, 2021; Leaver et al., 2021). We contribute to this literature by evaluating a unique recruitment policy with two relevant features. First, the reform combined entry barriers and financial incentives. Second, it targeted the earliest stage of the recruitment process, namely enrollment in teacher training programs. This addresses a relevant gap in the literature, as previous studies have focused on the recruitment of trained teachers - i.e., individuals who have already acquired a certain level of human capital, which pre-conditions occupational

[2]See, for example, Estrada (2019) for the case of Mexico, and Araujo et al. (2020) for the case of Ecuador.

sorting.[3] By analyzing the effects at the enrollment stage, we provide insights into an important margin; ignoring the impact on occupation sorting could lead to a significant underestimation of the long-term benefits of such interventions (De Ree et al., 2018).[4]

Recruitment policies directly targeting high school graduates require a screening device able to predict their quality as teachers. In this paper, we evaluate one such measure that is both easy to observe and use: college entry exams. Despite the general consensus on the weak correlation between observable characteristics and teaching quality (Staiger and Rockoff, 2010), college entry exams have been shown to predict it in some contexts. For example, Jacob et al. (2018), using classroom observation measures, argue that the SAT is predictive of teacher performance. More closely related to our paper, Gallegos et al. (2022) also focus on the case of Chile and analyze the adequacy of college entrance exams as a measure of teacher productivity, before the introduction of the reform. The authors use a wide range of teacher productivity measures, namely graduation, labor market outcomes, teachers' scores on college exit exams, and on-the-job teacher evaluations. They find a positive relation between college entrance test scores and their measures of teacher productivity. We contribute by focusing on value-added measures of teacher quality for different subjects, which take into account potential biases arising from student characteristics or the teachers' test-taking skills.[5] More importantly, focusing on the entire population of primary school students we directly test and quantify the effect of the 2011 reform on teacher quality. This stands in contrast to Gallegos et al. (2022), who infer the potential impact of the reform from the correlation between pre-reform admission test scores and teacher productivity. Similarly to Gallegos et al. (2022), a recent working paper by Pal (2025) also analyzes the 2011 reform, using it to estimate a structural model of higher-education admissions aimed at improving teacher recruitment and taking compositional changes based on admission test scores as the only channel at play.[6] In contrast to these two papers, and in line with the literature showing a weak correlation between observable characteristics and value added, we challenge the idea that college entry exams may serve alone as a tool to increase teaching quality. Indeed, we attribute most of the effect of the reform to unintended selection effects.

This paper also contributes to the more general literature on recruitment policies in contexts where financial incentives may attract individuals with lower intrinsic motivation or prosociality. Studies focusing on this topic have not reached a clear consensus. Focusing on health care workers, Ashraf et al. (2020) find that financial incentives attract less prosocial (or intrinsically motivated) individuals, but only among the low-talented. On the contrary, Dal Bó et al. (2013) who study this trade-off among public agents, find that higher wages attract high-ability workers who are

[3] Focusing on the Chilean setting, Behrman et al. (2016) show that this aspect plays an important role for wage policies.

[4] Recent studies estimating positive effects of financial incentives on teacher effort and student achievement include Lavy (2009); Duflo et al. (2012); Biasi (2021); Leaver et al. (2021); Tincani (2021).

[5] See Cohen and Goldhaber (2016); Steinberg and Garrett (2016); Bacher-Hicks et al. (2019); Bacher-Hicks (2022).

[6] Pal (2025) also presents reduced-form evidence of the reform's effect on teacher quality, focusing solely on changes in PSU score composition and comparing programs that accepted BVP enrollees to those that did not. This approach assumes that changes in the PSU composition are the only relevant ones. Note that our paper predates Pal (2025) and was first made publicly available in 2024 (see `https://dx.doi.org/10.2139/ssrn.4874361`).

also more motivated. Finally, Leaver et al. (2021) find that while financial incentives may attract less motivated teachers, this does not translate into worse job performance. We contribute to this literature by showing that scholarships do not attract teachers with lower intrinsic motivation or productivity levels. We show, however, that disincentives such as limiting access to training programs discourage the least motivated from pursuing a teaching career.

Road map. The rest of the paper proceeds as follows. Section 2 presents the institutional setting and the reform we focus on. The section additionally presents a theoretical framework that describes the potential effects of the reform on teacher quality and the mechanisms at play. Section 3 presents the data and results on enrollment, graduation, and employment at schools. In Section 4, we estimate the change in teacher quality due to the reform, focusing on the role of college admission test scores as a predictor of productivity. Section 5 discusses the effects of the reform beyond its selection of new teachers based on test scores. Finally, in Section 6 we conclude.

2 Background

In this section, we present some relevant features of the Chilean institutional setting in the years surrounding the reform and discuss some key aspects of it. Additionally, we present a theoretical framework that illustrates how the reform may affect teaching quality.

2.1 Chilean institutional setting

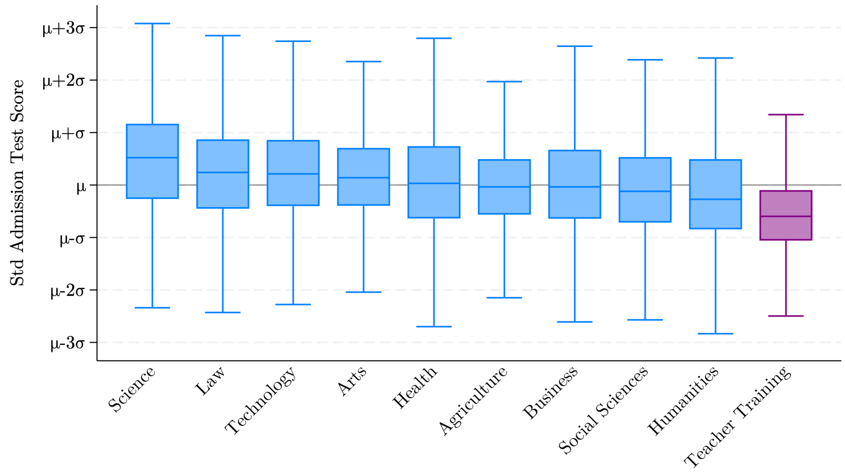
The higher education system. In Chile, there are three types of higher education institutions (HEIs): vocational institutions, technical institutes, and universities. Moreover, universities can be categorized based on whether they are part of a network called CRUCH (Consejo de Rectores de Universidades Chilenas). They are generally considered of higher prestige and are able to attract better students on average. To access most universities, students have to take a nationwide standardized test which, between 2004 and 2020, was called *Prueba de Selección Universitaria (PSU)*.

The PSU was offered once a year in December. It consisted of four components: mathematics and Spanish, which were compulsory, and science and/or history. For each component, the raw scores were standardized at the national level to produce a normal distribution of scores ranging from 150 to 850, with a mean of 500 and a standard deviation of 110. Admission to each program was determined by the number of places available and the number of applicants. The score used to determine admission was calculated as a weighted average of high school GPA and all sub-components of the PSU test, with program-specific weights. Programs would rank students based on the program-specific score, and fill the available spots based on their ranking.[7]

In 2011, annual tuition fees posted by HEIs were high relative to other OECD countries. For this reason, Chilean higher-education students relied heavily on financial aid. The Chilean government

[7]See Kapor et al. (2024) for a detailed description of the algorithm matching students to programs.

Figure 1: PSU scores of higher-education students enrolled in HE in 2010 by field of study



Note: The figure plots the distribution of PSU scores for high school graduates enrolling in different university programs in 2010. The gray horizontal line represents the mean test score of all university enrollees who took the PSU.

provided access to loans and scholarships, covering up to 90% of the tuition fees. Therefore scholarships were a useful policy tool to attract individuals into specific university programs.[8] Access to most financial aid was determined by the weighted average of the PSU test scores obtained in mathematics and Spanish.

The teaching profession. In order to work in the school system, teachers need to have a higher education degree in teaching.[9] These degrees are level-specific and can only be granted by some institutions: for preschool and elementary teachers, this can be a university or a vocational institution, while the degree to become a high-school teacher can only be granted by universities. At the time of the reform, 86% of higher-education students enrolled in a teacher training program were enrolled in a university program and 14% were enrolled in vocational institutions.

While at the time of the reform there was no shortage of teachers in Chile (OECD, 2010), individuals who pursued teacher training were historically negatively selected in terms of their PSU score (Gallegos et al., 2022). Figure 1 plots the distribution of PSU scores for students enrolling in universities in 2010 by field of study. The figure shows that students enrolling in teacher training programs performed significantly worse than students enrolling in other fields, scoring on average almost one standard deviation below the mean of all university enrollees taking the PSU.

After graduation, new teachers can decide to apply for a teaching job. In Chile, there are three

[8]Recent papers highlight how, in our setting, students are credit-constrained (Solis, 2017) and price-sensitive to tuition changes (Bucarey et al., 2020; De Falco and Reichlin, 2025).

[9]There are a few rare exceptions, such as teachers recruited through Enseña Chile, a program modeled similarly to Teach For America, which places university graduates in underserved schools. However, they represent only a negligible share of the teaching workforce (De Gregorio and Neilson, 2021).

types of schools determined by their funding and administration authority: *public* schools which are publicly funded and managed, representing 46.2% of the schools in 2010; *voucher* or *subsidized* schools which are publicly funded, but privately managed, representing 47.7% of the schools; and fully *private* schools, representing 5.4% of the schools. The recruitment of teachers varies by type of school. For public schools, municipal authorities organize the hiring of teachers. Vacancies are publicly posted, and a commission reviews applications by rating applicants based on professional performance, seniority, and other formal training. Wages are set nationally and are low compared to other occupations.[10] In the privately managed schools, schools have discretion in deciding how and whom to hire. There is, therefore, little scope for public schools to try to hire higher-quality teachers. In privately managed schools, however, wages can be negotiated and are generally more competitive than in public ones.

2.2 The 2011 reform

This negative selection into teacher training programs motivated the Ministry of Education to implement a reform in 2011 aimed at changing the distribution of PSU takers pursuing a teaching career. More precisely, the reform introduced a scholarship, called the *Beca Vocación de Profesor* (BVP) for high-achieving college students who enrolled in an *eligible* teacher training program.

The BVP offered several benefits, most notably a full tuition waiver, which was awarded solely on the basis of the average PSU score in mathematics and Spanish. Thus, all individuals who scored one standard deviation above the mean (i.e. above 600 points) and who enrolled in an eligible program, were offered the opportunity to study for free.[11] In addition, individuals who scored above 700 points could receive a monthly stipend of 80,000 Chilean pesos (almost 45% of the 2011 national minimum wage), and those who scored above 720 points were given the opportunity to study one semester abroad. Given the standardized distribution of PSU scores, around 20% of all PSU takers were potentially eligible for the tuition waiver and about 5% for the additional benefits. Our discussion of the impact of the reform will focus on the first threshold, mainly because its effects at higher PSU cutoffs are hard to detect due to the small number of PSU takers scoring above 700. Individuals scoring just below the 600 threshold were eligible for different types of financial aid depending on their family income, choice of institution, and cohort of study. Appendix Table B1 summarizes financial aid eligibility by these three dimensions.

The BVP required its beneficiaries to work in publicly funded schools for at least three years in the seven years following college graduation. This requirement was reduced to two years if the school was located in a rural area. Individuals who did not meet this requirement had to repay the scholarship.

[10] Tincani (2021) shows that non-teaching salaries in Chile are roughly 60% higher, conditional on the level of education.

[11] Note that while the standard deviation of each individual component is 110, the standard deviation of the combined mathematics and Spanish scores is around 100. This cutoff was set to be slightly lower (580) for students who were eligible for an additional scholarship, called *Beca Excelencia Académica* (BEA), which was awarded to students coming from public and subsidized schools, who were in the top 10% of students in their respective high schools.

Finally, the introduction of the BVP imposed some requirements on universities. To be considered *eligible*, programs had to meet two conditions. First, the program had to be *accredited*[12], and second, they could only admit up to 15% of their applicants with a PSU score below the standardized mean (i.e., below 500 points). In practice, this meant that low-scoring PSU takers had fewer places available to enroll in teacher training programs, as they could either try to enroll in the few slots offered by eligible programs or enroll in non-eligible programs. Overall, in our study period around 80% of the programs of CRUCH universities and 15% of the programs of non-CRUCH universities accepted BVP recipients, effectively limiting the access to their programs to applicants scoring below the mean. Additionally, almost 70% of the eligible programs did not accept any applicant scoring below 500. The other 30% of eligible programs accepted on average 7% of candidates scoring below 500.

2.3 The reform and teaching quality: a theoretical framework

We develop a simple framework to formalize how the reform may affect average teacher performance, or productivity. At its core, the policy leverages admission test scores under the assumption that academic ability is an important predictor of teaching effectiveness. While this channel captures a direct form of selection, it is unlikely to be the only one at play. Individuals whose decision to enter the teaching profession responds to the reform may differ along other unobserved traits that also influence teaching quality. Depending on how these traits correlate with academic ability and with the decision to teach, the reform's overall impact on productivity may be amplified or attenuated.

Setup. We model entry into teaching as a Roy-style occupational choice. The framework is deliberately minimal to make the mapping from primitives to composition effects transparent. It follows the classic occupational-choice logic of Roy (1951) and the more recent "motivated agents" literature (Besley and Ghatak, 2005; Dal Bó et al., 2013), adapted to teacher recruitment.

Productivity for an individual with academic ability A in teaching and in the outside (non-teaching) sector is given by

$$
\begin{aligned}
P_T &= \kappa_T A + \epsilon_T, \\
P_O &= \kappa_O A + \epsilon_O. \qquad (1)
\end{aligned}
$$

Productivity depends on academic ability (A), measured in our setting by the observed PSU test score; a teaching residual ability (ϵ_T), capturing effectiveness as a teacher beyond A; and an outside-sector residual ability (ϵ_O). The coefficients κ_T and κ_O measure how strongly academic ability translates into productivity in each sector.

[12]Since 2006, Chile has had an accreditation system for higher education institutions, designed to regularly evaluate and guarantee the quality of the programs offered by these higher education institutions. The accreditation process is carried out by an agency of the Ministry of Education. While accreditation is generally voluntary, this is not the case for medicine and teacher training programs, which since 2009 must go through the process.

The utilities from choosing teaching versus the outside sector are

$$\begin{aligned} U_T &= \bar{w}_T + \epsilon_M + r_T P_T + S(A), \\ U_O &= \bar{w}_O + r_O P_O. \end{aligned} \tag{2}$$

Here, $\bar{w}_T$ and $\bar{w}_O$ denote average wages in the teaching and outside sectors, respectively, and r_O is a pay-productivity slope, i.e. the return to a unit of productivity in the outside sector. We normalize $r_T = 0$.[13] The utility from teaching, U_T, also depends on a preference-for-the-job parameter (ϵ_M), or intrinsic motivation for teaching. We assume that $(\epsilon_M, \epsilon_T, \epsilon_O)$ follow a joint distribution that allows for correlation across the three dimensions.

The policy term $S(A)$ captures the benefits and costs of choosing the teaching career based on test scores. In our setting, $S(A) > 0$ for $A \geq 600$, representing the scholarship eligibility conditional on entering a teaching training program, while $S(A) < 0$ for $A < 500$, representing the cost incurred by individuals facing a narrower choice set.

A student of academic ability A observes $(\epsilon_T, \epsilon_O, \epsilon_M)$ and chooses teaching if $U_T \geq U_O$. Thus, our model nests two strands of literature: (i) selection when outside wages are more responsive to productivity than public-sector wages (Bacolod, 2007; Tincani, 2021), and (ii) selection with non-pecuniary "mission" motives (ϵ_M) that may be correlated with productivity (Dal Bó et al., 2013; Ashraf et al., 2020). Collecting constants and terms that depend only on A, define

$$K(A) := (\bar{w}_T - \bar{w}_O) - r_O \kappa_O A, \tag{3}$$

and the selection index

$$Z := -r_O \epsilon_O + \epsilon_M, \tag{4}$$

so that

$$U_T - U_O = K(A) + Z + S(A). \tag{5}$$

An individual enters the teaching profession if and only if $Z \geq -K(A) - S(A) = C(A)$, where $C(A)$ represents the choice cutoff. Hence, the decision is driven by two forces: (i) the outside sector pulls away individuals whose productivity is well rewarded outside ($-\, r_O\, \epsilon_O$), and (ii) the motivation (ϵ_M) which directly raises the payoff from teaching.

Characterizing entrants and stayers. From a baseline without $S(A)$, the introduction of the policy term $S(A)$ shifts the choice cutoff $C(A)$ in the selection index $Z = -r_O \epsilon_O + \epsilon_M$ for the targeted ranges of academic ability. A scholarship ($S(A) > 0$) lowers the cutoff, attracting individuals who, in the absence of $S(A)$, would have fallen just below the threshold. Conversely, a reduction of the

[13]Without loss of generality, we set $r_T = 0$, as only relative returns are relevant for understanding selection patterns and characterizing those who enter the teaching profession. All results therefore hold under the assumption that $r_O > r_T$, which is reasonable for Chile during the studied period. When students made their enrollment decisions, teachers in public schools were paid according to a wage scale that was almost entirely determined by seniority. Wage dispersion in voucher schools was somewhat higher but still considerably lower than in the private sector (Tincani, 2021).

available choice set of university programs ($S(A) < 0$) raises the cutoff, retaining only a stricter upper tail of the distribution of Z. Under the assumption that $(\epsilon_T, \epsilon_O, \epsilon_M)$ are jointly normally distributed, the expected change in ϵ_T when $S(A)$ is introduced can be written as

$$\Delta\mathbb{E}[\epsilon_T] \;=\; \frac{\mathrm{Cov}(\epsilon_T, Z)}{\mathrm{Var}(Z)}\,\Delta\mathbb{E}[Z], \tag{6}$$

where Δ denotes "post-reform minus pre-reform" for the selected A group (see proof in Appendix Section D). The shift in the cutoff $C(A)$ implies $\Delta\mathbb{E}[Z] < 0$ with the introduction of the scholarship and $\Delta\mathbb{E}[Z] > 0$ under the case of choice set reduction. Thus, the sign of $\Delta\mathbb{E}[\epsilon_T]$ is determined by the covariance $\mathrm{Cov}(\epsilon_T, Z)$, which can be rewritten as:

$$\mathrm{Cov}(\epsilon_T, Z) = \sigma_T\big(\,\rho_{T,M}\sigma_M - r_O\,\rho_{T,O}\sigma_O\,\big), \tag{7}$$

where $\sigma_T^2 = \mathrm{Var}(\epsilon_T)$, $\sigma_O^2 = \mathrm{Var}(\epsilon_O)$, $\sigma_M^2 = \mathrm{Var}(\epsilon_M)$ and $\rho_{\cdot,\cdot}$ denote the corresponding correlations.

There are two channels that determine the sign of $Cov(\epsilon_T, Z)$. The first is an *outside pull* channel: if the traits that make someone a good teacher are also rewarded in the outside sector (i.e., large r_O and positive correlation $\rho_{T,O}$), then $\mathrm{Cov}(\epsilon_T, Z)$ is pulled downward. The second is a *motivation* channel: if intrinsic motivation to teach is positively related to teaching effectiveness, then $\mathrm{Cov}(\epsilon_T, Z)$ is pulled upward.

When the cutoff $C(A)$ is lowered (so $\Delta\mathbb{E}[Z] < 0$), a scholarship increases the average ϵ_T among entrants precisely when the outside pull dominates the motivation channel (i.e., when $r_O\rho_{T,O}\sigma_O > \rho_{T,M}\sigma_M$). The intuition is straightforward: in the absence of the policy, if the outside pull channel dominates, individuals who self-select into teaching are more likely to be those with a lower ϵ_T. Thus, when $S(A)$ is introduced, the marginal entrant is *positively selected* on ϵ_T.[14] By contrast, when the cutoff $C(A)$ is higher, average quality rises if the motivation channel dominates (i.e., $\rho_{T,M}\sigma_M > r_O\rho_{T,O}\sigma_O$).

Overall change in value-added. What are the implications for the overall change in P_T? In general, it reflects two components: (i) the direct effect of shifting the distribution of A, which raises productivity through the $\kappa_T A$ term, and (ii) the change in residual teaching quality, ϵ_T, which depends on how the reform alters the composition of entrants.

For example, when the motivation channel prevails, scholarships increase the share of high-A individuals, but at the same time the marginal entrants are negatively selected on ϵ_T. In this case the reform generates an academic ability–motivation trade-off: average value-added rises through stronger academic ability, yet falls at the margin due to lower residual quality. By contrast, when the policy relies on disincentives, the marginal exit is negatively selected on ϵ_T, so the remaining teachers combine higher academic ability with stronger residual quality, and P_T increases unambiguously.

[14] This intuition closely parallels that in Borjas (1987). See also Bacolod (2007) for an application of the model proposed by Borjas (1987) to selection into teaching.

3 Recruitment

In this section, we study the effectiveness of the reform in terms of incentivizing (or discouraging) prospective college students from the top (bottom) of the PSU distribution to enroll in a teacher training program and subsequently pursue careers in publicly funded schools. To quantify the impact of the reform, we exploit the local variation of its two components in a Regression Discontinuity (RD) framework and provide causal estimates of their local impact. We complement this analysis with an across-cohort comparison of individuals enrolling before and after 2011, with the aim of providing suggestive evidence on the overall effects of the reform beyond the RD cutoffs.

3.1 Data

We use rich administrative data on the universe of PSU test takers and HEI enrollees between 2009 and 2014, provided by the Chilean Ministry of Education (MINEDUC) and the PSU administrative body (DEMRE). In addition to PSU scores, DEMRE provided information on the sociodemographic characteristics of test takers, including gender, high school GPA, and parental education. We merge this information with MINEDUC data on enrollment, graduation, and a rich set of program characteristics, including institution type, field of study, and data on application and take-up of financial aid. Finally, we combine this with administrative records on the universe of teachers working in primary and secondary education, also provided by MINEDUC. This results in a unique panel dataset that tracks PSU takers through their university years and into their teaching careers up to 2024, allowing us to identify key aspects of their higher-education outcomes and subsequent career trajectories as teachers.

3.2 Empirical strategy

We carry out a regression discontinuity analysis on first-time PSU takers between 2011 and 2014. This restriction avoids including individuals who may retake the test in order to improve their scores, either to become eligible for the BVP or to expand their choice set of institutions. We estimate the following model

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 D_i \times PSU_i + \beta_3 PSU_i + \nu_i \tag{8}$$

where Y_i is our outcome of interest, D_i is either an indicator of BVP eligibility or of being below the restriction cutoff for eligible institutions. We allow Y_i to depend linearly on the PSU score (PSU_i), with different slopes above and below the relevant threshold. Our parameter of interest, β_1, captures the difference in outcomes for individuals just affected by the scholarship eligibility or the restrictions, relative to those marginally unaffected. For both cutoffs, we apply a triangular kernel-weighting around the threshold. For the BVP eligibility cutoff, bandwidths are optimally chosen following Calonico et al. (2019). For the 500-point cutoff, bandwidths are set to 24 in order to avoid capturing the effects of an important policy for prospective college students in that period, namely the student loan program, which was available for those scoring above 475 points.[15] Summary statistics of the analyzed sample are presented in Appendix Table B2.

Identification. For the sharp RD estimate β_1 to identify the causal effect of the scholarship and the restrictions, standard identification assumptions must hold. Most importantly, potential outcomes must be continuous around the cutoff. This assumption could be violated if PSU takers were able to manipulate their scores. Results from a McCrary test (McCrary, 2008), presented in Appendix Figure A1, provide no evidence of manipulation. We complement this with a balance test, where we estimate model (8) for some sociodemographic characteristics. Results in Appendix Table B3 show that being marginally eligible is not significantly correlated with observables.

As mentioned above, individuals who are marginally below the cutoff may retake the test in subsequent years to improve their scores. Appendix Figure A2 plots the probability of becoming eligible in the second, third, and fourth years after high school graduation as a function of the first-time PSU score. Only about 10% of retakers manage to improve enough to change their eligibility status. For this reason, our results would remain virtually unchanged under a fuzzy RD design. Adopting a sharp RD specification allows us to analyze a longer time frame.

3.3 Results

Figure 2 plots the likelihood of enrolling in a teacher training program and of being employed as a teacher two years after graduation, as a function of the PSU score. These two relationships discontinuously change at both cutoffs of interest.[16]

Table 1 reports the RD parameters at the 600 cutoff (Panel A) and the 500 cutoff (Panel B). For enrollment in teacher training programs, individuals marginally eligible for the scholarship are 27% more likely to enroll (1.9 p.p.), while those below the 500 cutoff are 36% less likely to enroll (2.9 p.p).[17] Estimates relative to the baseline are therefore substantial, even though the absolute numbers are modest, given that teacher training enrollees constitute a small proportion of the overall population of PSU takers. Column (2) of Table 1 shows results for graduation, defined as obtaining a degree from a teacher training program within eight years of the first PSU attempt. At the 600 cutoff, the effect corresponds to a 13% increase relative to the baseline. For those scoring below 500, the relative effect is 17%.[18]

These findings closely replicate those of Gallegos et al. (2022) for enrollment and graduation. The authors also examine the impact on employment immediately after graduation. We extend

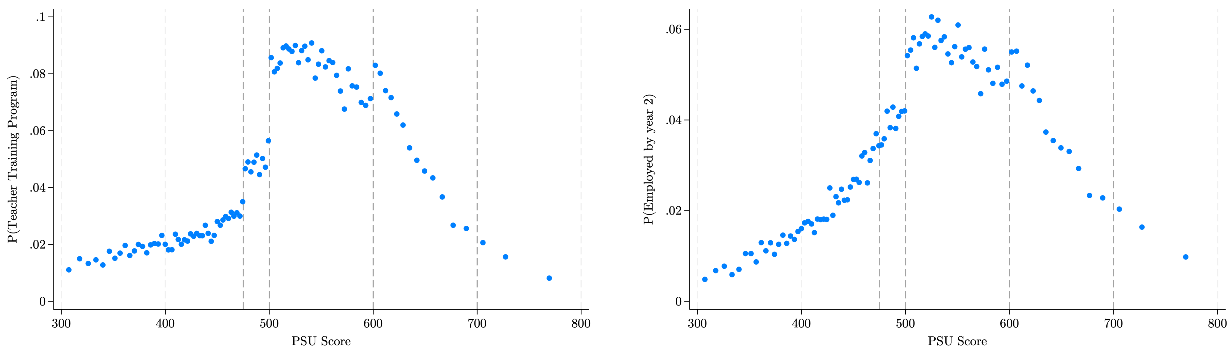
[15]See Solis (2017) for a discussion of the loans programs.

[16]The discontinuity at 475 can be explained by the fact that students scoring above this cutoff become eligible for student loans, which largely affects the general enrollment decision margin (Solis, 2017). The discontinuity is also visible before the reform, as shown in Appendix Figure A7.

[17]In Appendix Figures A3 and A4 we show heterogeneous results by type of institution. Results are mostly driven by the most prestigious universities (CRUCH), which completely prevent students scoring below the 500 cutoff from enrolling and are more attractive to BVP-eligible test takers.

[18]Although we pool all cohorts, Appendix Figure A5 documents some heterogeneity across cohorts. For the BVP treatment, effects are particularly pronounced in 2011, most likely due to the rollout of more generous financial aid from 2012 onward (De Falco and Reichlin, 2025)

Figure 2: Enrollment and employment variables as a function of PSU score



(a) Enrollment Teacher Training Program

(b) Employed as Teacher

Note: Panel (a) shows the probability of enrolling in a teacher training program one year after high school graduation, as a function of the PSU score. Panel (b) shows the probability of ever being employed as a teacher two years after college graduation, as a function of the PSU score. The sample includes all PSU takers between 2011 and 2014. The four horizontal lines refer to the cutoff for loan eligibility (475), the cutoff preventing students from enrolling in some BVP-eligible teacher training programs (500), the cutoff defining the eligibility for the BVP scholarship (600), and the one giving access to a monthly stipend together with the scholarship (700).

the analysis in two important ways. First, we examine the types of schools where graduates begin working, which is a relevant factor given that the scholarship specifically incentivizes employment in publicly funded schools. Second, we examine employment outcomes up to five years after graduation. This medium-term perspective is important since the scholarship-induced incentives to work in specific schools cease after three years, so any short-term effects may fade over time.

We analyze the role of policy-induced sorting incentives into different types of schools by tracking graduates into their teaching careers. In column (3), we construct an indicator equal to one if an individual was ever employed in a school within the first two years after graduating from a teacher training program. We repeat this exercise separately for each type of school: public, voucher, private, and rural. As shown in columns (3) to (7) of Table 1, we find differences in the types of schools where teachers work after graduation, consistent with the rules and incentives faced by BVP recipients. Focusing on PSU takers who were eligible for the scholarship, the effect on school employment is about 19% relative to the baseline. This effect is concentrated among public and voucher schools. Voucher schools are generally viewed as more attractive options, as reflected in their higher baseline employment rates (3% vs 1.4%), yet the reform still produces a sizable effect for public schools. Although the absolute effect is somewhat smaller for public schools, the relative effect is larger given their lower baseline. Regarding teachers affected by the restricted choice set, we also see a 21% reduction relative to the baseline, which is again mostly driven by public and voucher schools. In the short run, the policy was effective in leading high-academic-ability teachers into schools that the Ministry of Education considered to be in greater need of them. Figure 3 shows employment results up to five years after graduation, focusing on the role of the scholarship in shaping longer term employment decisions. The results suggest that incentivized teachers are

Table 1: Regression Discontinuity results

	(1) Enrolled	(2) Graduated	(3) Employed	(4) Public	(5) Voucher	(6) Private	(7) Rural
Panel A. 600 cutoff							
RD Estimate	0.019***	0.008**	0.009***	0.004***	0.006***	0.000	0.001
	(0.003)	(0.003)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Bandwidth	43	47	53	62	56	56	62
Observations	125,458	135,437	152,497	178,390	162,619	162,619	178,390
Baseline Mean	0.068	0.061	0.047	0.014	0.030	0.008	0.004
Percentage Change	27.4	13.1	19.3	29.0	20.4	5.7	23.3
Panel B. 500 cutoff							
RD Estimate	-0.029***	-0.013***	-0.012***	-0.005***	-0.009***	0.000	-0.002*
	(0.005)	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
Bandwidth	24	24	24	24	24	24	24
Observations	113,468	113,468	113,468	113,468	113,468	113,468	113,468
Baseline Mean	0.082	0.075	0.054	0.021	0.036	0.003	0.008
Percentage Change	-35.7	-16.7	-21.4	-22.6	-24.6	13.1	-22.6

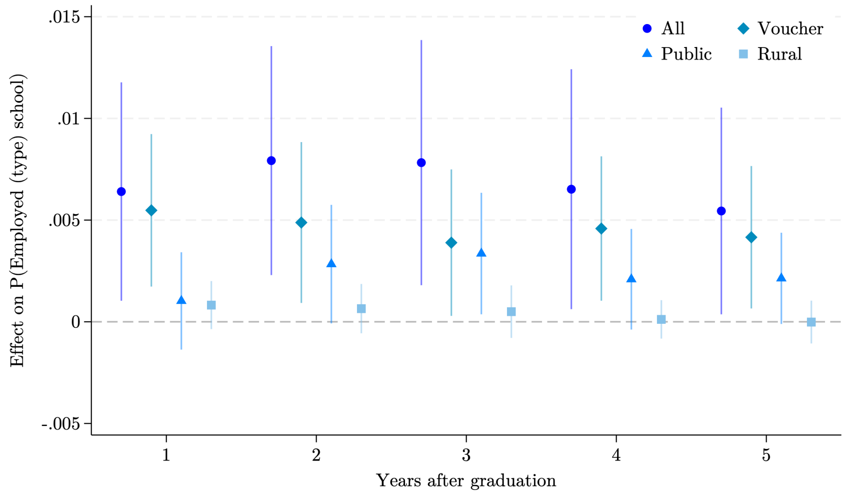
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table shows Regression Discontinuity parameters where all specifications are estimated using weighted local linear regressions and consider high school graduates taking the PSU between 2011 and 2014. For the 600 cutoff, the RD Estimate corresponds to the effect of being marginally *above* 600 or marginally eligible for the scholarship; bandwidths are chosen optimally according to Calonico et al. (2019). For the 500 cutoff, the RD Estimate corresponds to the effect of being marginally *below* 500 or marginally affected by the restriction; bandwidths are set to 24 to avoid capturing the effect of the loan eligibility cutoff at 475 points. Standard errors are clustered at the PSU test score level and reported in parentheses. *Observations* correspond to the number of observations with non-zero weight given the chosen bandwidth. The *Baseline Mean* refers to the outcome mean for marginally unaffected students. The *Percentage Change* is computed as the coefficient over the Baseline Mean (times 100). *Enrolled* is an indicator for individuals enrolling in a teacher training program in the year after high school graduation. *Graduated* is an indicator for individuals graduating from a teacher training program within eight years of high school graduation. *Employed* is an indicator for individuals observed working in any school within 2 years of their graduation from a teacher training program. *Public*, *Voucher*, *Private*, and *Rural* are indicators for individuals observed working in each particular type of school within 2 years of their graduation from a teacher training program. Results are conditional on taking the college admissions test.

more likely to remain in the profession beyond the requirements imposed by the scholarship, and this holds for both public and voucher schools.

Robustness and sensitivity. We perform a series of robustness checks to test the validity of our results. Supporting the identification of the RD parameters, Appendix Table B5 and Appendix Figure A6 show that the RD estimates are robust to different local polynomials and bandwidth choices.[19] Additionally, Appendix Figure A7 shows that the documented discontinuities were not present in the years prior to the reform. We test formally whether these discontinuities existed in Appendix Table B4, where we estimate model (8) for cohorts of study before the reform.

[19] Not surprisingly, given that observables are balanced at the cutoff, the results remain virtually unchanged when adjusting for observable characteristics (such as cohort fixed effects or student covariates) and are robust to alternative choices of standard errors. We remain conservative and report clustered standard errors.

Figure 3: BVP eligibility and employment at school over time



Note: The figures show RD parameters at the 600 cutoff using as outcomes the probability of being employed in a school, and the probability of being employed by type of school up to five years after graduation. For the RD results, the sample is restricted to those taking the college admissions test between 2011 and 2014.

Effects away from the cutoff and change in PSU distribution. While the RD design provides local estimates on the effects of the reform, it does not directly test for the change in the PSU distribution of the entire population of teacher training enrollees. To address this, we complement the RD evidence with a simple across-cohort comparison, which allows us to analyze effects away from the cutoffs and examine PSU statistics of teacher training enrollees over time.

Specifically, we compare the outcomes of cohorts 2009–2010 and 2011-2012 and replicate the results from Table 1 and Figure 3 in Appendix Table B6 and Appendix Figure A9, respectively. The estimates relative to the baseline are broadly consistent with the local RD results. Naturally, this before-after approach rests on a stronger assumption - crucially, the absence of underlying trends in outcomes. Appendix Figure A10 provides supporting evidence that for most outcomes, pre-trends are limited. Nevertheless, these results should be interpreted with caution, as they are less likely to be identified than the RD estimates. Yet, they remain informative for understanding the overall effects of the reform.[20]

After 2011, teacher training enrollees had, on average, higher PSU scores. As shown in Appendix Figure A11, the mean PSU increased by 10 points, corresponding to 15% of the overall variation in PSU among teacher training enrollees. Similar patterns are observed for other distributional statistics, such as the 10th and 90th percentiles. Taken together, this section shows that the reform successfully attracted the targeted high school graduates to the teaching profession. As a result, the average PSU scores of newly hired teachers also increased, especially in publicly funded schools. Whether this translated into an increase in their average performance as teachers is the

[20] See Appendix Figure A8 for a non-parametric representation of the probability of working in different types of school, as a function of PSU. The probabilities are computed conditional on graduating from a teacher training program, to show how the graduates are distributed in different school types, and how this changed after the reform.

question we seek to answer in the remainder of this paper.

4 Teacher quality

In this section, we quantify the change in teacher quality due to the reform, as measured by the value added by teachers to their elementary school students' test scores. The analysis focuses on elementary mathematics and Spanish teachers. We begin by discussing our data and sample. We describe the empirical model and the procedure for estimating the standard deviation of teacher value added, which will serve as a benchmark to understand the effect of the policy. Then, we present the results on the value-added model quantifying the change in teaching quality induced by the reform. Finally, we discuss the role of the PSU as a screening device.

4.1 Data and population of interest

We complement the data described in Section 3 with administrative data provided by the Ministry of Education (MINEDUC) covering the universe of Chilean students enrolled in primary education. The data include information on students' average GPA, gender, and date of birth, as well as the school, grade, and classroom in which they are enrolled each calendar year. In addition, we use individual-level results of a standardized test called SIMCE, provided by the *Agencia de Calidad de la Educación*, for the years 2012 to 2018. This test is conducted annually for all Chilean students in a specific grade. Appendix Table B7 summarizes the grades tested each year. Information on parental education and family income quintiles is collected by the *Agencia de Calidad de la Educación* through a survey. Crucially, we can merge these data with information on each student's teachers, including their years of experience, PSU test scores, and higher education enrollment and graduation.

We restrict our sample of interest to primary school teachers of mathematics and Spanish. We apply two further sample restrictions closely related to our identification strategy. First, we limit the sample to teachers for whom PSU test score data are available. Information on PSU takers is available starting in 2004. Thus, we focus on seven teacher cohorts who took the PSU before the reform - i.e. whose decisions were not affected by it - and four cohorts who took it afterward. Second, we focus on teachers whose students have complete information on all variables included in the value-added model, which we describe in the next section. This restriction implies that our analysis focuses only on 6th- and 8th-grade teachers, since we lack data on both mathematics and Spanish lagged test scores for students in other grades (see Appendix Table B7). For these students, lagged test scores are measured two years before the current test. Taken together, this means that our population of interest is the universe of mathematics and Spanish teachers trained between 2004 and 2014 who taught in 6th and 8th grade in a year when SIMCE was administered.

Appendix Tables B8 and B9 provide some summary statistics of our sample of teachers. Consistent with the findings discussed in the previous section, teachers who pursued their training after the reform and obtained a PSU score above 600 are more likely to work in public and subsidized schools. Unsurprisingly, teachers who enrolled after the reform have, on average, less than one year of teaching experience, compared to roughly 2.7 years for those who enrolled before. Finally, Appendix Figure A12 shows the distribution of PSU scores before and after the reform for teachers in our sample. On average, standardized PSU scores increased by 0.31 for math teachers and by 0.40 for Spanish teachers after the reform.

4.2 Estimation of teacher quality

We begin by estimating the standard deviation of the teacher effects. This provides an insight into the heterogeneity among teachers in terms of their impact on student learning. More precisely, it gives a measure of teacher effectiveness in terms of the return to test score from being taught by a teacher whose impact is one standard deviation higher. Moreover, it serves as a useful benchmark against which to contrast our results on the change in TVA estimates induced by the reform.

4.2.1 The variance of teacher effects

Consider the following model

$$Y_{isjgtc} = \lambda Y_{i,t-1} + \gamma \mathbf{X}_{i,t} + \theta_{cjst} + \theta_s + \theta_{js} + \alpha_t + \mu_g + v_{isjgtc} \tag{9}$$

where Y_{isjgt} is the (mathematics or Spanish) test score of student i, in school s, taught by teacher j, in grade g, at time t, and in classroom c. The variable $Y_{i,t-1}$ corresponds to the lagged value of student i's test score, and $\mathbf{X}_{i,t}$ is a vector of individual characteristics measured at time t. They include indicators for being female, belonging to a specific family income quintile, and parental education. Parameters θ_{cjst}, θ_s and θ_{js} represent respectively classrooms, schools and teachers effects. Finally, α_t and μ_g are year and grade fixed effects, and v_{isjgtc} are idiosyncratic random shocks. Estimating the variance of θ_{js}, defined as σ^2_{js}, comes with some complications. In our setting, many parameters from model (9) are not identifiable. For example, as long as teachers teach in one school and one class, it is not possible to separate the teacher effect from the school or classroom effects. The empirical variance of $\hat{\sigma}^2_{js}$ is therefore an upward biased estimator for the true variance σ^2_{js}. In order to produce an unbiased estimator for σ^2_{js} we start by estimating the following value-added model

$$Y_{isjgtc} = \lambda Y_{i,t-1} + \gamma \mathbf{X}_{i,t} + \delta_{cjst} + \alpha_t + \mu_g + v_{isjgtc} \tag{10}$$

where δ_{cjst} combines classroom, teacher, and school effects. It can be shown using model (9) that the variance of the school effects can be expressed as $\sigma^2_s = Cov(\delta_{cjst}, \delta_{cj'st})$. That is, the covariance of classroom effects between classes taught by different teachers, during the same year. Using a similar argument, the covariance of classroom effects taught by the same teachers identifies the sum of the school and teacher variances, that is $Cov(\delta_{cjst}, \delta_{c'jst}) = \sigma^2_s + \sigma^2_{js}$ (Kane and Staiger, 2008).[21]

Appendix Figure A13 shows the frequency with which we observe different teachers in the

[21]The assumption behind this result is the independence of the error term v_{isjgt} and $v_{isjgt'}$. This would be arguably violated if teachers teach the same students across years. In our sample, this happens in less than 1% of the cases.

same school, teachers in more than one classroom and those observed in more than one year. Note that within the same year we do not often observe different teachers in the same school. However, our data have the advantage that across years, in 40% of the cases we observe different teachers working in the same school, and 55% of the teachers are observed in more than one classroom. This implies that we can estimate σ_s^2 as the covariance of $\hat{\delta}$ for two different teachers, in the same school, regardless of the year ($Cov(\hat{\delta}_{cjst}, \hat{\delta}_{cj'st'})$). Similarly, we estimate $\sigma_s^2 + \sigma_{js}^2$ as the covariance of $\hat{\delta}$ for different classrooms taught by the same teacher, in the same school, regardless of the year.[22] We estimate σ_{js}^2 and σ_s^2 separately for mathematics and Spanish.

Table 2: Effect of 1 SD increase in school and teacher effects

	With weights		Without weights	
	Math (1)	**Spanish** (2)	**Math** (3)	**Spanish** (4)
School	0.22	0.12	0.22	0.13
Teacher	0.10	0.11	0.09	0.11

Note: The table reports the effect of receiving a one standard deviation better school and teacher on students' subject-level test scores. School variance σ_s^2 is computed as $Cov(\hat{\delta}_{cjst}, \hat{\delta}_{cj'st})$, and the teacher variance σ_{js}^2 is computed as the difference between $Cov(\hat{\delta}_{cjst}, \hat{\delta}_{c'jst})$ and $Cov(\hat{\delta}_{cjst}, \hat{\delta}_{cj'st})$, where $\hat{\delta}_{cjst}$ is the OLS estimate of δ_{cjst} in model (10). In columns (1) and (2), we use the sum of the sizes of both classrooms involved as analytical weights when computing the covariance. In columns (3) and (4), we exclude the weights such that each classroom pair has the same weight when computing the covariance.

Table 2 summarizes the results.[23] Estimates for the standard deviation of teacher effects are similar to those found in the United States (Bacher-Hicks, 2022; Chetty et al., 2014a). However, they appear somewhat smaller in magnitude when compared to research conducted in developing countries, such as the studies by Bau and Das (2020) and Buhl-Wiggers et al. (2022).

4.2.2 Teacher value-added and the 2011 reform

To estimate the impact of the reform on teaching quality, we estimate a value-added model that compares teachers trained before and after the reform. Consider the following simple value-added model:

$$Y_{isjgtc} = \lambda_0 + \lambda_1 Y_{i,t-2} + \gamma_1 \mathbf{X}_{it} + \gamma_2 \mathbf{C}_{ct} + \gamma_3 \mathbf{S}_{st} + \tau_j + \alpha_t + \mu_g + \varepsilon_{isjgtc}, \tag{11}$$

[22] Similar procedures have been implemented in the literature. See Araujo et al. (2016); Bau and Das (2020).

[23] Appendix Table B10 shows that our estimates remain almost unchanged when the covariance of $\hat{\delta}$ is computed using different classes for the same teacher, in the same school, in the same year, and when we estimate $\hat{\delta}$ in model (10) without including $\mathbf{X}_{i,t}$.

where Y_{isjgt} is the (mathematics or Spanish) test score of student i, in school s, classroom c, taught by teacher j, in grade g in year t. The term $Y_{i,t-2}$ refers to the student's lagged test scores in both subjects, measured two years earlier. $\mathbf{X}_{it}$ is a vector of student-level covariates (female, family income quintile, and parental education). $\mathbf{C}_{ct}$ captures classroom characteristics (class size, share of female students, average lagged test scores for both subjects, the distribution of students by family income quintile, and the distribution of parental education). $\mathbf{S}_{st}$ is a vector of school-level covariates (indicators for public, subsidized, technical, and rural schools). α_t and μ_g denote year and grade fixed effects, respectively. The coefficient τ_j recovers the teacher value-added (TVA) of each individual teacher, conditional on the included covariates.

We now turn to how individual teacher quality might have been affected by the reform. Our parameter of interest is the difference in τ_j across teacher cohorts. We model this as

$$\tau_j = \Phi_0 + \sum_{\substack{y=2004 \\ y\neq 2010}} \Phi_y \mathbb{1}\{PSUcohort_j = y\} + e_j, \tag{12}$$

where $\mathbb{1}\{PSUcohort_j = y\}$ are indicators for the PSU cohort of teacher j. Plugging this expression into equation (11) yields our final estimating equation:

$$\begin{aligned} Y_{isjgtc} = \lambda_0 + \lambda_1 Y_{i,t-2} + \gamma_1 \mathbf{X}_{it} + \gamma_2 \mathbf{C}_{ct} + \gamma_3 \mathbf{S}_{st} + \alpha_t + \mu_g \\ + \sum_{\substack{y=2004 \\ y\neq 2010}} \Phi_y \mathbb{1}\{PSUcohort_j = y\} + \eta_{jt} + \epsilon_{isjgtc}, \end{aligned} \tag{13}$$

which combines the standard value-added framework with our cohort comparison. Here, Φ_y captures the difference in teacher quality for each cohort relative to the 2010 baseline cohort. Alternatively, we estimate model

$$\begin{aligned} Y_{isjgtc} = \lambda_0 + \lambda_1 Y_{i,t-2} + \gamma_1 \mathbf{X}_{it} + \gamma_2 \mathbf{C}_{ct} + \gamma_3 \mathbf{S}_{st} + \alpha_t + \mu_g \\ + \phi_1 \mathbb{1}\{AfterReform_j\} + \eta_{jt} + \epsilon_{isjgtc}, \end{aligned} \tag{14}$$

where we include a single indicator for teachers having enrolled in a teacher training program after 2011.

In models (13) and (14), η_{jt} additionally controls non-parametrically for teachers' years of experience. This is crucial, since comparing teachers trained before and after the reform, even after residualizing by student, classroom, and school characteristics, would otherwise omit one of the most important determinants of teacher quality: teaching experience. Appendix Figure A14 illustrates this point by showing the non-parametric relationship between TVA estimates and years of experience, which follows a clear upward trend and peaks at around five years. Mechanically, teachers who enrolled in teacher training programs before 2011 had, on average, more years of teaching experience at their disposal. However, Appendix Figure A16, which plots the distribution of years of experience by study cohort, shows that there is considerable variation in teaching experience within each cohort.

We estimate models (13) and (14) by weighting observations by the inverse of the number of students taught by each teacher, and cluster the standard errors at the teacher-by-year level. Reweighting ensures that each teacher contributes equally to the estimation, avoiding giving more weight to those with more students. This one-step procedure avoids any measurement error that could arise from first estimating (11) and then estimating (12) in a separate step.[24]

Identification. Identification of the parameters of interest relies on a conditional independence assumption. In other words, once we adjust for all the variables in the value-added model and years of teaching experience, any remaining bias in the comparison of teachers trained across PSU cohort would be negligible. It is important to stress that this comparison is carried out controlling for students' baseline test scores and a rich set of student, classroom and school characteristics. Crucially, this is a within-year and within-grade comparison. Any confounders that vary over time, such as the introduction of new policies, are absorbed by year fixed effects, as long as they do not differentially affect teachers belonging to different PSU cohorts.[25] As highlighted in Appendix Figure A15, there is substantial variation within the year of entering the labor force across PSU cohorts, providing variation to identify our parameters of interest.

4.3 Results

Table 3 presents our main findings. Column (1) shows that, after accounting for teaching experience and a rich set of student, classroom and school covariates, mathematics teachers who enrolled in college after the reform are, on average, more effective in improving students' test scores. The size of the effect is about 38% of the standard deviation of math teachers' value added. By contrast, column (2) shows that the estimated effect of the reform for Spanish teachers is an order of magnitude smaller and statistically indistinguishable from zero.

It is interesting to note that this heterogeneity in effects across subjects cannot be explained by differences in the change in the composition of teachers. As shown in Table 3, both groups of teachers experienced an increase in PSU scores of approximately one-third of a standard deviation, and this increase was even larger for Spanish teachers.

Trends in teaching quality and other policies. A potential threat to the identification strategy is the presence of trends in teaching quality across cohorts if university enrollment. If, for example, teacher training programs are becoming more effective over time, our estimates would be biased. Figure 4 plots the estimates of Φ_y from model (12). To increase precision, we pool together cohorts

[24]Value-added effects are generally estimated with error (Herrmann et al., 2016), and using it as an outcome in a two-step estimation process would likely induce attenuation bias.

[25]Among the reforms worth mentioning, there are the *Sistema de Desarrollo Docente* (Law 20.903) and the *Ley de Inclusión Escolar* (Law N 20.845). These reforms introduced changes in the salary structure for public school teachers and in the allocation of students to schools, respectively.

Table 3: Effect of the reform on TVA

	(1) Math teachers	(2) Spanish teachers
Post-reform cohort	0.038**	-0.004
	(0.016)	(0.014)
# Students	206,094	225,584
# Teachers	4,386	4,966
Δ Std PSU	0.320	0.387

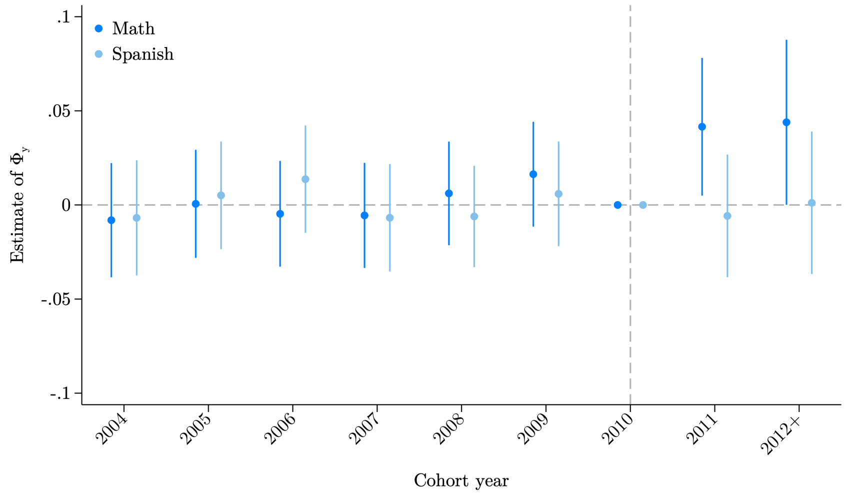
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table reports OLS estimates for the coefficients of interest in model (14). Δ *Std PSU* corresponds to the change in the average standardized teacher test scores. Standard errors, clustered at the teacher-by-year level, are shown in parentheses.

2012, 2013 and 2014 due to the small number of teachers trained after 2012 in our sample.[26] As the figure shows, before 2010 there were no significant differences in teaching quality across cohorts. For math teachers, the positive effects appear only after 2011, the year of the reform's implementation. A natural concern is that other changes in the Chilean higher education system after 2011 may differentially affect the cohorts we study. Two events are worth mentioning. First, the expansion of the centralized application system in 2012 (Kapor et al., 2024). Second, the introduction of free tuition in 2015. The latter does not enter our analysis, as our sample ends in 2014 precisely to avoid confounding from the free tuition reform, which effectively eliminated the BVP's effect on enrollment in teaching training programs (Castro-Zarzur et al., 2022). Regarding the 2012 application reform, the evidence suggests that this is unlikely to drive our results: the change in TVA is already visible in 2011 and remains similar in 2012. Potential contemporaneous effects in 2011, for example the student protests documented in González and Prem (2024), may also raise concerns. Our theoretical framework, however, provides a clear benchmark for evaluating whether such events could confound our estimates. The reform operates only through two channels: (i) changes in the composition of prospective teachers in terms of PSU, and (ii) changes in incentives for individuals with PSU below 500 and above 600. The framework predicts that the average TVA for individuals with PSU scores between 500 and 600 should not be affected by the reform. We test and discuss this result in the following section.

Sensitivity to TVA estimates and sample selection. We further test the robustness of our results in several ways. In Appendix Table B11, we test the sensitivity of our estimates to different specifications. We estimate model (14) controlling for more restrictive sets of covariates (including GPA, share of days at school) and less restrictive (including only school or class characteristics) sets of covariates. Results remain fairly stable. Despite controlling non-parametrically for years of experi-

[26]Among math and Spanish teachers, the number trained in each cohort after the reform is as follows: 2011 — 263 math and 310 Spanish teachers; 2012 — 129 math and 139 Spanish teachers; 2013 — 54 math and 53 Spanish teachers; 2014 — 13 math and 8 Spanish teachers.

Figure 4: Average change in TVA by cohort



The figure reports OLS estimates of Φ_y from (13). Confidence intervals at the 10% significance level are shown, with standard errors clustered at the teacher-by-year level.

ence in the specification, there may be concerns about the model being estimated using teachers who studied before the reform with more years of experience than those in the treatment group, namely, those who studied after. The relevant comparison, in fact, is at low levels of experience, since teachers in the treatment groups have, on average, less than one year of experience. As shown in Appendix Table B11 and Appendix Figure A17, the results remain essentially unchanged when restricting the sample to teachers with fewer than five, three, or two years of experience.[27] Finally, newer cohorts have had less time to appear in the dataset, and if higher-value-added teachers are more likely to complete their degrees on time, this could generate cohort differences in who is observed in the data. To assess whether this type of sample selection is driving our results, we re-estimate the model, restricting the sample to individuals who finish their studies on time and whom we first observe six or seven years after taking the PSU. As reported in Appendix Table B12, math teachers from post-reform cohorts exhibit higher TVA, while the estimates for Spanish remain close to zero, even in this subsample of teachers.

4.4 Heterogeneity

The BVP aimed not only to attract potentially better teachers into the profession, but also to match them with more vulnerable schools. The small proportion of rural and private schools does not allow us to test for heterogeneity of effects using these natural sample divisions. Therefore, we test whether the reform was more effective in the targeted schools by analyzing the heterogeneity of effects by school vulnerability. We define vulnerable schools using the IVE-SINAE index. This

[27]The reform may also have affected the returns to experience. We cannot directly test this hypothesis because very few teachers trained after the reform have more than one year of experience. However, this limitation is not critical for our analysis, since we non-parametrically compare teachers before and after the reform.

measure is calculated annually by the Ministry of Education for all publicly funded schools in order to identify children who, without appropriate support, are at high risk of dropping out of the school system. The IVE-SINAE corresponds to the percentage of students at risk in a given school (Cornejo, 2005).[28] We define high-vulnerability schools as those with an IVE-SINAE above the median (78% of vulnerable students).

Results in Table 4 show that the average change in math TVA is similar for both types of school (column (2)). Similarly to our main results by subject, the heterogeneity does not reflect differential changes in the average PSU of the relevant group of teachers.

Table 4: Effect of the reform on TVA - Heterogeneity by vulnerability

	Math teachers		Spanish teachers	
	(1) Low vulnerability	(2) High vulnerability	(3) Low vulnerability	(4) High vulnerability
Post-reform cohort	0.035*	0.044*	0.004	-0.009
	(0.020)	(0.024)	(0.020)	(0.022)
# Students	125,176	70,542	137,834	72,460
# Teachers	2,051	2,042	2,287	2,277
Δ Std PSU	0.45	0.27	0.47	0.45
Δ % PSU<500	-0.20	-0.15	-0.26	-0.25
Δ % PSU≥600	0.23	0.08	0.14	0.13
Baseline % PSU<500	0.31	0.49	0.41	0.60
Baseline % PSU≥600	0.18	0.09	0.12	0.04
Average IVE	0.63	0.88	0.63	0.88
Average very high IVE	0.38	0.69	0.37	0.69

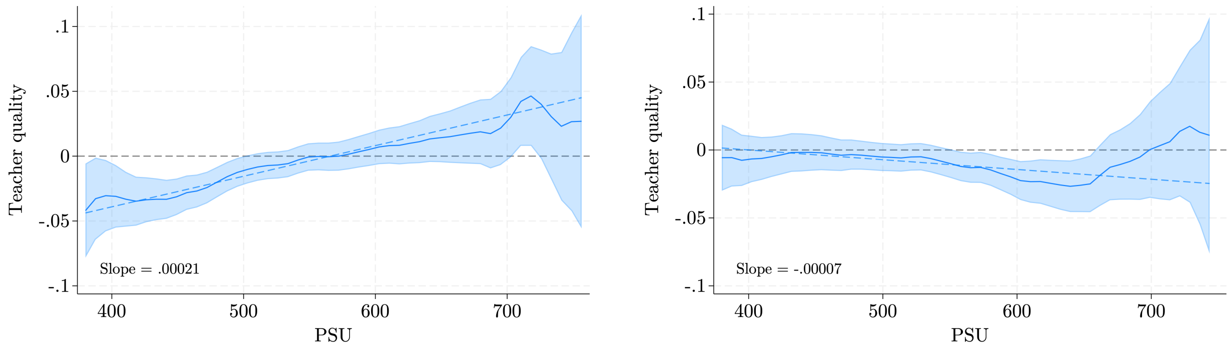
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table shows OLS estimates for the coefficients of interest in model (14). High vulnerability schools are defined as those with an IVE-SINAE above the median (78%). Standard errors are clustered at the teacher-by-year level, and shown in parentheses.

4.5 TVA and PSU: teaching quality and its screening device

Next, we test the reform's implicit assumption about the predictive power of PSU scores for an individual's teacher quality. In Figure 5, we plot the non-parametric fit alongside the linear fit for a teacher's value-added (estimates of τ_j from model (11)) as a function of their PSU score, for mathematics and Spanish teachers who began their studies before the 2011 reform. The figure shows that before the reform, PSU scores were predictive of teacher quality for mathematics teachers in primary school, but not for Spanish teachers. Specifically, a one standard deviation increase (100 points) in PSU score was associated with a 0.02 standard deviation increase in TVA for mathematics teachers (or about 1 year of experience). For Spanish primary school teachers, however, this coefficient is smaller in magnitude and not statistically different from zero.

[28]The IVE-SINAE identifies three priority groups: 1) children of medium priority, i.e. children living below the poverty line; 2) children of high priority, i.e. children living below the poverty line with additional risk factors for dropping out of school; and 3) children of very high priority, i.e. children at risk of extreme poverty. The IVE-SINAE index of the school is calculated as the sum of these three priority groups.

Figure 5: Pre-reform local polynomial fit of TVA on PSU



(a) Math

(b) Spanish

Note: The figures show the non-parametric relationship between TVA and PSU described by a local polynomial regression together with their linear fit, for teachers that started their studies before the 2011 reform. The shaded area represents the 90% confidence interval. On the left-hand side, the focus is on the mathematics teachers before the reform, while on the right-hand side, the focus is on Spanish teachers. The numbers in the left-bottom corner refer to the coefficient of the linear fit.

While these results are in line with the findings by Gallegos et al. (2022), they challenge the idea that PSU scores can serve as an effective recruitment screening device to enhance students' learning outcomes across all subjects. Gallegos et al. (2022) use several measures of teacher quality, namely, graduation from college, college exit exams, government evaluation, employment, wages, and students' achievement gains. While the authors provide interesting insights on the correlation between PSU and these measures, our measure allows us to disentangle the role of sorting - between and within schools - in partially explaining this correlation. This is crucial when thinking of measuring teacher productivity as sorting might amplify the predictive power of pre-college achievement.[29]

Several studies have investigated the role of teachers' cognitive ability in explaining their performance, using TVA as an outcome. Its predictive power differs depending on the country studied and how cognitive ability is measured. Most papers have analyzed the role of teachers' contemporaneous content knowledge, as measured by standardized tests. Focusing on Peru, Metzler and Woessmann (2012) find that a one standard deviation increase in teachers' test scores, increases their students' math test scores by about 0.06, but they find no correlation for Spanish teachers. For lower income countries, Bau and Das (2020) and Bietenbeck et al. (2018) find that a one standard deviation increase in test scores is associated with roughly a 0.25-0.3 increase in value-added, regardless of the subject. When measuring teachers' cognitive ability using proxies different than content knowledge, correlations are lower. For example, Araujo et al. (2016) find that one standard deviation increase in teachers' IQ is associated with a 0.04 higher TVA across subjects. Papers studying the role of pre-college test scores, such as the one we study, are rare and never

[29]The authors also use measures of students' achievement gain. However, they are limited to a small number of teachers (around 100), which also limits the richness of potential controls in the estimation of the value-added model. Additionally, they only focus on mathematics.

related to value-added measures (Bardach and Klassen, 2020).

5 Beyond selection on academic ability

The change in the average performance of mathematics teachers cannot solely be explained by a compositional effect, that is, by the higher share of high PSU scoring teachers. The average teacher's test scores increased by 0.32 of a standard deviation, and given our estimates of the correlation between PSU and TVA, we would predict an increase of around 0.01 in value-added. Therefore, the role of PSU scores appears marginal in explaining the overall effect of the reform.

Table 5: Before-After comparison beyond academic ability - teacher quality

	Math teachers		Spanish teachers	
	(1)	(2)	(3)	(4)
Post-reform cohort	0.034**		-0.001	
	(0.016)		(0.015)	
Post-reform cohort × PSU < 500		0.065*		0.004
		(0.036)		(0.027)
Post-reform cohort × PSU $\in [500, 600)$		0.001		-0.014
		(0.021)		(0.019)
Post-reform cohort × PSU ≥ 600		0.065**		0.030
		(0.026)		(0.027)
# Students	206,094	206,094	225,584	225,584
# Teachers	4,386	4,386	4,966	4,966

Note: * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. The table shows OLS estimates for the coefficients of interest in model (14). We additionally control for a polynomial of order two of PSU score. Standard errors are clustered at the teacher-by-year level, and shown in parentheses.

A natural question that follows is where the remainder of the effect comes from. According to the mechanisms outlined in the theoretical framework in Section 2.3, we could expect teachers who responded to the reforms' incentives to be positively selected on value-added, conditional on academic ability.

To corroborate this intuition, we estimate model (14) including PSU scores as a control. Column (1) in Table 5 shows that even after adjusting for compositional effects, math teachers who studied after 2011 exhibit, on average, a higher TVA compared to those who studied before the reform. We then perform the same analysis, but interacting the post-reform dummy with each subgroup of PSU scores based on the restrictions and incentives imposed by the reform. Results are presented in column (2) in Table 5. We observe that math teachers scoring below 500 and above 600 are associated with a 0.065 higher value-added for post-reform cohorts.

We test the robustness of our results in the same spirit as we proceed for the average effect. Namely, we test whether they are sensitive to different value-added specifications, or to restricting the years of experience of teachers. Results, shown in Appendix Table B13, remain stable across specifications. Moreover, Appendix Figure A18 shows the effect not being driven by any specific cohort before 2011. The change in the relation between teacher quality and PSU scores is also illustrated in Figure 6, where we plot the non-parametric fit of the TVA estimates τ_j and PSU scores. For mathematics teachers, we observe an increase in the share of high-value-added teachers among those scoring below 500, no differences in teacher quality for those scoring between 500 and 600, and a steeper positive relationship for teachers scoring above 600. Consistent with our previous findings, the relationship between TVA and PSU is not altered for Spanish teachers.

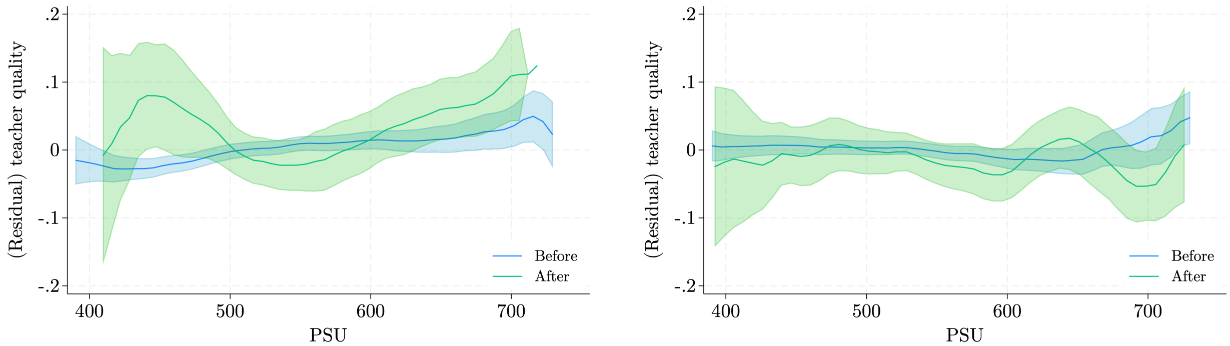
It is important to stress that finding no differences for teachers not directly targeted by the reform (those scoring between 500 and 600) has also implications for the assumption on our identification strategy, which relies on differences across cohorts. For math teachers, the absence of changes in teaching quality over the long pre-reform window is suggestive that the results are driven by the reform. As discussed in the previous section, the policy directly targeted teachers scoring below 500 or above 600. Since the mechanisms we discuss apply only to these teachers, the fact that the TVA does not change for those scoring between 500 and 600 after the reform aligns with the idea that the results are not confounded by other contemporaneous events or reforms.

Why are entrants and stayers better teachers? The framework described in Section 2.3 explains why we may observe higher average quality for both marginal entrants (PSU$\geq$ 600) and marginal remainers (PSU$<$ 500). Recalling that the decision to become a teacher is characterized by two opposite forces (i) the *outside pull* and (ii) the *motivation* channel, the framework would predict that, given the results we observe on value-added, at high levels of academic ability the former prevails, while at lower levels the latter is more relevant.

In other words, individuals with high academic ability face better outside options, which makes it more likely that high–value–added teachers, despite the correlation between motivation and teaching productivity, pursue alternative careers. The introduction of incentives draws in marginal entrants with higher value-added, as these teachers are recruited from individuals who would otherwise choose different career paths. By contrast, among those with lower academic ability, motivation plays a stronger role in career choice, leading those who enter teaching to be positively selected on value-added. As a result, the restrictions primarily exclude teachers with low value-added and low motivation.

Why do these two forces operate differently across levels of academic ability? A plausible explanation is a heterogeneous parameter r_O, i.e., how productivity in the non-teaching sector is valued. There are several reasons why r_O may differ by PSU. The first is related to the role of PSU itself: individuals with low scores may be less able to convert their skills into high-return outside opportunities because their choice set of higher education programs is more limited. Appendix Figure A20 documents a strong positive relationship between PSU scores and the number of programs available in the centralized admission system. Another explanation, supported by evidence in the literature, is that academic ability interacts with correlated characteristics, such as

Figure 6: Before and after local polynomial fit of TVA on PSU



(a) Math

(b) Spanish

Note: The figures show the non-parametric relationship between the residual TVA and PSU described by a local polynomial regression together with their linear fit, for teachers that started their studies before the reform (2004-2010) and after the reform (2011-2014). The shaded area represents the 90% confidence interval. The *(residual) teacher quality* measure is obtained as the residual from a non-parametric regression of TVA on years of experience. Panel (a) shows the relationship for mathematics teachers, while Panel (b) shows it for Spanish teachers.

parental resources, in the production function of human capital or wages (Carneiro et al., 2011)[30].

5.1 Competing explanations: the role of other inputs

We discuss potential confounders to our interpretation. In particular, we examine the role of teacher sorting across schools and higher education programs that may account for the observed effects, even after adjusting for PSU.

As discussed in Section 3, the reform altered the sorting of high school graduates into teaching training programs. Teachers studying with the BVP were more likely to pursue more prestigious higher education programs, whereas those scoring below 500 faced restrictions to access them. If teachers are trained more effectively in these programs, we might expect this to explain part of the estimated effect. Therefore, we re-estimate model (11) controlling for program and classmate characteristics. More precisely, indicators for whether the institution is a public or private CRUCH university, the average PSU score of the same-cohort enrollees in the program, the proportion of enrollees who took the PSU, and the program size.

Additionally, the reform changed the types of schools that some teachers sort into. This may influence their TVA in two ways. First, it may be more or less difficult to teach students from different backgrounds in order to improve their test scores. For instance, in Appendix Figure A19 we show how TVA is negatively correlated with the share of vulnerable students taught. Second, newly hired teachers may be exposed to different colleagues as a consequence of the reform's

[30]While it may seem natural to use an RD strategy to characterize who enters or remains in the teaching profession, such an approach is not feasible in this setting. The number of teachers close to the PSU cutoffs after 2011 is small, preventing any credible RD estimation. In addition, Figure 6 shows that the differences in TVA after the reform are larger for individuals who are further away from the cutoffs.

incentive to work in public and rural schools. In columns (5) to (7) of Appendix Table B14, we show results additionally adjusting for the vulnerability index of the school where the teacher works, the number of colleagues, their average PSU scores, the percentage of teachers without PSU scores, and their average years of experience.

The results indicate that including these factors leaves the effect on the TVA mostly unchanged. The change in TVA, conditional on PSU scores, decreases from 0.034 to 0.028 and remains statistically significant at the 10% confidence level. Therefore, a large share of the overall effect remains unexplained.

5.2 The role of motivation

From the theoretical framework described in Section 2.3, we can derive predictions about how the reform's incentives affected the average motivation among teachers. In particular, it is straightforward to show that individuals facing restrictions on entry into the teaching profession should exhibit higher levels of motivation if they are measured to be of high value-added. By contrast, for those incentivized to enter through the scholarship, the model yields no clear prediction regarding changes in motivation if higher value-added is measured among marginal entrants.[31]

To directly test this prediction, we would ideally need to complement our analysis with measures of motivation collected at the moment of enrollment. We proxy them with on-the-job survey data on a sub-sample of primary school math and Spanish teachers. These data are regularly collected by the Chilean Ministry of Education during the days in which the standardized test for their students takes place. While the teacher survey takes place every year, the specific questions asked change regularly. In 2018 and 2019, teachers were asked some questions about their satisfaction, dedication and motivation towards their role as teachers. Our main variable of interest is an indicator variable for teachers strongly agreeing to the statement *"If I could decide, I would choose this job again"*. Appendix Table B15 shows that our measure of motivation is positively correlated with teacher quality: strongly agreeing with the statement is associated with a 0.033 increase in TVA.

Table 6 presents the average change in motivation after the reform, controlling for a second-order polynomial of PSU, as well as years of experience, year, gender, and additional school covariates. School covariates adjust for the possible effect of sorting happening after graduation. Column (1) in Panel A shows that math teachers enrolling before and after the reform are not differentially likely to strongly agree with the statement. However, column (2) reveals some heterogeneity across PSU scores, as it presents evidence of positive selection based on motivation for teachers who scored below 500 points: those enrolling after the reform are almost 21 p.p (36%) more likely to strongly agree with the statement. For teachers scoring above 600, we do not find evidence of negative selection on motivation induced by the reform.

[31] See Appendix D for a simple proof.

Table 6: Before-After comparison beyond academic ability - motivation

	Math teachers		Spanish teachers	
	(1)	(2)	(3)	(4)
Post-reform cohort	0.028		0.009	
	(0.045)		(0.043)	
Post-reform cohort $\times$ PSU < 500		0.207***		-0.009
		(0.074)		(0.067)
Post-reform cohort $\times$ PSU $\in [500, 600)$		0.010		0.005
		(0.065)		(0.057)
Post-reform cohort $\times$ PSU ≥ 600		-0.070		0.055
		(0.071)		(0.079)
Observations	1079	1079	1210	1210
Baseline mean	0.573		0.572	

Note: * $p < 0.10$,** $p < 0.05$, *** $p < 0.01$. The table shows OLS estimates for the coefficients of a post-reform cohort indicator, in a model that additionally controls for a polynomial of order two on (standard) PSU score, as well as fixed effects for years of experience, gender, year, public, private, and rural schools. Robust standard errors are reported in parentheses.

6 Conclusion

This paper examines the impact of a Chilean educational reform that sought to improve the quality of teachers by changing the composition of students who enrolled in higher education teacher training programs. The reform used the national standardized university entrance exam (PSU) to set entry standards for teacher education programs. It limited the admission of students who scored below the mean while offering scholarships to incentivize the enrollment of those who scored more than one standard deviation above the mean.

Our analysis shows that this reform was successful in attracting new teachers from the top of the PSU score distribution and effectively prevented lower-scoring individuals from entering the teaching profession. To test how this translated into teaching quality, we focus on mathematics and Spanish primary school teachers and estimate the effect based on a Teacher Value-Added (TVA) framework using students' test scores. We show that the reform increased the average teacher quality for mathematics teachers, but not for Spanish teachers. Moreover, we estimate that the change in academic ability explains only about 10% of the positive effect of the reform on TVA.

We furthermore show that after the reform, both graduates who were offered the scholarship and those who experienced admissions restrictions are associated with higher TVA relative to similar teachers from earlier cohorts. We rationalize these findings through the lens of an occupational sorting model à la Roy (1951), that predicts that scholarships can draw in high-ability individuals who would otherwise sort into outside options, while restrictions can screen out candidates with low motivation.

Although successful, most of the positive effects of the reform were unintended. The selection of graduates on the basis of a single observable dimension, namely college admission test scores, played only a minor role. Our results underscore that recruitment policies targeting an observable trait can nonetheless generate large improvements when they also shift unobservable attributes.

References

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., and Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics*, 131(3):1415–1453.

Araujo, M. D., Heineck, G., and Cruz-Aguayo, Y. (2020). Does test-based teacher recruitment work in the developing world? Experimental evidence from Ecuador. *IZA Discussion Paper*.

Ashraf, N., Bandiera, O., Davenport, E., and Lee, S. S. (2020). Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services. *American Economic Review*, 110(5):1355–1394.

Bacher-Hicks, A., Chin, M. J., Kane, T. J., and Staiger, D. O. (2019). An experimental evaluation of three teacher quality measures: Value-added, classroom observations, and student surveys. *Economics of Education Review*, 73:101919.

Bacher-Hicks, K. (2022). Estimation and interpretation of teacher value added in research applications. *Handbook of the Economics of Education 6, 93-134*.

Bacolod, M. 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(4):737–751.

Bardach, L. and Klassen, R. M. (2020). Smart teachers, successful students? a systematic review of the literature on teachers' cognitive abilities and teacher effectiveness. *Educational Research Review*, 30:100312.

Bau, N. and Das, J. (2020). Teacher value added in a low-income country. *American Economic Journal: Economic Policy*, 12(1):62–96.

Behrman, J. R., Tincani, M. M., Todd, P. E., and Wolpin, K. I. (2016). Teacher quality in public and private schools under a voucher system: The case of Chile. *Journal of Labor Economics*, 34(2):319–362.

Besley, T. and Ghatak, M. (2005). Competition and incentives with motivated agents. *American Economic Review*, 95(3):616–636.

Biasi, B. (2021). The labor market for teachers under different pay schemes. *American Economic Journal: Economic Policy*, 13(3):63–102.

Bietenbeck, J., Piopiunik, M., and Wiederhold, S. (2018). Africa's skill tragedy: Does teachers' lack of knowledge lead to low student performance? *Journal of Human Resources*, 53(3):553–578.

Borjas, G. (1987). Self-selection and the earnings of immigrants. *American Economic Review*, 77(4):531–53.

Bucarey, A., Contreras, D., and Muñoz, P. (2020). Labor market returns to student loans for university: Evidence from Chile. *Journal of Labor Economics*, 38(4):959–1007.

Buhl-Wiggers, J., Kerwin, J. T., Smith, J., and Thornton, R. (2022). Learning More about Teachers: Estimating Teacher Value-Added and Treatment Effects on Teacher Value-Added in Northern Uganda. Working Paper.

Busso, M., Montaño, S., Muñoz-Morales, J., and Pope, N. G. (2024). The unintended consequences of merit-based teacher selection: Evidence from a large-scale reform in Colombia. *Journal of Public Economics*, 239:105238.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2019). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2):192–210.

Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating marginal returns to education. *American Economic Review*, 101(6):2754–2781.

Castro-Zarzur, R., Espinoza, R., and Sarzosa, M. (2022). Unintended consequences of free college: Self-selection into the teaching profession. *Economics of Education Review*, 89(102260):10.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–79.

Cohen, J. and Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, 45(6):378–387.

Cornejo, A. (2005). SINAE Sistema Nacional de Asignación con Equidad para Becas JUNAEB Una nueva visión en la construcción de igualdad de oportunidades en la infancia. Technical report, Junta Nacional de Jardines Infantiles (JUNJI).

Dal Bó, E., Finan, F., and 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 Falco, A. and Reichlin, Y. (2025). *Grants vs. loans: the role of financial aid in college major choice*. SSRN working paper available at https://ssrn.com/abstract=4581235 or http://dx.doi.org/10.2139/ssrn.4581235.

De Gregorio, S. and Neilson, C. (2021). Alternative pathways into teaching and student learning in chile.

De Ree, J., Muralidharan, K., Pradhan, M., and Rogers, H. (2018). Double for nothing? Experimental evidence on an unconditional teacher salary increase in Indonesia. *The Quarterly Journal of Economics*, 133(2):993–1039.

Deserranno, E. (2019). Financial incentives as signals: experimental evidence from the recruitment of village promoters in Uganda. *American Economic Journal: Applied Economics*, 11(1):277–317.

Duflo, E., Hanna, R., and Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4):1241–78.

Estrada, R. (2019). Rules versus discretion in public service: Teacher hiring in Mexico. *Journal of Labor Economics*, 37(2):545–579.

Gallegos, S., Neilson, C., Calle, F., and Karnani, M. (2022). *Screening and Recruiting Talent At Teacher Colleges Using Pre-College Academic Achievement*. Princeton University, Industrial Relations Section.

González, F. and Prem, M. (2024). Police violence, student protests, and educational performance. *Review of Economics and Statistics*, 106(3):712–727.

Herrmann, M., Walsh, E., and Isenberg, E. (2016). Shrinkage of value-added estimates and characteristics of students with hard-to-predict achievement levels. *Statistics and Public Policy*, 3(1):1–10.

Jacob, B. A., Rockoff, J. E., Taylor, E. S., Lindy, B., and Rosen, R. (2018). Teacher applicant hiring and teacher performance: Evidence from dc public schools. *Journal of Public Economics*, 166:81–97.

Kane, T. J. and Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Technical report, National Bureau of Economic Research.

Kapor, A., Karnani, M., and Neilson, C. (2024). Aftermarket frictions and the cost of off-platform options in centralized assignment mechanisms. *Journal of Political Economy*, 132(7):2346–2395.

Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, 99(5):1979–2011.

Leaver, C., Ozier, O., Serneels, P., and Zeitlin, A. (2021). Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools. *American Economic Review*, 111(7):2213–2246.

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

Metzler, J. and Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics*, 99(2):486–496.

OECD (2010). *Education at a glance 2010: OECD indicators*. OECD Paris.

Pal, J. (2025). Optimal policy design for teacher recruitment. *Available at SSRN 5373707*.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2):247–252.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146.

Solis, A. (2017). Credit access and college enrollment. *Journal of Political Economy*, 125(2):562–622.

Staiger, D. O. and Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, 24(3):97–118.

Steinberg, M. P. and Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis*, 38(2):293–317.

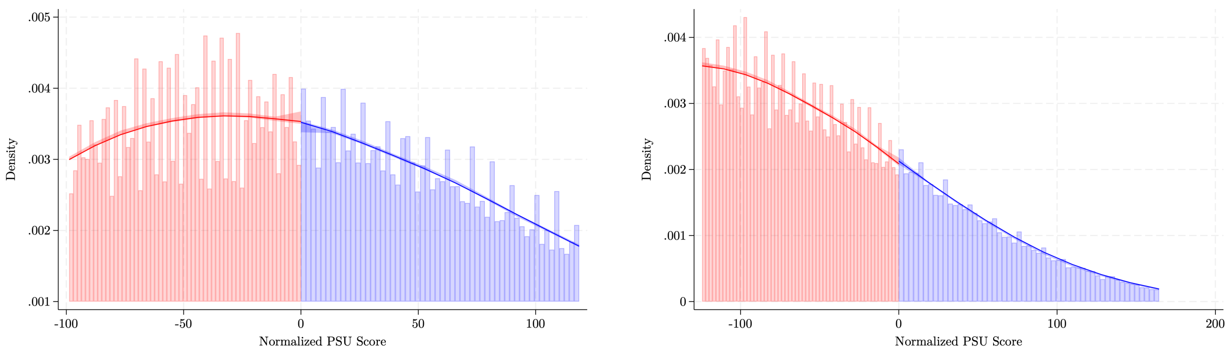
Tincani, M. M. (2021). Teacher labor markets, school vouchers, and student cognitive achievement: Evidence from chile. *Quantitative Economics*, 12(1):173–216.

World Bank (2017). *World development report 2018: Learning to realize education's promise*. The World Bank.

Recruiting Better Teachers?
Evidence from a Higher Education Reform in Chile
Supplemental Appendices

A Additional Figures

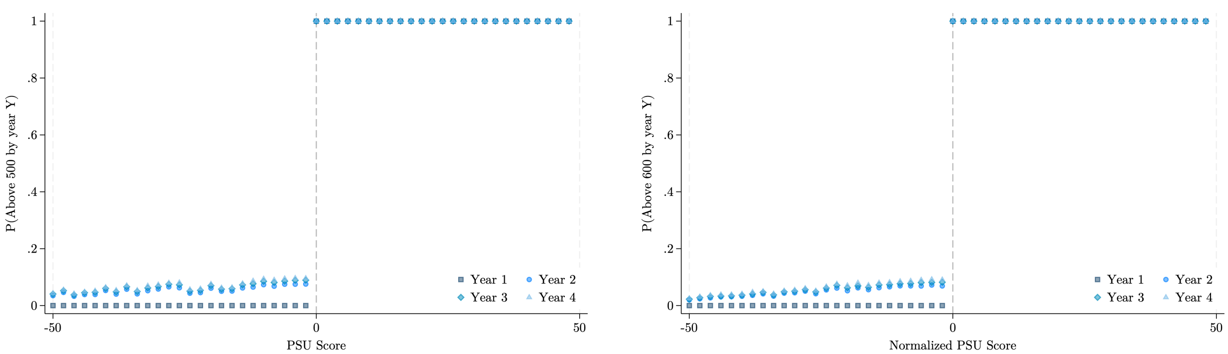
Figure A1: McCrary test



(a) 500 cutoff (b) 600 cutoff

Note: The figures display the McCrary test for discontinuity in the running variable for the 500 cutoff (a) and 600 cutoff (b).

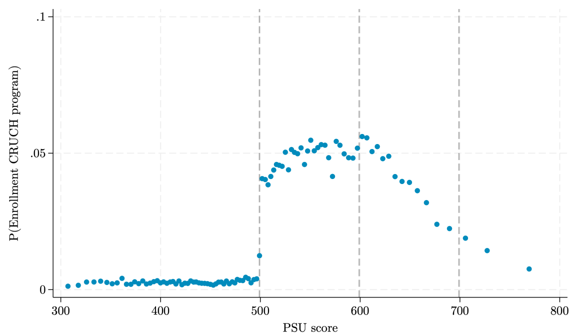
Figure A2: $P(Eligible)$ by year y as a function of normalized PSU score



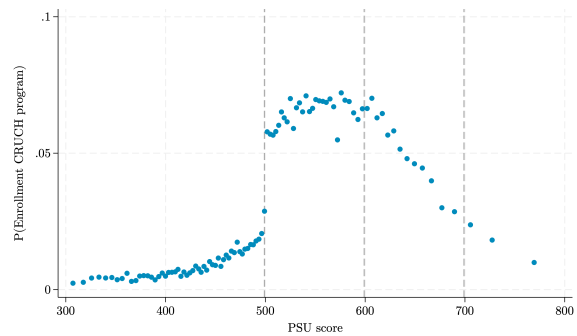
(a) 500 cutoff (b) 600 cutoff

Note: The figure presents shares of individuals eligible to enroll in an eligible institution (Panel (a)) and for the BVP (Panel (b)) and as a function of the normalized PSU score. Shares are shown in a 2-point bin.

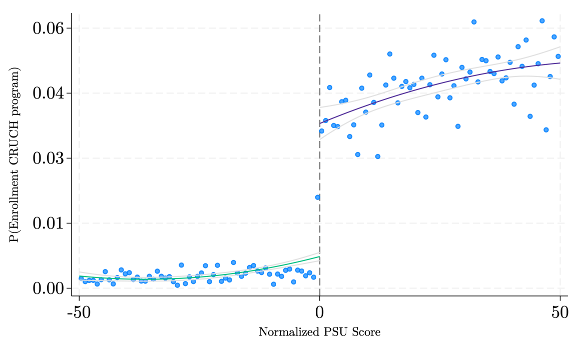
Figure A3: Enrollment in CRUCH teacher training program as a function of the PSU



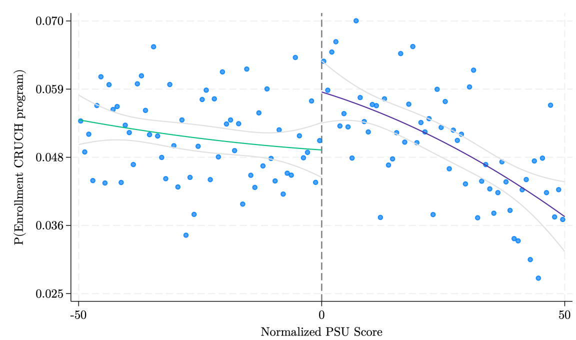
(a) Year 1



(b) Year 2



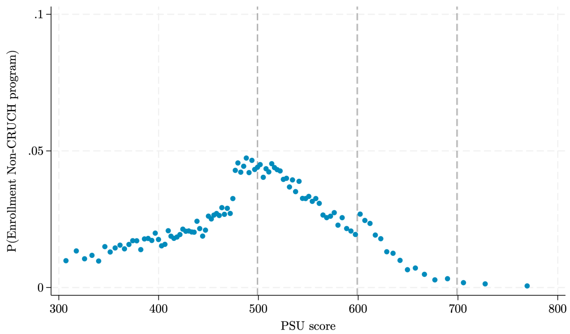
(c) 500 cutoff



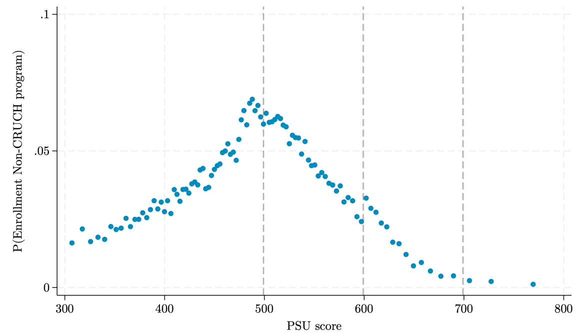
(d) 600 cutoff

Note: The figures show probabilities of enrollment in teacher training undergraduate programs in CRUCH institutions as a function of the PSU score. CRUCH institutions are the most prestigious, and almost all of them are *eligible* institutions in which to use the BVP. Figure (a) shows enrollment in the same year of the first PSU attempt; Figure (b) shows enrollment by the second year; Figures (c) and (d) show the probability of enrollment in the year of the first PSU attempt at each cutoff.

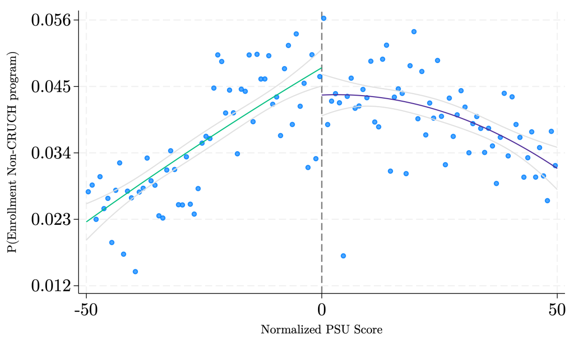
Figure A4: Enrollment in Non-CRUCH teacher training program as a function of the PSU



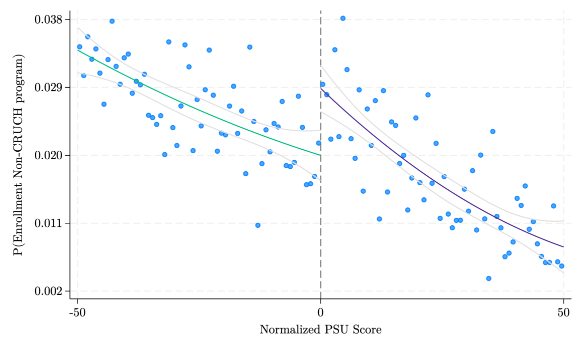
(a) Year 1



(b) Year 2



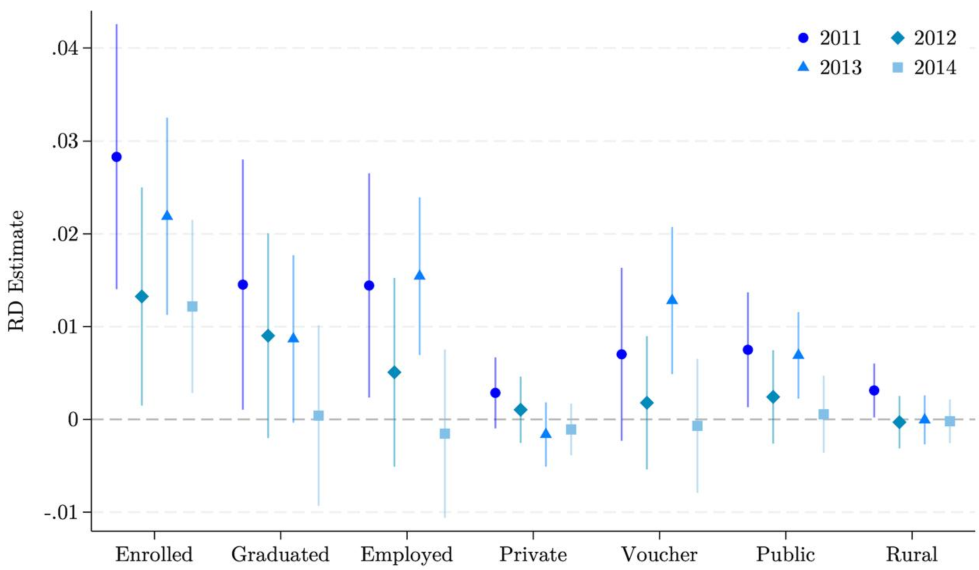
(c) 500 cutoff



(d) 600 cutoff

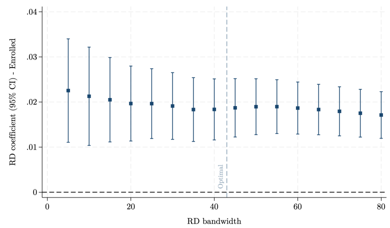
Note: The figures show probabilities of enrollment in teacher training undergraduate programs in non-CRUCH institutions as a function of the PSU score. Figure (a) shows enrollment in the same year of the first PSU attempt; Figure (b) shows enrollment by the second year; Figures (c) and (d) show the probability of enrollment in the year of the first PSU attempt at each cutoff.

Figure A5: RD estimates at the 600 cutoff by cohort of study

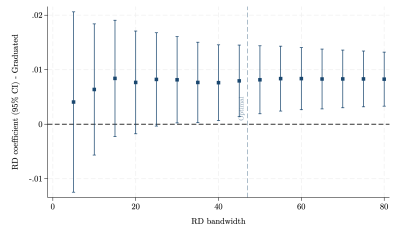


Note: The figure shows RD estimates from model (8) at the 600 cutoff, by cohorts of study. The outcomes are the same analyzed in Table 1.

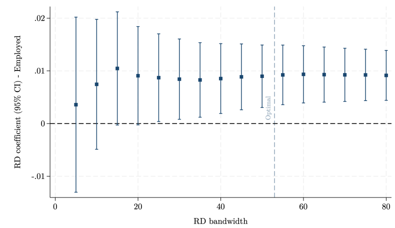
Figure A6: Sensitivity of RD estimates to different bandwidths



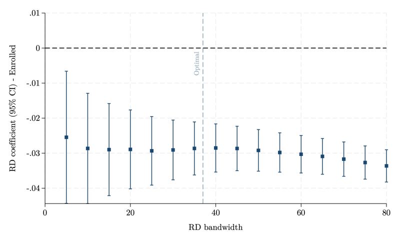
(a) Enrolled
(600 cutoff)



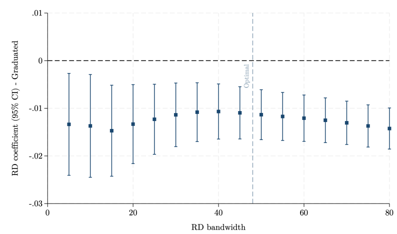
(b) Graduated
(600 cutoff)



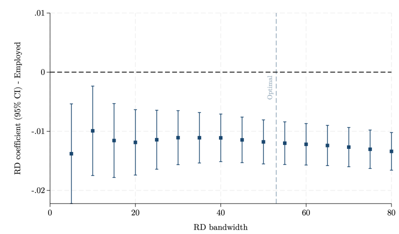
(c) Employed
(600 cutoff)



(d) Enrolled
(500 cutoff)



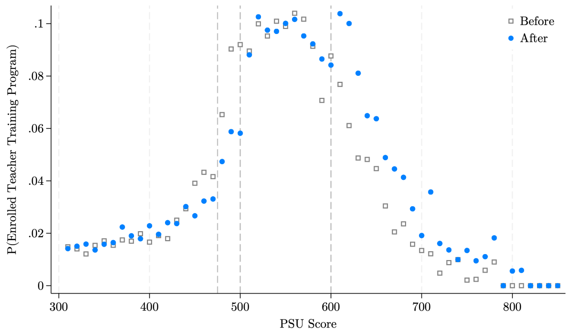
(e) Graduated
(500 cutoff)



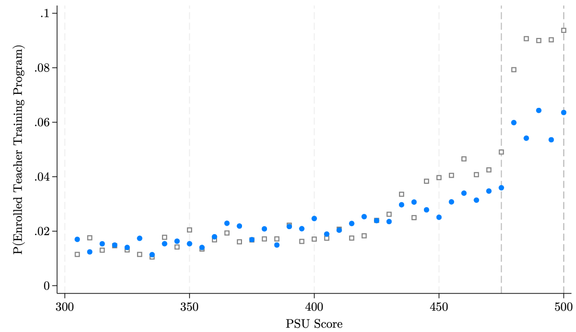
(f) Employed
(500 cutoff)

Note: The figure shows the sensitivity of our main RD estimates to different bandwidths. The estimation was performed using a triangular kernel and a polynomial of order 1. Standard errors are clustered at the PSU score level.

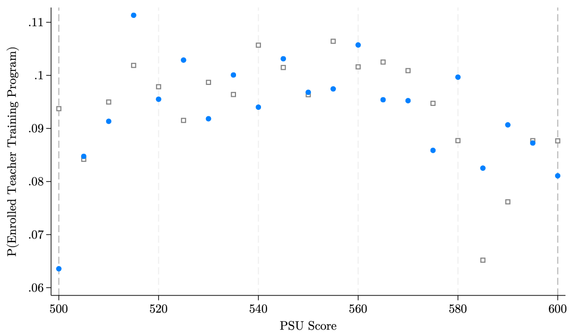
Figure A7: Enrollment in a teacher training program as a function of the PSU Score



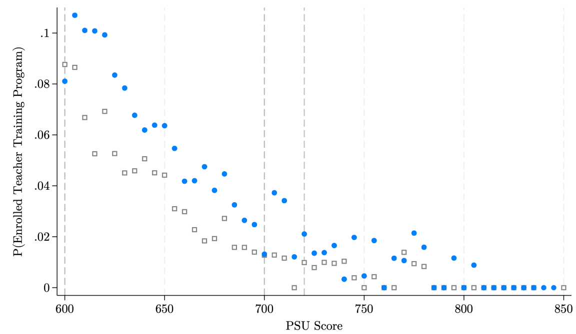
(a) Full distribution



(b) Affected by Restrictions



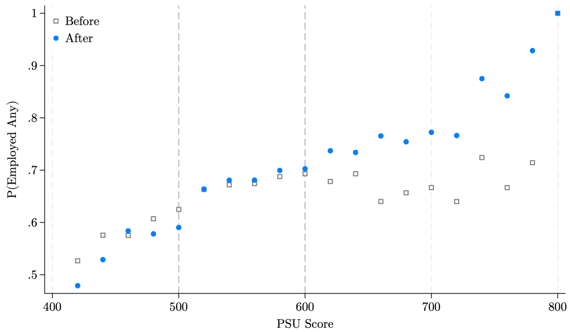
(c) Not Directly Affected



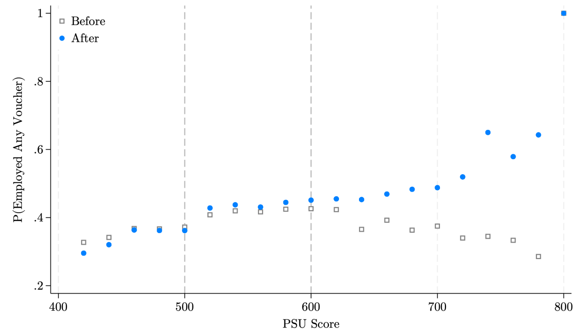
(d) Grant Eligible

Note: The figure shows the probability of enrolling in a teacher training program one year after high school graduation, as a function of the PSU score, before and after the reform. Panel (a) plots the entire PSU distribution, Panels (b) to (d) zoom in on the specific groups that were differently affected by the reform. The gray squares represent teachers who took the PSU for the first time before the reform was implemented, and the blue dots represent those who took the PSU after the reform was implemented. The two horizontal lines represent several relevant cutoffs, namely the loan eligibility cutoff (475), the cutoff preventing students from enrolling in BVP-eligible teacher training programs (500), the cutoff defining the eligibility for the BVP scholarship (600), and the two cutoffs for additional BVP benefits (700 and 720).

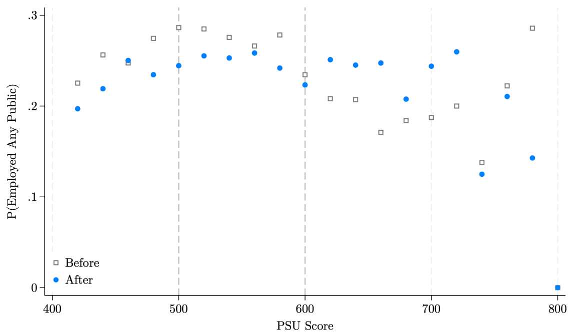
Figure A8: Employment in publicly funded schools as a function of the PSU Score



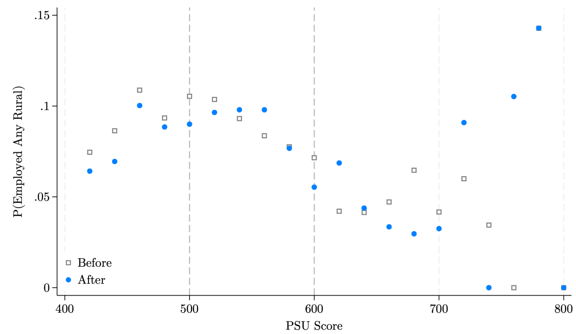
(a) Any School



(b) Voucher Schools



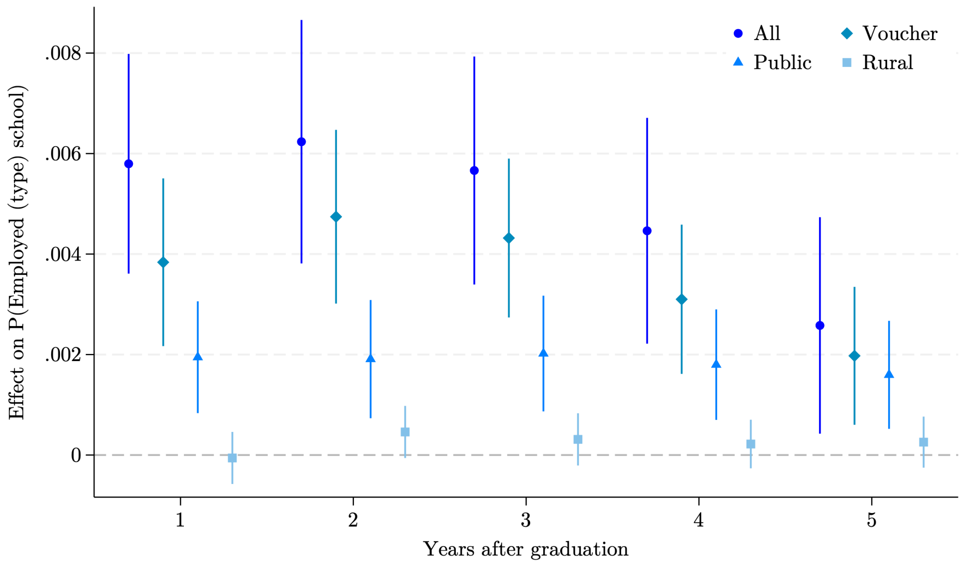
(c) Public School



(d) Rural Schools

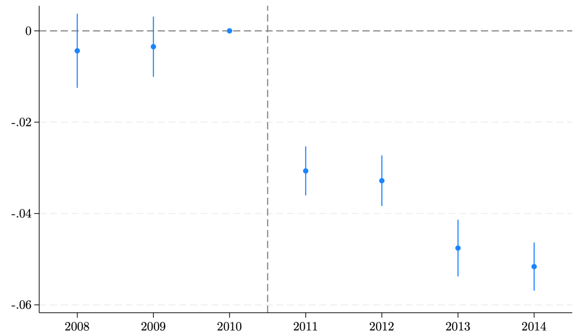
Note: The figure shows the probability of being employed in different types of schools in the first two years after graduation as a function of the PSU score. Panel (a) plots the probability of being employed in any school, Panel (b) plots the probability of being employed in a voucher school, Panel (c) plots the probability of being employed in a public school, and Panel (d) plots the probability of being employed in a rural school. The gray squares represent teachers who took the PSU for the first time before the reform was implemented, and the blue dots represent those who took the PSU after the reform was implemented. The two horizontal lines represent the cutoff preventing students from enrolling in BVP-eligible teacher training programs (500) and the cutoff defining the eligibility for the BVP scholarship (600).

Figure A9: BVP Eligibility and employment at school - Before-After analysis

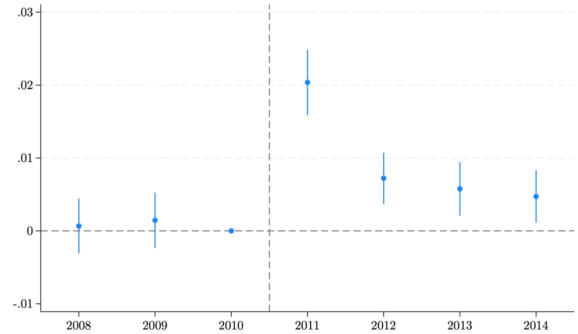


Note: The figure shows before-after difference for students scoring above 600 using as outcomes the probability of being employed in a school, and the probability of being employed by type of school up to five years after graduation. The sample is restricted to those taking the college admissions test between 2009 and 2014.

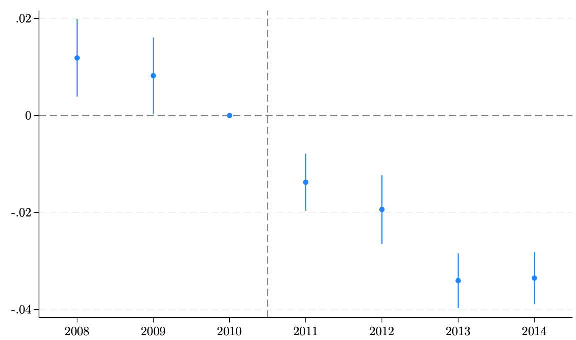
Figure A10: Trends on recruitment variables



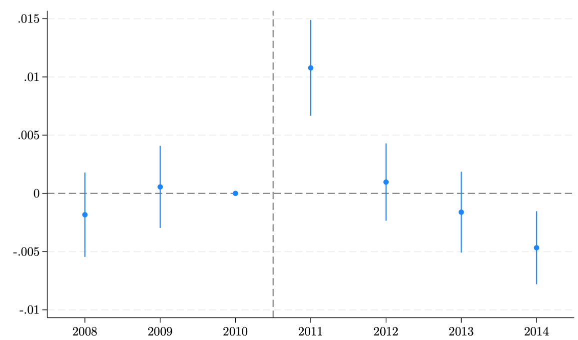
(a) Enrollment - below 500



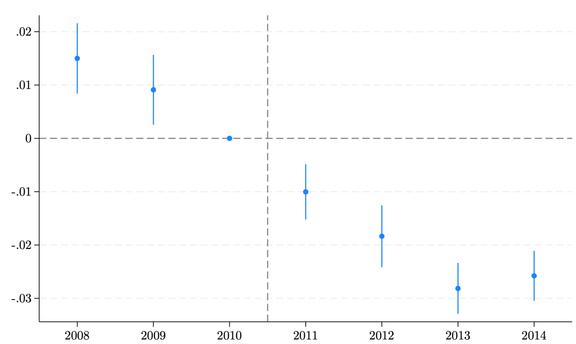
(b) Enrollment - above 600



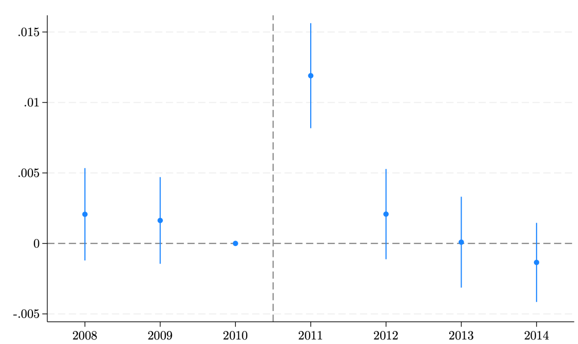
(c) Graduation - below 500



(d) Graduation - above 600



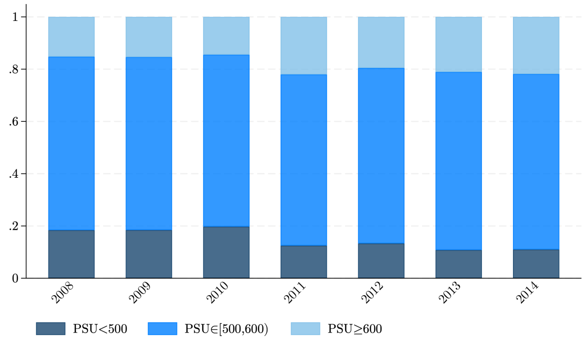
(e) Employment - below 500



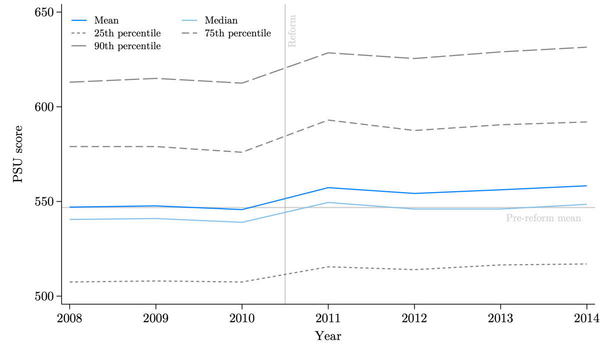
(f) Employment - above 600

Note: The figure presents the trends on our main recruitment variables of interest. On the left, the focus is on PSU takers scoring between 475 and 500, while on the right, on those scoring above 600. *Enrollment* is an indicator for individuals enrolling in a teacher training program on the year after high school graduation. *Graduation* is an indicator for individuals graduating from a teacher training program within eight years of high school graduation. *Employment* is indicator for individuals observed working in any school within 2 years of their graduation from a teacher training program.

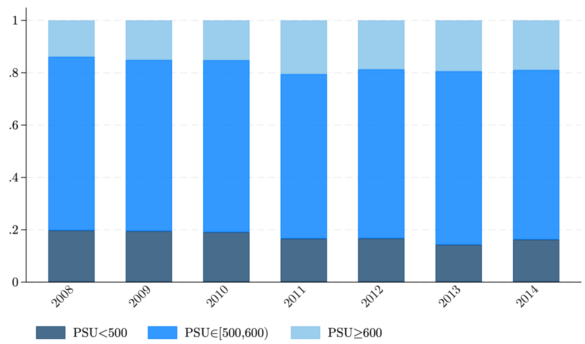
Figure A11: Trends in ability composition of enrollment and employment



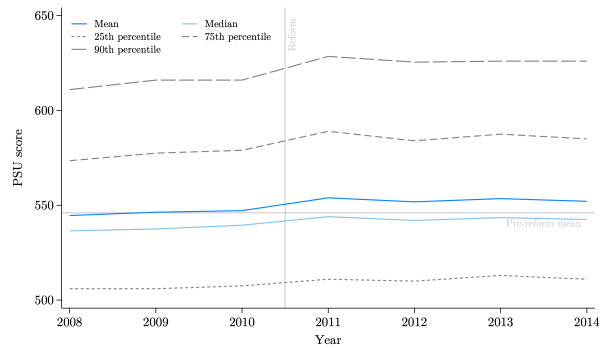
(a) Share of enrollees by score



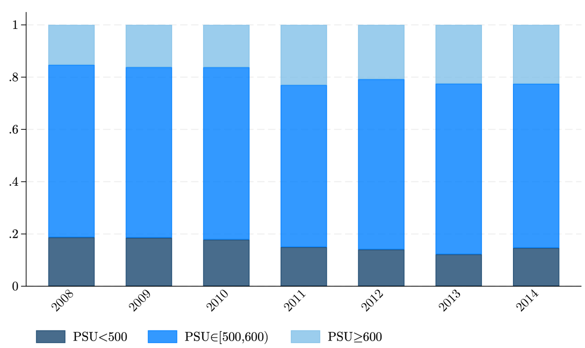
(b) Score statistics - Enrollment



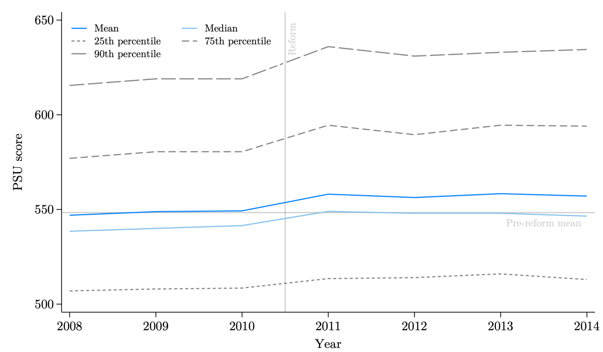
(c) Share of graduates by score



(d) Score statistics - Graduation



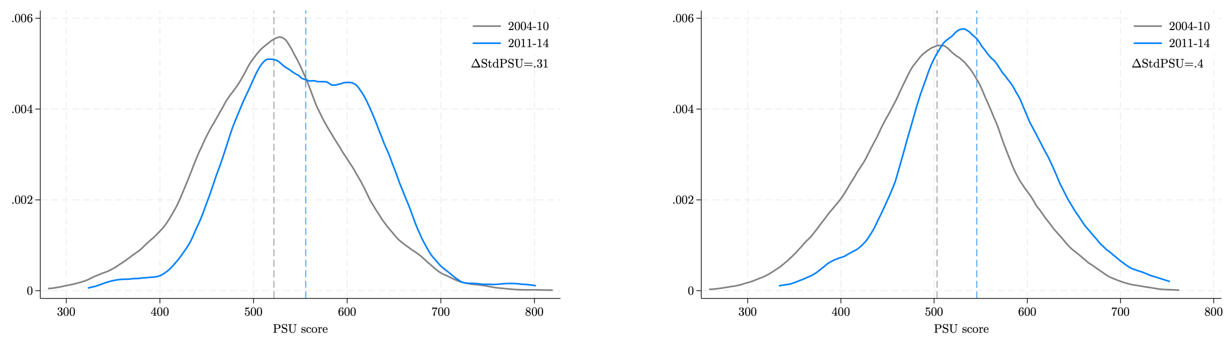
(e) Share of employed teachers by score



(f) Score statistics - Employment

Note: The figures show the trends in PSU statistics by cohort of higher education enrollment. The two plots on the top focus on enrollment in teacher training programs. The two plots on the middle focus on graduation after 8 years from a teacher training programs. The two plots on the bottom focus on teachers employed in schools 2 years after graduation. All results are conditional on scoring above 475 points.

Figure A12: Density of PSU of teachers before and after the reform

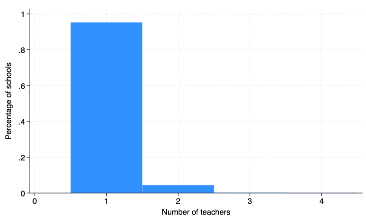


(a) Math Teachers

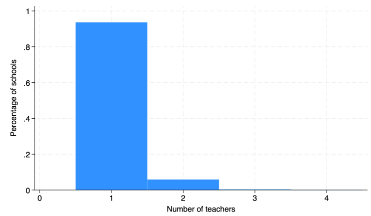
(b) Spanish Teachers

Note: The figures show the density of mathematics in Panel (a) and Spanish teachers in Panel (b) who enrolled in higher education before and after the reform and for whom we can reliably compute TVA. The dotted lines represent the mean PSU score for each group.

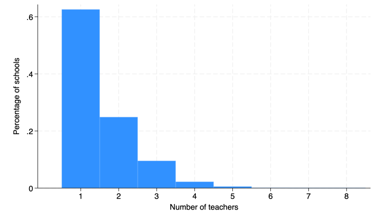
Figure A13: Variation for TVA Variance



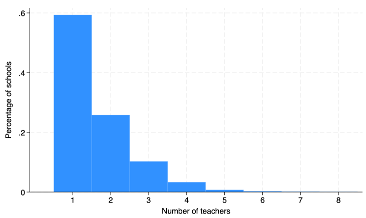
(a) # of teachers per school-year - Math



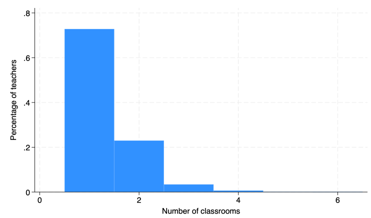
(b) # of teachers per school-year - Spanish



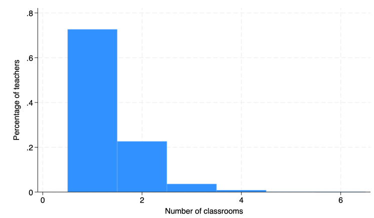
(c) # teachers per school - Math



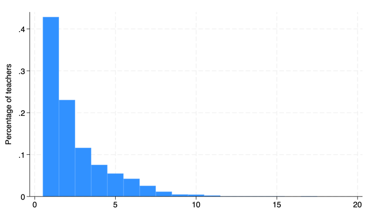
(d) # teachers per school - Spanish



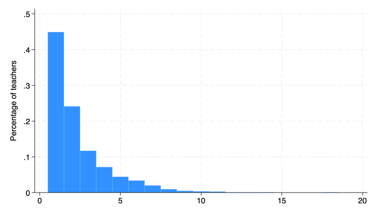
(e) # of classes per teacher-year - Math



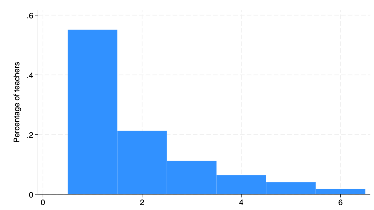
(f) # of classes per teacher-year - Spanish



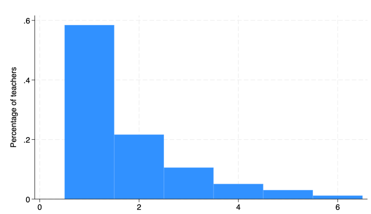
(g) # number of classes per teacher - Math



(h) # number of classes per teacher - Spanish



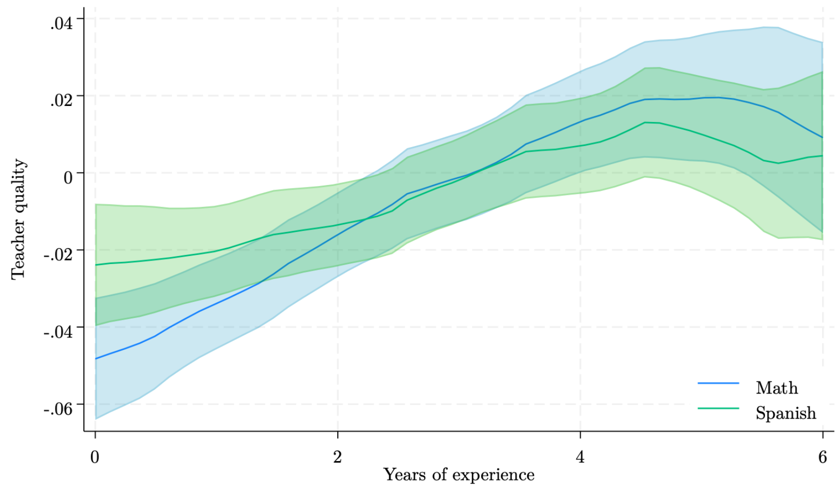
(i) # years by teacher



(j) # years by teacher

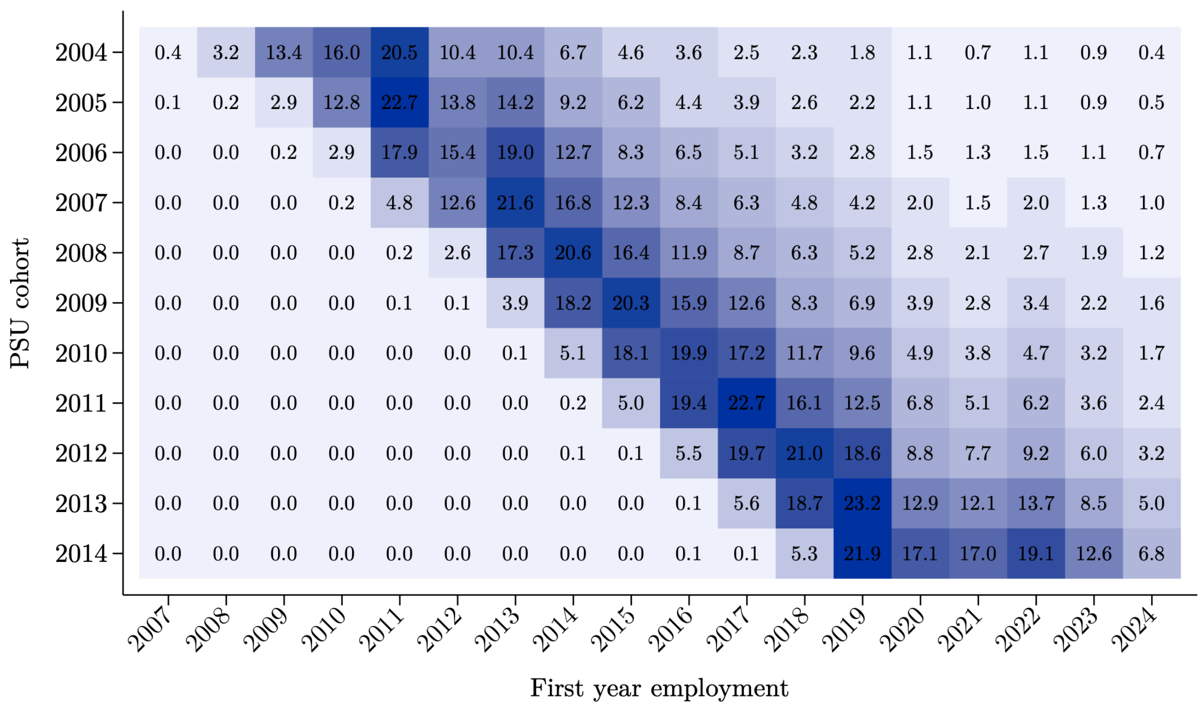
Note: Focusing on the teachers for which we can estimate class fixed effects, the figures show frequency plots by subjects. From the first to the last row, they show frequencies of teachers per school-year ((a) and (b)), and per school ((c) and (d)), frequencies of classes per teacher-year ((e) and (f)) and per school ((g) and (h), and frequencies of years ((i) and (j)).

Figure A14: TVA estimates and experience



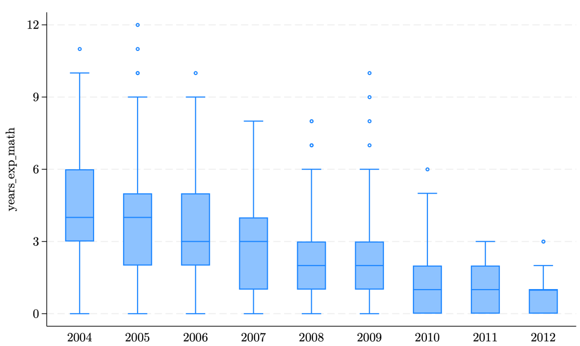
Note: The figures show the non-parametric relationship between TVA estimates and average year of experience, for mathematics and Spanish teachers.

Figure A15: Variation in the year of entry to the labor market within cohort

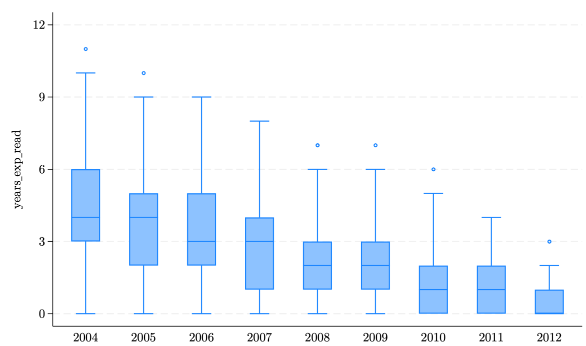


Note: The figure shows the frequency with which individuals of each cohort entered the labor market in each particular year. For illustrative purposes, cells with higher frequencies are colored in darker shades of blue.

Figure A16: Years of experience by PSU Cohort



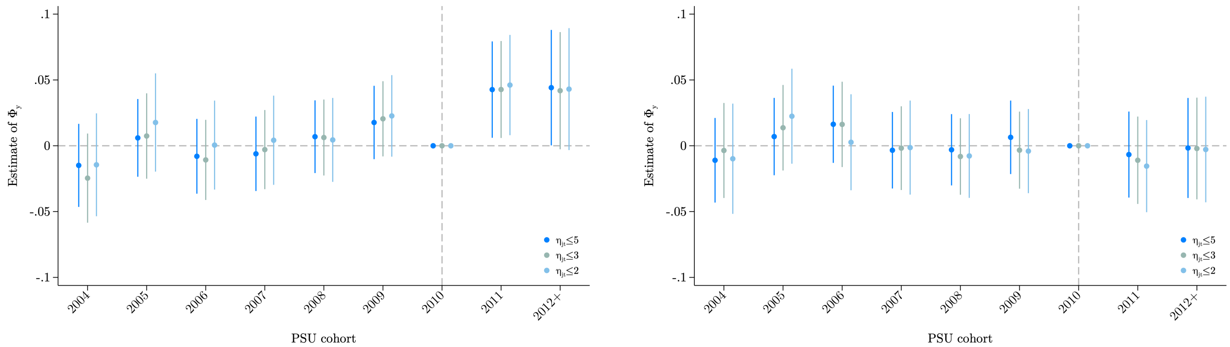
(a) Mathematics



(b) Spanish

Note: The figures show the distribution of years of experience by cohorts of study. Figure (a) refers to Mathematics teachers, while Figure (b) refers to Spanish teachers.

Figure A17: Average change in TVA by cohort - Robustness check by years of experience

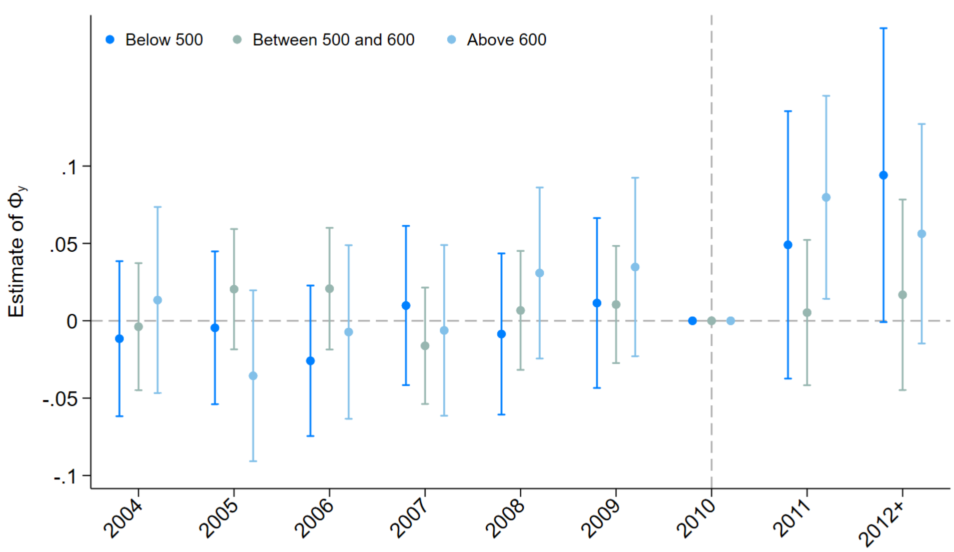


(a) Math

(b) Spanish

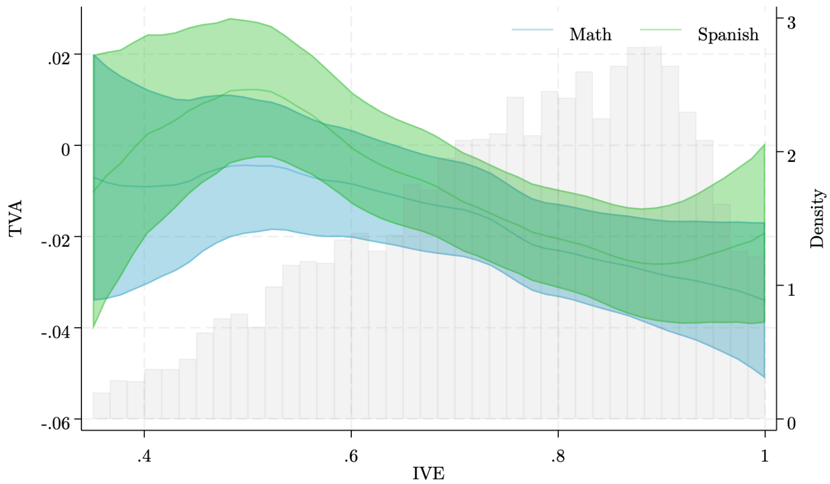
Note: The figures show OLS estimates for the coefficients Φ_y from (13). Confidence intervals at the 10% level are displayed and computed using standard errors clustered at the teacher-by-year level. On the left panel, the focus is on math teachers, while on the right panel is on Spanish. The figures present three sets of coefficients, estimated by restricting the sample to teachers with less than five, three, and two years of experience.

Figure A18: Average change in TVA by cohort - Results by PSU score



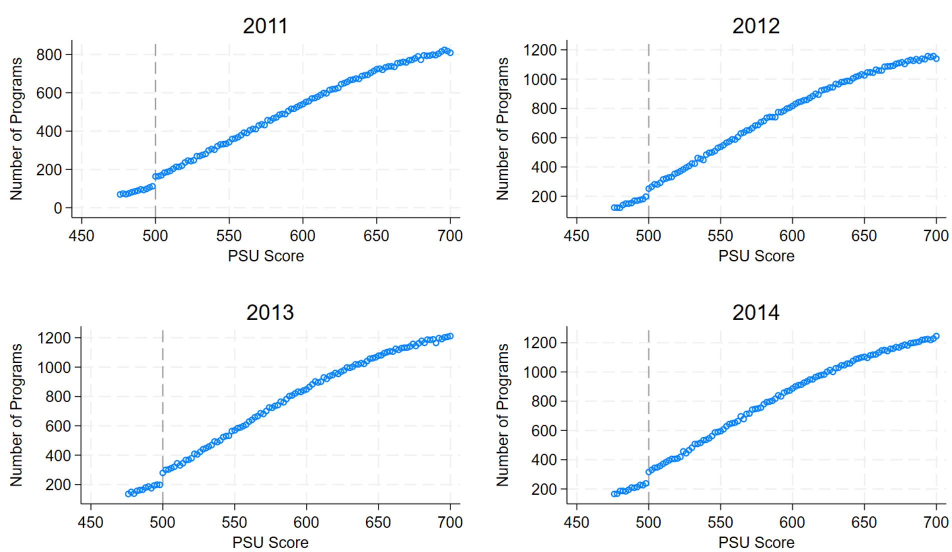
Note: The figure shows OLS estimates for the coefficients Φ_y from (13) for math teachers. Confidence intervals at the 10% level are displayed and computed using standard errors clustered at the teacher-by-year level. The figures present three sets of coefficients, estimated by restricting the sample to teachers with PSU below 500, between 500 and 600, and above 600.

Figure A19: TVA estimates and IVE-SINAE



Note: The figures show the non-parametric relationship between TVA estimates and average IVE-SINAE of the schools in which teachers work. The blue figure represents the relationship for math teachers and the green figure for Spanish teachers. The gray bars represent the density distribution of the IVE-SINAE index in our sample of teachers.

Figure A20: Number of available programs in the CAS as a function of PSU



Note: The figure shows the size of the choice set of PSU takers, as a function of the PSU score. The choice set is defined as the number of higher education programs within the centralized admission system where a PSU taker can enroll.

B Additional Tables

Table B1: Access to financial aid based on income and cohort just below the 600 Cut-off

	2011	*2012*	*2013*	*2014*
Family Income				
Quintile 1	*BIC*	*BIC + JGM*	*BIC + JGM*	*BIC + JGM*
Quintile 2	*BIC*	*BIC + JGM*	*BIC + JGM*	*BIC + JGM*
Quintile 3	FSCU + CAE	*BIC + JGM*	*BIC + JGM*	*BIC + JGM*
Quintile 4	*FSCU + CAE*	*FSCU + CAE*	*FSCU + CAE*	*FSCU + CAE*
Quintile 5	N.E.	N.E.	N.E.	N.E.

Note: Displayed are the eligibility criteria in terms of family income and the cohort of study for scholarships and loans. Scholarships refer to the Bicentennial (BIC) and Juan Gòmez Millas (JGM), which can be used in CRUCH and non-CRUCH institutions, respectively. Juan Gòmez Millas was introduced in 2012. Loans refer to the FSCU (*Fondo Solidario de Crédito Universitario*) and CAE (*Crédito con Aval del Estado*) non-CRUCH, which can be used in CRUCH and non-CRUCH institutions, respectively. Note that the vast majority of students enrolled in CRUCH institutions rely on the grant *Bicentenario* (BIC) and the loan *Fondo Solidario de Crédito Universitario* (FSCU), covering up to 90% of tuition in CRUCH institutions. In non-CRUCH institutions, students applying for financial aid mostly rely on the loan *Crédito con Aval del Estado* (CAE), and on the scholarship *Juan Gómez Millas*, which was introduced in 2012. The former covers up to 90% of fees, while the latter is a fixed amount equal to two-thirds of the median tuition.

Table B2: Summary statistics PSU Takers

	Mean	SD	Min	Max	N
Demographic					
High School GPA	5.597	0.493	4	7	677,275
Female	0.529	0.499	0	1	683,633
Father High School	0.546	0.498	0	1	683,633
Mother High School	0.600	0.490	0	1	683,633
Any Parent College	0.379	0.485	0	1	564,107
# Family Members	4.561	1.608	1	30	646,982
# Family Members Working	1.232	0.782	0	30	683,633
Single Parent	0.166	0.372	0	1	640,067
Father Working	0.654	0.476	0	1	587,549
Mother Working	0.389	0.488	0	1	625,760
Applied Financial Aid	0.695	0.461	0	1	683,633
Income Quintile 1	0.354	0.478	0	1	474,858
Income Quintile 2	0.243	0.429	0	1	474,858
Income Quintile 3	0.166	0.372	0	1	474,858
Income Quintile 4	0.134	0.341	0	1	474,858
Income Quintile 5	0.103	0.304	0	1	474,858
Enrollment					
Enrolled HEI	0.611	0.488	0	1	683,633
University	0.388	0.487	0	1	683,633
Traditional university	0.204	0.403	0	1	683,633
Enrolled	0.045	0.208	0	1	683,633
BEA Eligibility	0.076	0.266	0	1	683,633
BVP by year 1	0.007	0.084	0	1	683,633
BVP by year 4	0.011	0.105	0	1	683,633

Note: The table shows summary statistics of demographic and variables related to enrollment decisions and relevant aid take-up. The sample consists of the universe of PSU takers from 2011 to 2014.

Table B3: Balance check

	Baseline Mean	**PSU≥600**	**Baseline Mean**	**PSU≥500**
High School GPA	5.899	-0.008	5.547	0.005
	(0.004)	(0.006)	(0.003)	(0.004)
Female	0.503	0.001	0.534	0.001
	(0.004)	(0.005)	(0.004)	(0.006)
Father High School	0.712	0.004	0.564	-0.005
	(0.004)	(0.005)	(0.002)	(0.004)
Mother High School	0.781	0.009	0.625	0.002
	(0.004)	(0.005)	(0.003)	(0.004)
Any Parent College	0.635	0.002	0.345	-0.008
	(0.006)	(0.008)	(0.004)	(0.005)
# Family Members	4.444	-0.007	4.500	-0.006
	(0.011)	(0.016)	(0.010)	(0.014)
# Family Members Working	1.212	0.016	1.219	0.003
	(0.006)	(0.009)	(0.004)	(0.006)
Single Parent	0.131	0.005	0.176	0.003
	(0.003)	(0.004)	(0.003)	(0.004)
Father Working	0.718	0.001	0.645	0.004
	(0.005)	(0.006)	(0.003)	(0.004)
Mother Working	0.471	0.001	0.384	0.003
	(0.004)	(0.006)	(0.002)	(0.004)
Applied Financial Aid	0.779	0.006	0.771	0.005
	(0.004)	(0.005)	(0.004)	(0.005)
Income Quintile 1	0.213	0.006	0.367	-0.012
	(0.004)	(0.007)	(0.005)	(0.006)
Income Quintile 2	0.224	-0.004	0.254	0.013
	(0.004)	(0.005)	(0.004)	(0.005)
Income Quintile 3	0.191	-0.004	0.174	0.003
	(0.004)	(0.005)	(0.003)	(0.005)
Income Quintile 4	0.197	-0.002	0.128	-0.002
	(0.004)	(0.005)	(0.002)	(0.004)
Income Quintile 5	0.175	0.004	0.076	-0.002
	(0.003)	(0.006)	(0.002)	(0.003)

Note: The table presents results from equation (8) using predetermined characteristics as outcome. The bandwidth is fixed in any specification to 50 for the 600 cutoff, and to 24 for the 500 cutoff. The model is estimated using a triangular weighted kernel. Standard errors are clustered at the PSU level.

Table B4: RD estimates - cohorts 2008-2010

	(1) Enrolled	(2) Graduated	(3) Employed	(4) Private	(5) Voucher	(6) Public	(7) Rural
Panel A. 600 cutoff							
RD Estimate	0.001	-0.003	-0.004	0.000	-0.003	0.000	0.001
	(0.004)	(0.004)	(0.003)	(0.001)	(0.003)	(0.002)	(0.001)
Bandwidth	45	45	49	59	54	45	56
Observations	93,439	94,769	102,938	122,933	112,348	93,439	117,653
Panel B. 500 cutoff							
RD Estimate	0.004	0.001	0.003	0.001	0.002	0.001	-0.001
	(0.004)	(0.003)	(0.004)	(0.001)	(0.003)	(0.002)	(0.001)
Bandwidth	24	24	24	24	24	24	24
Observations	82,676	82,676	82,676	82,676	82,676	82,676	82,676

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table shows Regression Discontinuity parameters where all specifications are estimated using weighted local linear regressions and consider high school graduates taking the PSU before the reform, that is between 2008 and 2010. For the 600 cutoff, the RD Estimate corresponds to the effect of being marginally *above* 600 or marginally eligible for the scholarship; bandwidths are chosen optimally according to Calonico et al. (2019). For the 500 cutoff, the RD Estimate corresponds to the effect of being marginally *below* 500 or marginally affected by the restriction; bandwidths are set to 24 to avoid capturing the effect of the loan eligibility cutoff at 475 points. Standard errors are clustered at the PSU test score level and reported in parentheses. *Observations* correspond to the number of observations with non-zero weight given the chosen bandwidth. *Enrolled* is an indicator for individuals enrolling in a teacher training program in the year after high school graduation. *Graduated* is an indicator for individuals graduating from a teacher training program within eight years of high school graduation. *Employed* is an indicator for individuals observed working in any school within 2 years of their graduation from a teacher training program. *Voucher*, *Public*, and *Rural* are indicators for individuals observed working in each particular type of school within 2 years of their graduation from a teacher training program. All results are unconditional.

Table B5: Sensitivity of RD estimates to different local polynomials

	(1) Enrolled	(2) Enrolled	(3) Graduated	(4) Graduated	(5) Employed	(6) Employed
Panel A. 600 cutoff						
RD Estimate	0.020***	0.019***	0.008**	0.006	0.009**	0.008*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Polynomial order	2	3	2	3	2	3
Bandwidth	66	91	68	96	71	98
Observations	190,924	262,198	195,460	275,067	205,334	280,956
Panel B. 500 cutoff						
RD Estimate	-0.028***	-0.029***	-0.015***	-0.015**	-0.012***	-0.010**
	(0.008)	(0.009)	(0.006)	(0.006)	(0.004)	(0.004)
Polynomial order	2	3	2	3	2	3
Bandwidth	24	24	24	24	24	24
Observations	113,468	113,468	113,468	113,468	113,468	113,468

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table presents the RD estimates for the effect of the reform on enrollment, graduation, and employment. The table shows Regression Discontinuity parameters where all specifications are estimated using weighted local linear regressions, consider high school graduates taking the PSU between 2011 and 2014, and are estimated using a polynomial of order 2 or 3. For the 600 cutoff, the RD Estimate corresponds to the effect of being marginally above 600 or marginally eligible for the scholarship; bandwidths are chosen optimally according to Calonico et al. (2019). For the 500 cutoff, the RD Estimate corresponds to the effect of being marginally below 500 or marginally affected by the restriction; bandwidths are set to 24 to avoid capturing the effect of the loan eligibility cutoff at 475 points. Standard errors are clustered at the PSU test score level and reported in parentheses. *Enrolled* is an indicator for individuals enrolling in a teacher training program in the year after high school graduation. *Graduated* is an indicator for individuals graduating from a teacher training program within eight years of high school graduation. *Employed* is an indicator for individuals observed working in any school within 2 years of their graduation from a teacher training program.

Table B6: Across-cohort comparison

	(1) Enrolled	(2) Graduated	(3) Employed	(4) Public	(5) Voucher	(6) Private	(7) Rural
Panel A. Above 600							
After Reform	0.013***	0.006***	0.006***	0.002***	0.005***	0.000	0.000
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Observations	108,092	108,092	108,092	108,092	108,092	108,092	108,092
Baseline Mean	0.04	0.04	0.03	0.01	0.02	0.01	0.00
Percentage Change	34.3	14.4	19.4	25.4	27.7	0.9	12.9
Panel B. Below 500							
After Reform	-0.006***	-0.007***	-0.008***	-0.004***	-0.004***	-0.000**	-0.002***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
Observations	390,297	390,297	390,297	390,297	390,297	390,297	390,297
Baseline Mean	0.04	0.04	0.03	0.01	0.02	0.00	0.01
Percentage Change	-18.0	-16.0	-24.0	-27.3	-23.1	-19.7	-29.4
Panel C. Full sample							
After Reform	-0.002**	-0.006***	-0.006***	-0.004***	-0.003***	-0.000	-0.001***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
Observations	694,874	694,874	694,874	694,874	694,874	694,874	694,874
Baseline Mean	0.05	0.06	0.04	0.02	0.03	0.00	0.01
Percentage Change	-3.9	-11.3	-14.1	-20.0	-11.4	-5.5	-20.6

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table shows differences in means between outcomes for PSU takers before and after 2011, restricting the cohorts of interest to 2009 to 2012. Standard errors are clustered at the PSU test score level and reported in parentheses. The *Baseline Mean* refers to the outcome mean for students enrolling before the 2011 reform. The *Percentage Change* is computed as the coefficient over the Baseline Mean (times 100). *Enrolled* is an indicator for individuals enrolling in a teacher training program in the year after high school graduation. *Graduated* is an indicator for individuals graduating from a teacher training program within eight years of high school graduation. *Employed* is an indicator for individuals observed working in any school within 2 years of their graduation from a teacher training program. *Voucher*, *Public*, and *Rural* are indicators for individuals observed working in each particular type of school within 2 years of their graduation from a teacher training program.

Table B7: Cohorts taking the SIMCE by year

	Cohort										
Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
2019									8		
2018						10				**6**	
2017					10		**8**				4
2016				10				**6**		4	
2015			10		**8**		**6**		4		2^*
2014		10		8		**6**		4		2^*	
2013	10		8		**6**		4		2^*		
2012						4		2^*			
2011	8				4						
2010				4							
2009			4								
2008		4									
2007	4										

Note: Cohorts are identified by the year in which students start 1^{st} grade of elementary school. The table shows in which grades different cohorts took the SIMCE. Grades 2,4,6, and 8 correspond to elementary school grades, and 10th grade corresponds to high school. The years in which students could potentially be taught by a BVP-eligible teacher have been highlighted in gray. Finally, note that second graders are only tested on their Spanish knowledge.

Table B8: Summary statistics mathematics teachers in the sample

	All			PSU< 500		PSU≥ 600	
	(1) All	(2) Before	(3) After	(4) Before	(5) After	(6) Before	(7) After
Teacher characteristics							
Female	0.71	0.71	0.74	0.75	0.74	0.60	0.68
	(0.45)	(0.46)	(0.44)	(0.43)	(0.44)	(0.49)	(0.47)
Father has college education	0.17	0.17	0.18	0.11	0.11	0.33	0.34
	(0.37)	(0.37)	(0.39)	(0.31)	(0.32)	(0.47)	(0.48)
Mother has college education	0.13	0.13	0.12	0.09	0.04	0.22	0.26
	(0.33)	(0.33)	(0.33)	(0.29)	(0.21)	(0.42)	(0.44)
Quintile 1 Family Income	0.44	0.46	0.41	0.53	0.44	0.31	0.30
	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)	(0.46)	(0.46)
Quintile 2 Family Income	0.21	0.20	0.23	0.22	0.30	0.21	0.19
	(0.41)	(0.40)	(0.42)	(0.42)	(0.46)	(0.41)	(0.39)
Quintile 3 Family Income	0.15	0.15	0.15	0.12	0.10	0.17	0.19
	(0.35)	(0.35)	(0.35)	(0.32)	(0.31)	(0.38)	(0.39)
Quintile 4 Family Income	0.12	0.11	0.14	0.09	0.08	0.17	0.17
	(0.32)	(0.32)	(0.34)	(0.28)	(0.27)	(0.37)	(0.37)
Quintile 5 Family Income	0.08	0.08	0.08	0.04	0.08	0.14	0.16
	(0.27)	(0.27)	(0.28)	(0.21)	(0.27)	(0.35)	(0.37)
Years of experience	2.55	2.72	0.83	2.86	1.01	2.36	0.79
	(1.92)	(1.92)	(0.78)	(1.96)	(0.85)	(1.91)	(0.68)
Works in public school	0.34	0.35	0.30	0.45	0.44	0.21	0.23
	(0.47)	(0.47)	(0.45)	(0.49)	(0.49)	(0.40)	(0.41)
Works in private school	0.07	0.08	0.05	0.04	0.03	0.20	0.08
	(0.26)	(0.26)	(0.22)	(0.19)	(0.18)	(0.39)	(0.26)
Works in subsidized school	0.58	0.58	0.65	0.51	0.53	0.59	0.69
	(0.48)	(0.48)	(0.47)	(0.49)	(0.49)	(0.48)	(0.45)
Works in rural school	0.17	0.17	0.16	0.23	0.28	0.06	0.06
	(0.37)	(0.37)	(0.36)	(0.42)	(0.45)	(0.24)	(0.24)
# classrooms taught	2.20	2.29	1.32	2.25	1.44	2.01	1.26
	(2.11)	(2.17)	(0.92)	(2.15)	(1.02)	(1.89)	(0.83)
# classrooms taught per year	1.14	1.15	1.03	1.11	1.06	1.17	1.02
	(0.71)	(0.72)	(0.60)	(0.69)	(0.66)	(0.75)	(0.58)
Classroom characteristics							
Class size	18.56	18.49	19.26	16.84	15.83	20.85	21.44
	(12.25)	(12.21)	(12.59)	(11.93)	(12.64)	(12.28)	(12.12)
% female students	0.42	0.42	0.43	0.41	0.42	0.42	0.43
	(0.24)	(0.24)	(0.25)	(0.24)	(0.27)	(0.26)	(0.23)
% students Q1 family income	0.17	0.17	0.14	0.21	0.19	0.10	0.11
	(0.21)	(0.21)	(0.19)	(0.23)	(0.22)	(0.15)	(0.16)
% students Q5 family income	0.17	0.16	0.19	0.11	0.09	0.28	0.27
	(0.26)	(0.26)	(0.26)	(0.20)	(0.18)	(0.34)	(0.31)
% students with college educated mother	0.11	0.11	0.12	0.08	0.06	0.20	0.17
	(0.17)	(0.17)	(0.17)	(0.13)	(0.11)	(0.25)	(0.20)
% students with college educated father	0.13	0.13	0.14	0.09	0.07	0.23	0.20
	(0.20)	(0.20)	(0.20)	(0.15)	(0.14)	(0.28)	(0.24)
Average lagged SIMCE score Math	-0.03	-0.03	-0.03	-0.14	-0.16	0.19	0.10
	(0.48)	(0.48)	(0.48)	(0.45)	(0.43)	(0.52)	(0.53)
Average lagged SIMCE score Spanish	0.01	0.01	0.01	-0.08	-0.09	0.17	0.11
	(0.40)	(0.40)	(0.40)	(0.39)	(0.41)	(0.42)	(0.43)
Observations	4386	3981	405	1525	93	608	117

Note: The table presents the mean and standard deviation (in parenthesis) of several variables for mathematics teachers in our main TVA analysis sample. Column 1 presents the statistics for the full sample. Columns 2 and 3, divide the sample between teachers that started their studies before and after the reform. Columns 4 and 5 present the summary statistics for those scoring below 500 points. Columns 6 and 7 present the summary statistics for those scoring above 600 points.

Table B9: Summary statistics Spanish teachers in the sample

	All			PSU$<$ 500		PSU$\geq$ 600	
	(1) All	(2) Before	(3) After	(4) Before	(5) After	(6) Before	(7) After
Teacher characteristics							
Female	0.81	0.81	0.84	0.83	0.85	0.73	0.80
	(0.39)	(0.39)	(0.37)	(0.37)	(0.36)	(0.44)	(0.40)
Father has college education	0.16	0.17	0.13	0.12	0.06	0.38	0.27
	(0.37)	(0.37)	(0.33)	(0.32)	(0.24)	(0.49)	(0.45)
Mother has college education	0.13	0.13	0.11	0.09	0.06	0.30	0.28
	(0.33)	(0.33)	(0.31)	(0.29)	(0.23)	(0.46)	(0.45)
Quintile 1 Family Income	0.47	0.49	0.41	0.56	0.50	0.21	0.21
	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)	(0.41)	(0.41)
Quintile 2 Family Income	0.21	0.19	0.26	0.20	0.24	0.21	0.26
	(0.41)	(0.39)	(0.44)	(0.40)	(0.43)	(0.41)	(0.44)
Quintile 3 Family Income	0.15	0.14	0.17	0.11	0.14	0.18	0.24
	(0.35)	(0.35)	(0.37)	(0.31)	(0.35)	(0.38)	(0.43)
Quintile 4 Family Income	0.11	0.11	0.11	0.07	0.09	0.21	0.15
	(0.31)	(0.31)	(0.31)	(0.25)	(0.29)	(0.41)	(0.36)
Quintile 5 Family Income	0.07	0.08	0.06	0.06	0.03	0.19	0.14
	(0.26)	(0.27)	(0.24)	(0.24)	(0.17)	(0.39)	(0.35)
Years of experience	2.63	2.81	0.80	2.93	0.85	2.63	0.86
	(1.97)	(1.96)	(0.76)	(2.00)	(0.86)	(1.89)	(0.78)
Works in public school	0.33	0.33	0.31	0.43	0.45	0.13	0.24
	(0.46)	(0.46)	(0.46)	(0.49)	(0.50)	(0.33)	(0.42)
Works in private school	0.09	0.09	0.06	0.04	0.03	0.29	0.09
	(0.28)	(0.28)	(0.23)	(0.19)	(0.16)	(0.45)	(0.29)
Works in subsidized school	0.58	0.57	0.63	0.53	0.52	0.58	0.67
	(0.48)	(0.48)	(0.48)	(0.49)	(0.50)	(0.48)	(0.47)
Works in rural school	0.16	0.16	0.18	0.21	0.28	0.03	0.06
	(0.36)	(0.36)	(0.38)	(0.40)	(0.45)	(0.16)	(0.23)
# classrooms taught	2.07	2.13	1.42	2.08	1.42	2.19	1.25
	(1.99)	(2.05)	(1.06)	(1.99)	(1.08)	(2.11)	(0.93)
# classrooms taught per year	1.14	1.14	1.10	1.11	1.04	1.23	1.06
	(0.74)	(0.74)	(0.68)	(0.72)	(0.61)	(0.81)	(0.67)
Classroom characteristics							
Class size	18.70	18.56	20.18	17.17	17.19	21.14	21.86
	(12.38)	(12.33)	(12.79)	(11.90)	(12.62)	(12.77)	(12.70)
% female students	0.41	0.41	0.43	0.41	0.44	0.42	0.44
	(0.24)	(0.24)	(0.24)	(0.24)	(0.24)	(0.26)	(0.25)
% students Q1 family income	0.17	0.17	0.13	0.21	0.19	0.07	0.09
	(0.21)	(0.21)	(0.18)	(0.23)	(0.24)	(0.11)	(0.11)
% students Q5 family income	0.18	0.17	0.18	0.11	0.13	0.36	0.27
	(0.27)	(0.27)	(0.24)	(0.20)	(0.19)	(0.37)	(0.31)
% students with college educated mother	0.12	0.12	0.11	0.08	0.07	0.26	0.17
	(0.18)	(0.18)	(0.15)	(0.13)	(0.11)	(0.28)	(0.21)
% students with college educated father	0.14	0.14	0.12	0.09	0.08	0.30	0.20
	(0.21)	(0.21)	(0.17)	(0.14)	(0.12)	(0.31)	(0.24)
Average lagged SIMCE score Math	-0.01	-0.01	-0.01	-0.12	-0.11	0.31	0.15
	(0.49)	(0.49)	(0.50)	(0.45)	(0.47)	(0.54)	(0.52)
Average lagged SIMCE score Spanish	0.01	0.01	0.03	-0.08	-0.01	0.24	0.13
	(0.42)	(0.42)	(0.42)	(0.39)	(0.41)	(0.45)	(0.46)
Observations	4966	4515	451	2145	113	472	96

Note: The table presents the mean and standard deviation (in parenthesis) of several variables for Spanish teachers in our main TVA analysis sample. Column 1 presents the statistics for the full sample. Columns 2 and 3, divide the sample between teachers that started their studies before and after the reform. Columns 4 and 5 present the summary statistics for those scoring below 500 points. Columns 6 and 7 present the summary statistics for those scoring above 600 points.

Table B10: Effect of a 1 Standard deviation increase in teacher and school effects

	With weights		**Without weights**	
	Math	**Spanish**	**Math**	**Spanish**
Including controls and only contemporaneous classrooms				
School	.24	.19	.24	.2
Teacher	.09	.09	.08	.08
Without controls and all different classrooms				
School	.22	.13	.23	.14
Teacher	.11	.11	.09	.11
Without controls and only contemporaneous classrooms				
School	.25	.19	.26	.21
Teacher	.11	.09	.09	.08

Table B11: Robustness - effect of the reform on TVA using different before/after specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Math						
Post-reform cohort	0.036**	0.037**	0.035**	0.047***	0.038**	0.037**
	(0.017)	(0.016)	(0.016)	(0.017)	(0.016)	(0.016)
Observations	193143	206094	206094	206094	179445	130412
Sample						
Teachers	4,133	4,386	4,386	4,386	4,160	3,556
Panel B. Spanish						
Post-reform cohort	-0.001	-0.004	-0.005	0.007	-0.007	-0.007
	(0.015)	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)
# Students	207,617	225,584	225,584	225,584	196,655	142,267
# Teachers	4,619	4,966	4,966	4,966	4,662	3,891
Additional covariates	Yes	No	No	No	No	No
Individual covariates	Yes	No	No	No	Yes	Yes
School Covariates	Yes	Yes	Yes	No	Yes	Yes
Class covariates	Yes	Yes	No	No	Yes	Yes
Years of Experience	All	All	All	All	<6	<3

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table shows OLS estimates from model (14). Standard errors are clustered at the teacher-by-year level, and shown in parentheses. Additional covariates include the lagged GPA of the student and the share of days in which she went to school the previous year.

Table B12: Robustness - effect of the reform on TVA for on-time-finishers

	Math teachers		Spanish teachers	
	(1)	(2)	(3)	(4)
Post-reform cohort	0.074*	0.064**	0.004	-0.000
	(0.042)	(0.026)	(0.051)	(0.028)
# Students	22,383	39,359	26,693	44,562
# Teachers	937	1,601	1,060	1,810
Δ Std PSU	0.284	0.254	0.272	0.294
% Post-reform teachers	0.319	0.239	0.316	0.235
Years after PSU	6	7	6	7

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table shows OLS estimates from model (14), restricting the sample only to those teachers that finish their studies on time and are first observed as teachers six (in columns (1) and (3)) or seven (in columns (2) and (4)) years after taking the PSU. Additionally, we restrict the student observations to those taught on the first year the teacher is observed. Standard errors are clustered at the teacher-by-year level, and shown in parentheses.

Table B13: Robustness - effect beyond ability - Math teachers

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. PSU < 500						
Post-reform cohort	0.058	0.067*	0.061*	0.048	0.062*	0.061*
	(0.036)	(0.036)	(0.035)	(0.038)	(0.036)	(0.036)
# Students	67,992	70,092	70,092	70,092	59,507	41,529
# Teachers	1,577	1,618	1,618	1,618	1,521	1,263
Panel C. PSU $\in [500, 600)$						
Post-reform cohort	0.005	-0.000	-0.001	0.019	0.004	0.002
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
# Students	98,506	103,880	103,880	103,880	91,258	66,436
# Teachers	1,941	2,043	2,043	2,043	1,950	1,678
Panel D. PSU ≥ 600						
Post-reform cohort	0.054*	0.064**	0.063**	0.054*	0.065**	0.062**
	(0.028)	(0.027)	(0.026)	(0.029)	(0.027)	(0.027)
# Students	26,645	32,122	32,122	32,122	28,680	22,447
# Teachers	615	725	725	725	689	615
Additional covariates	Yes	No	No	No	No	No
Individual covariates	Yes	No	No	No	Yes	Yes
School Covariates	Yes	Yes	Yes	No	Yes	Yes
Class covariates	Yes	Yes	No	No	Yes	Yes
Years of Experience	All	All	All	All	<6	<4

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table shows OLS estimates from model (14). Standard errors are clustered at the teacher-by-year level, and shown in parentheses. Additional covariates include the lagged GPA of the student and the share of days in which she went to school the previous year.

Table B14: Beyond ability - adjusting for higher education and school characteristics

	Baseline	Program characteristics			School characteristics			All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Full sample								
Post-reform cohort	0.034**	0.034**	0.032**	0.032**	0.035**	0.029*	0.030*	0.028*
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Panel B. By score								
Post-reform cohort × PSU < 500	0.065*	0.065*	0.062*	0.063*	0.066*	0.065*	0.066*	0.064*
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.035)	(0.035)	(0.035)
Post-reform cohort × PSU $\in [500, 600)$	0.001	0.001	-0.002	-0.001	0.002	-0.005	-0.004	-0.006
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.020)
Post-reform cohort × PSU $\geq$ 600	0.065**	0.065**	0.065**	0.064**	0.067**	0.057**	0.058**	0.057**
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.025)	(0.025)	(0.025)
# Students	206,094	206,094	206,094	206,094	206,094	206,094	206,094	206,094
# Teachers	4,386	4,386	4,386	4,386	4,386	4,386	4,386	4,386
PSU	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HE program type	No	Yes	No	No	No	No	Yes	Yes
Program size & peer PSU	No	No	Yes	No	No	No	Yes	Yes
Colleague characteristics	No	No	No	Yes	No	Yes	Yes	Yes
IVE-SINAE	No	No	No	No	Yes	Yes	Yes	Yes
All controls	No	No	No	No	No	No	No	Yes

Note: * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. The table shows OLS estimates for the coefficients of interest in model (14). We control for a polynomial of order two of PSU score. Standard errors are clustered at the teacher-by-year level, and shown in parentheses.

Table B15: Correlation between teacher quality and inputs

	Math			Spanish		
	(1)	(2)	(3)	(4)	(5)	(6)
PSU score	0.034***	0.035***	0.024***	-0.013*	-0.011	-0.016**
	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)
Years of experience	0.016	0.017	0.016	0.032***	0.033***	0.032***
	(0.011)	(0.011)	(0.011)	(0.009)	(0.009)	(0.009)
Choose again		0.033**	0.037**		0.029**	0.031**
		(0.016)	(0.016)		(0.014)	(0.014)
Observations	812	812	812	905	905	905
R^2	0.046	0.051	0.072	0.018	0.023	0.031
Controls	No	No	Yes	No	No	Yes

Note: * $p < 0.05$, *** $p < 0.01$. The table shows the estimates of an OLS regression where we estimate the correlation of teacher quality with motivation, years of experience, and PSU score. PSU scores here are standardized to have a mean of zero and a standard deviation of one. All regressions control for a polynomial of order two on years of experience, a polynomial of order two on PSU scores. Columns 3 and 6 additionally control for the type of school the teacher works at (public, private, rural). Robust standard error reported in parentheses.

C Recruitment - Additional Results

In this section, we present additional results related to the section on recruitment.

Characterizing enrollees Appendix Table B2 presents the results of performing our RD analysis, but using observable characteristics as dependent variables. We observe that marginally incentivized students ($PSU \geq 600$) are similar to their peers right below the threshold. The only noteworthy difference relates to the family income quintile. Students who are enrolled in teacher training programs just above the scholarship cutoff are more likely to come from poorer families, compared to those just below. Similarly, we observe that students enrolling in teacher training programs right below the 500 cutoff come from richer families.

Where do they come from? Students scoring above 600 select different *majors* within the pool of teaching colleges. When dividing degrees by the grade levels they prepare individuals to teach (e.g., secondary or primary school), we see that they are more (less) likely to choose a high-school (primary school) career (see Table C17). The same holds true for those marginally eligible at the 500 cutoff. Finally, regarding the counterfactual college career path, we show in Figure C21 that BVP takers would not have opted to forgo enrolling in college. The effect on general enrollment is estimated to be close to zero at the cutoff. Students at the 500 cutoff are instead more likely to enroll. Therefore, some students who see their choice set reduced decide not to pursue a higher education career.

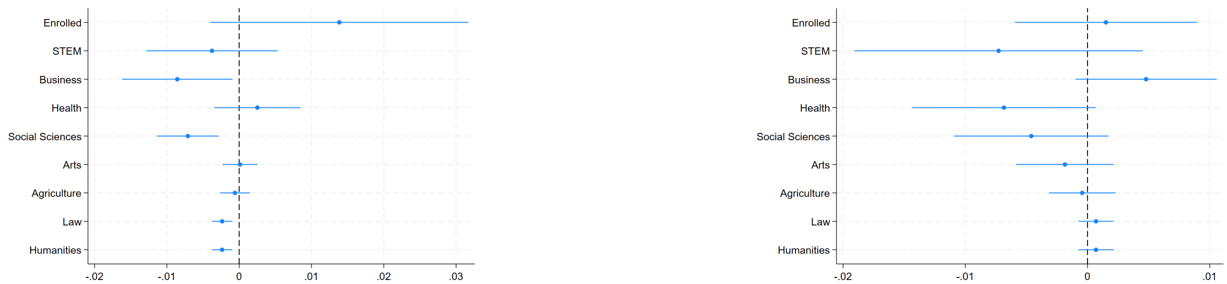
Do they choose different types of teacher training programs? It is difficult to infer from the name of the program in which students enroll whether this choice preconditions them to become mathematics or Spanish teachers. Therefore, in the recruitment section, we cannot directly test whether students are more likely to select into teaching math or Spanish. What we can test, however, is whether they choose programs that naturally track into teaching at the pre-school, primary, or high-school level. For a subset of programs, it is still not possible to determine the eventual teaching track. In Table C17, we examine whether, conditional on enrolling in a teacher training program, marginally eligible scholarship recipients are more likely to select one type of program over another. The results indicate that the distribution of program types changes slightly above the cutoff: enrollment in high-school–oriented programs increases, while enrollment in primary school programs decreases.

Table C16: Characteristics enrollees

	(1) PSU ≤ 500	(2) PSU ≥ 600
High School GPA	-0.072***	-0.031**
	(0.023)	(0.015)
	[5.590]	[5.868]
Female	-0.008	-0.021
	(0.035)	(0.017)
	[0.671]	[0.624]
Father High School	0.054**	0.044**
	(0.023)	(0.020)
	[0.516]	[0.655]
Mother High School	0.041**	-0.002
	(0.016)	(0.018)
	[0.617]	[0.746]
Any Parent College	0.053**	0.020
	(0.025)	(0.019)
	[0.302]	[0.527]
# Family Members	0.126**	0.038
	(0.062)	(0.052)
	[4.366]	[4.471]
Single Parent	0.018	0.014
	(0.016)	(0.012)
	[0.171]	[0.135]
Father Working	0.005	0.001
	(0.024)	(0.017)
	[0.638]	[0.691]
Mother Working	0.005	0.025
	(0.022)	(0.018)
	[0.391]	[0.426]
Income Quintile 1	-0.029*	0.048**
	(0.016)	(0.022)
	[0.369]	[0.259]
Income Quintile 2	-0.015	-0.004
	(0.018)	(0.016)
	[0.290]	[0.240]
Income Quintile 3	-0.004	-0.040**
	(0.019)	(0.016)
	[0.167]	[0.210]
Income Quintile 4	0.017	-0.014
	(0.017)	(0.017)
	[0.124]	[0.192]
Income Quintile 5	0.031***	0.010
	(0.009)	(0.012)
	[0.050]	[0.099]

Note: The table presents regression discontinuity parameters using predetermined characteristics as outcomes. The bandwidth is fixed to 24 for the 500 cutoff, to 50 for the 600 cutoff, and the model is estimated using a triangular weighted kernel for all specifications. Standard errors are clustered at the PSU level and are shown in parentheses. Baseline means are presented below the standard errors, in brackets.

Figure C21: Effect on general enrollment and enrollment in other fields



(a) 500 cutoff (b) 600 cutoff

Note: All outcomes are binary indicators. All specifications are estimated using weighted local linear regressions. Bandwidths are chosen optimally according to Calonico et al. (2019). Standard errors are clustered at the PSU test score level.

Table C17: Categories of programs within teacher colleges

	(1) Preschool	(2) Primary	(3) High School	(4) Other
Panel A. 600 cutoff				
RD Estimate	-0.010	-0.022*	0.039**	-0.005
	(0.010)	(0.013)	(0.018)	(0.023)
Baseline Mean	0.061	0.123	0.337	0.475
Bandwidth	43	48	58	47
Observations	9,184	10,391	12,472	10,076

Note: ** $p < 0.05$, *** $p < 0.01$. The Table presents estimated discontinuity in the relationship between PSU and outcomes of interest at the 600 and 500 cutoffs. The sample is composed of individuals who enrolled in a teacher college, and the outcomes are binary indicators for enrolling in a program that leads to teaching in preschool, elementary school, and high school. We add an indicator for when the program does not specify the specialization (others).

D Theoretical Framework - Additional Results

Proof of equation (6). **Distributional assumption.** Let $(\epsilon_T, \epsilon_O, \epsilon_M)$ be jointly normally distributed with mean zero and some positive definite covariance matrix Σ. Define the selection index

$$Z := -r_O\,\epsilon_O + \epsilon_M.$$

Because Z is a linear combination of jointly normal variables, (ϵ_T, Z) is jointly normal. Let $\mu_T := \mathbb{E}[\epsilon_T] = 0$, $\mu_Z := \mathbb{E}[Z] = 0$, $\sigma_Z^2 := \text{Var}(Z)$, and $\sigma_{TZ} := \text{Cov}(\epsilon_T, Z)$.

Policy, cutoff, and selection event. Fix an ability bin A so that $K(A)$ is constant within the bin. The policy $S(A)$ shifts the cutoff in the selection rule:

$$\text{choose teaching} \iff Z \geq C(A) := -K(A) - S(A).$$

Let $c_0 := -K(A)$ be the pre-reform cutoff and $c_1 := C(A)$ the post-reform cutoff (with $c_1 \neq c_0$ whenever $S(A) \neq 0$). For any cutoff c, denote the selected set by $\{Z \geq c\}$ and write

$$\mathbb{E}_c[\cdot] := \mathbb{E}[\cdot \mid Z \geq c].$$

Step 1: Conditional mean of a bivariate normal is linear. Since (ϵ_T, Z) is jointly normal,

$$\mathbb{E}[\epsilon_T \mid Z = z] = \mu_T + \frac{\text{Cov}(\epsilon_T, Z)}{\text{Var}(Z)}(z - \mu_Z) = \frac{\sigma_{TZ}}{\sigma_Z^2}z,$$

where we used $\mu_T = \mu_Z = 0$. This follows from the least-squares projection property of the multivariate normal.

Step 2: Law of iterated expectations under selection. Applying the tower property,

$$\mathbb{E}_c[\epsilon_T] = \mathbb{E}\Big[\mathbb{E}[\epsilon_T \mid Z]\Big|Z \geq c\Big] = \mathbb{E}\left[\frac{\sigma_{TZ}}{\sigma_Z^2}Z\middle|Z \geq c\right] = \frac{\sigma_{TZ}}{\sigma_Z^2}\mathbb{E}_c[Z].$$

Therefore, for any two cutoffs c_0, c_1,

$$\mathbb{E}_{c_1}[\epsilon_T] - \mathbb{E}_{c_0}[\epsilon_T] \;=\; \frac{\sigma_{TZ}}{\sigma_Z^2}\Big(\mathbb{E}_{c_1}[Z] - \mathbb{E}_{c_0}[Z]\Big).$$

□

Remark. *Explicit form for* $\mathbb{E}_c[Z]$*).* Because $Z \sim \mathcal{N}(0, \sigma_Z^2)$, the truncated normal mean is

$$\mathbb{E}_c[Z] = \sigma_Z\,\lambda(\alpha)$$

with

$$\alpha := \frac{c}{\sigma_Z} \qquad \text{and} \qquad \lambda(\alpha) := \frac{\phi(\alpha)}{1 - \Phi(\alpha)},$$

where ϕ and Φ are the standard normal PDF and CDF. This confirms that changing the cutoff c, via the introduction of $S(A)$, changes $\mathbb{E}_c[Z]$ and hence $\mathbb{E}_c[\epsilon_T]$, exactly in proportion to $\frac{\text{Cov}(\epsilon_T, Z)}{\text{Var}(Z)}$.

Prediction on changes in motivation

The theoretical framework also yields predictions for changes in ϵ_M, which follow the same logic and are likewise governed by the outside–option parameter r_O. In fact, equation (6) can be rewritten replacing ϵ_T with ϵ_M.

$$\Delta\mathbb{E}[\epsilon_M] \;=\; \frac{\text{Cov}(\epsilon_M, Z)}{\text{Var}(Z)}\, \Delta\mathbb{E}[Z], \tag{15}$$

In this case, the statistic that would determine the sign of $\Delta\mathbb{E}[\epsilon_M]$ would be $\text{Cov}(\epsilon_M, Z) = \sigma_M - r_O\rho_{M,O}\sigma_O$. The key difference is that motivation enters the index Z directly: its "own" term σ_M contributes one–for–one rather than being scaled by $\rho_{T,M}$. By contrast, the pull–out force on motivation operates through the correlation between motivation and outside productivity, $\rho_{M,O}$. Under the plausible assumption $\rho_{M,O} < \rho_{T,O}$, it is easy to show that the framework delivers a sharp prediction whenever $Cov(\epsilon_T, Z) > 0$. First,

$$\text{Cov}(\epsilon_T, Z) > 0 \iff r_O \;<\; \frac{\rho_{T,M}}{\rho_{T,O}} \cdot \frac{\sigma_M}{\sigma_O}.$$

Second,

$$\text{Cov}(\epsilon_M, Z) > 0 \iff r_O \;<\; \frac{1}{\rho_{M,O}} \cdot \frac{\sigma_M}{\sigma_O}.$$

Since $\rho_{M,O} < \rho_{T,O}$, it follows that

$$\frac{\rho_{T,M}}{\rho_{T,O}} \;\leq\; \frac{1}{\rho_{T,O}} \;\leq\; \frac{1}{\rho_{M,O}},$$

and therefore

$$r_O < \frac{\rho_{T,M}}{\rho_{T,O}} \cdot \frac{\sigma_M}{\sigma_O} \implies r_O < \frac{1}{\rho_{M,O}} \cdot \frac{\sigma_M}{\sigma_O}.$$

Hence $\text{Cov}(\epsilon_T, Z) > 0$ implies $\text{Cov}(\epsilon_M, Z) > 0$. When $\text{Cov}(\epsilon_T, Z) < 0$, the condition only implies

$$r_O > \frac{\rho_{T,M}}{\rho_{T,O}} \cdot \frac{\sigma_M}{\sigma_O},$$

which does not determine the sign of $\text{Cov}(\epsilon_M, Z)$. The prediction for motivation is therefore ambiguous in this case.