The Unintended Consequences of Merit-Based Teacher Selection: Evidence from Large-Scale Reform in Colombia

Matias Busso, Sebastián Montaño, Juan Muñoz-Morales, and Nolan Pope*

February 23, 2023

Abstract

Teacher quality is a key factor in improving student academic achievement. As such, education policymakers are incentivized to design screening systems to hire the most effective teachers. In this paper, we study the effects of introducing a merit-based screening system for all Colombian public schools that ties teacher hiring decisions to candidates' performance on an exam evaluating subject-specific knowledge and teaching aptitude. This policy had the intended result of increasing the pre-college test scores of teachers by 17 percentile points. However, it also unintentionally increased the number of novice teachers and decreased the overall stock of teacher experience. Using a difference-in-differences strategy that compares the outcomes of students from public and private schools over two decades, we find that instead of improving student outcomes, this screening system decreased students' performance on high school exit exams by 8 percent of a standard deviation, and reduced the likelihood of enrolling and graduating from college by over 10 percent. These results suggest that using ex-ante teacher quality measures to screen teacher candidates may have a limited effect and can even result in negative selection of teacher characteristics important for students' achievement.

Keywords: Teachers, teaching experience, teacher screening, test scores, college enrollment

JEL codes: I25, I28, J24

*Busso: Inter-American Development Bank (mbusso@iadb.org). Montaño: The University of Maryland (montano@umd.edu). Muñoz: IESEG School of Management, Univ. Lille, CNRS, UMR 9221-LEM-Lille Économie Management, F-59000 Lille, France (j.munoz@ieseg.fr) Pope: The University of Maryland (npope@umd.edu). The opinions expressed in this document are those of the authors and do not necessarily reflect the views of the University of Maryland, IÉSEG School of Management, or the Inter-American Development Bank, its Board of Directors, and the countries they represent. Errors are our own.

1 Introduction

Large disparities exist in students' educational outcomes across countries, states, and school districts (Hanushek and Woessmann, 2011; Blanden, Doeple and Stuhler, 2022). In an effort to improve educational outcomes and close the gaps with high-performing areas, countries and policymakers often implement sweeping education reforms. Due to the fact that teacher quality has been shown to be a main determinant of student human capital development (Chetty, Friedman and Rockoff, 2014; Hanushek and Rivkin, 2006; Rivkin, Hanushek and Kain, 2005), these reforms often focus on how to attract, select, and retain high-quality teachers. Many countries, particularly in Latin America, have implemented nationwide merit-based hiring systems to select new teachers based on an array of information that often includes teachers' scores on standardized exams (Elacqua et al., 2018; Cruz-Aguayo, Hincapie and Rodriguez, 2020).

The success of these teacher hiring systems requires that the information used to screen candidates is predictive of teacher quality. Unfortunately, as past work has shown, many of the observable characteristics of teachers fail to predict their teaching effectiveness (Hanushek and Rivkin, 2006; Rockoff et al., 2011). Therefore, teacher hiring systems may struggle to foster teacher quality or to improve students' learning outcomes (Kane and Staiger, 2005; Harris, Ingle and Rutledge, 2014). Furthermore, hiring systems that heavily weight specific characteristics of candidates – such as licensing, educational attainment, or the candidates' own performance on standardized exams – could cause negative selection on dimensions that are more predictive of teacher quality, such as experience (Staiger and Rockoff, 2010). In such cases, teacher hiring policies could unintentionally have a negative impact on students, especially when implemented on a large scale.

In this paper, we look at how the introduction of a merit-based teacher hiring system impacted both the attributes of teachers and the educational outcomes of students. Specifically, we assess the effect of a nationwide teacher hiring program implemented by the Colombian government starting in 2005 for all public schools. The re-

form changed the teacher hiring process from a decentralized system to a system that ties hiring to performance on a standardized entrance exam. Due to the government's large-scale implementation of the reform, nearly half of all public school teachers were hired under the new regulation within ten years. Additionally, the reform also tried to attract a higher-quality pool of teacher candidates by increasing entry-level salaries.

To estimate the impact of the reform, we use administrative data on teachers and students spanning two decades. The teacher data allow us to measure how the reform changed the composition of the staff at public schools for many characteristics such as pre-college test scores, education level, age, gender, and experience. The student data allow us to observe the performance of students on high school exit exams and subsequent college outcomes. We estimate the impact of the reform on students' outcomes by leveraging the fact that the reform only changed the hiring process and wages for new public school teachers while having no direct impact on private schools. This allows us to compare students in public and private schools before and after the teacher selection policy was implemented using a difference-in-differences strategy.

We find that teachers hired under the new system have substantially higher cognitive skills than teachers hired under the previous system. Specifically, while the education level, gender, and hiring age of teachers remained stable, teachers pre-college test scores increased by 17 percentile points after the hiring reform was implemented. However, the reform also replaced many temporary teachers (i.e. teachers without a long-term contract) with several years of teaching experience. Within two years of implementation, the reform had replaced nearly 50 thousand temporary teachers hired pre-reform – or 17 percent of all teachers. In addition, after the first two years of the reform, over 4,000 temporary teachers were being replaced annually, equivalent to 37 percent of the teachers that left the profession each year. As a result, the newly hired teachers were significantly less experienced than those employed before the reform was passed. In particular, four years after the reform was implemented, the share of teachers with less than five years of experience tripled, rising from 10 to 30 percent.

Due to these changes in teacher composition, the reform had large negative im-

pacts on students' academic achievement. We find that the introduction of the teacher hiring system decreased the overall performance of public school students by 8.2 percent of a standard deviation on the high school exit exam. This is roughly equivalent to a one standard deviation decrease in teacher quality (Chetty, Friedman and Rockoff, 2014; Petek and Pope, 2023). The overall negative effect on students' performance is explained by large negative impacts on the mathematics and English subject exams; although negative effects are also estimated in all other subjects: reading, natural sciences, and social sciences.

In addition to the negative consequences of the reform on student test scores, we also provide estimates for the impact of the reform on college enrollment and graduation. The new teacher hiring system decreased college enrollment by 3.3 percentage points, equivalent to a 21 percent drop in the likelihood that a public school student attends college after graduating high school. These large declines in college enrollment are partially explained by the fact that most colleges in the country define cutoffs or assign significant weight to student performance on the high school exit exam to offer admissions to different college programs. Finally, the large decline in test scores and college enrollment resulted in a 0.9 percentage point (or 9 percent) reduction in the college graduation rates of public school students.

The reform negatively impacted students' outcomes despite the explicit attempt of the reform to improve teacher quality and subsequently student learning. These results call for caution when governments aim to implement screening policies that rely on ex-ante measures of teacher quality to hire new teachers. The Colombian reform increased new teachers' own test scores, which is not necessarily correlated with effectiveness in improving student learning. But the reform also reduced the stock of teaching experience, which can hurt students' academic achievement (Staiger and Rockoff, 2010; Araujo et al., 2016). As such, future teacher hiring policies should carefully evaluate what ex-ante teacher information they use when designing their hiring systems. Given how difficult it is to ex-ante predict teacher quality, future reforms focused on improving teacher quality may even want to avoid selecting teachers based on ex-ante

information in lieu of retaining and laying off teachers based on ex-post measures of teacher quality. In addition, these results indicate that teacher hiring systems and educational reforms reduce their efficacy when they rely on temporary teachers or have high levels of teacher turnover due to increased learning loses from more student being taught by novice teachers each year.

Our paper contributes to the literature studying the effects of personnel policies and educational reforms that aim to improve teacher quality. Such policies commonly define scoring systems that weight the background information of teacher candidates (e.g., degrees, experience, and licensure) along with additional data collected throughout interviews, in-person or video-recorded teaching samples, or even test scores from an entrance exam. Goldhaber, Grout and Huntington-Klein (2017) and Jacob et al. (2018) provide some evidence of a positive effect of such screening systems on teachers' value-added in the context of the United States. In the Latin American context, Cruz-Aguayo, Ibarrarán and Schady (2017), Estrada (2019), Araujo et al. (2020), and Araujo (2022) study the effects of policies in which teacher candidates are selected based on their performance in subject-specific knowledge and teaching aptitude exams. The evidence provided by these papers is mixed, documenting positive or no effect on test scores of students from Mexico and Ecuador. Our paper contributes to this literature by showing that well-intended teacher selection systems, that rely heavily on scoring schemes, could unintentionally lead to worse student outcomes when other important factors in the teaching production function, such as experience, are less rewarded.

Similar to us, Ome (2012, 2013) and Brutti and Sánchez-Torres (2022) study the Colombian reform that introduced a centralized merit-based system to hire new teachers. The estimation strategy of these papers exploits the variation that results from having, within the same school or school-subject, teachers hired through the new merit system and teachers employed before the reform was implemented. Ambiguous effects are documented. While Ome (2012, 2013) finds negligible effects on student test scores, Brutti and Sánchez-Torres (2022) find evidence of a small positive impact.

However, these results are potentially biased due to strong identification assumptions, namely that the share of teachers hired under the new regulation is orthogonal to unobserved factors within schools or school-subject.[1] Unlike these papers, we provide evidence regarding the aggregate effects on the education market that stem from the reform implementation. We also rely on weaker identification assumptions and provide evidence on validity.

Our paper is also related to empirical work regarding the effects of teacher quality, commonly measured by teacher value-added measures. Research in this area has found consistent evidence that teacher quality explains a significant fraction of the variation in students' academic performance, education attainment, and adulthood outcomes like savings, wages, and even participation in illegal activities (Rockoff, 2004; Rivkin, Hanushek and Kain, 2005; Hanushek, 2011; Chetty et al., 2011; Hanushek and Rivkin, 2012; Chetty, Friedman and Rockoff, 2014; Araujo et al., 2016; Jackson, 2018; Rose, Schellenberg and Shem-Tov, 2022). A common finding in this literature is the little predictive power that teachers' observable characteristics, such as education level or certifications, may offer to predict future teaching effectiveness (Rivkin, Hanushek and Kain, 2005; Hanushek and Rivkin, 2006; Jackson, Rockoff and Staiger, 2014; Araujo et al., 2016). According to the evidence, while teachers' education does not explain students' learning in most school systems, it is usual to find that teachers are less effective during the first few years of their careers (Hanushek and Rivkin, 2006). Moreover, students exposed to newly hired teachers can suffer negative learning impacts (Staiger and Rockoff, 2010).

The relationship between teacher quality and characteristics that are difficult to observe by education authorities has also been analyzed in recent research. Araujo et al. (2016) offer evidence using an experimental setting and rich information on teachers. These authors find that children assigned to teachers with less than three years of experience learn less than the average student. They also document how other char-

[1] In addition to potential bias of the estimates shown in Brutti and Sánchez-Torres (2022), their main results condition on teacher experience, an attribute of the staff at public schools that was affected by the reform.

acteristics of teachers – including skills, measured by IQ scores and the Big Five personality traits, and tenure – are not correlated with student performance in Ecuador. Additional evidence from Ecuador also suggests that students assigned to teachers with higher performance at an exam evaluating knowledge and pedagogical abilities do not experience a positive effect on learning (Cruz-Aguayo, Ibarrarán and Schady, 2017).

The rest of the paper is organized as follows. In Section 2, we contextualize our work, describing the Colombian education system, the teacher labor market, and the reform of the teacher hiring policy. In Section 3, we describe the data sources we use and present descriptive statistics. In Section 4, we show evidence of the reform's effect on different attributes of the teaching staff. The paper's main results are presented in Section 5. First, we introduce the research design and the identification assumptions we rely on to interpret our results as causal effects of the merit-based hiring system. Then, we present evidence of the reform's effect on students' academic outcomes. In Section 6, we discuss how the evidence we found aligns with previous work on teacher quality and provide policy recommendations. Section 7 concludes.

2 Education System, Teachers and Hiring Reform

This section presents relevant background about the Colombian education system, the teacher labor market, and the reform of the teacher screening and hiring system implemented in 2005. We focus on the government's effort to attract higher-skilled teacher applicants and the details and changes introduced by the new centralized hiring system. Finally, we describe the large increase in teacher turnover following the reform and the role of temporary teachers within Colombia .

2.1 The Colombian Education System

Colombia is a middle-income country where school enrollment rates have grown dramatically over the past several decades. By 2010, Colombia's elementary educa-

tion reached near universal enrollment, while secondary education enrollment rose from 35 to 77 percent in the prior four decades (Bassi, Busso and Muñoz, 2015). Although enrollment levels have increased, students' learning, as measured by student test scores, has seen little to no improvement. Colombian students' 75th percentile score on international standardized exams lies well below the 25th percentile of OECD country students. This low performance has been persistent over time and has pushed Colombia to rank low among participating countries.[2] This "learning crisis" is common in much of the developing world but is more pronounced among Latin American and African countries (World Bank, 2018).

In Colombia, schooling is divided into: i) preschool or kindergarten; ii) elementary school for grades 1 to 5; iii) lower secondary for grades 6 to 9; iv) upper secondary or high school for grades 10 and 11; and v) post-secondary or tertiary education, which is offered as vocational programs of two- and three-years or as bachelor's programs of four- and five-years. Education in the country is provided by both public and private schools. Private institutions represent an important share of the education supply at all levels, educating almost 30 percent of high school students and 69 percent of post-secondary students. Parents and students face considerable differences in tuition and quality when choosing between a private or public school. While public schools are free, private schools require a tuition payment that can vary substantially.[3] However, on average, private school students obtain higher test scores on standardized exams than their peers in public schools.

Every year, before high school students graduate, they take a standardized exam evaluating their knowledge in different subjects.[4] This *high school exit exam* is high stakes for several reasons. First, the exam is mandatory to enroll in any higher educa-

[2]Colombia ranked 47 out of 58 countries that took the Programme for International Student Assessment (PISA) reading exam in 2009, 54 out of 62 in 2012, 55 out of 72 in 2015, and 58 out of 77 in 2018. Rankings in mathematics and sciences show a similar pattern, where Colombia is among the lowest ranked countries.

[3]In 2014, the annual tuition among private secondary schools could range between a few hundred and 16,000 US dollars (Las 2 Orillas, 2014).

[4]The exam is known as *Saber 11* (formerly, ICFES exam). Five subjects have been consistently evaluated across time in the exam: reading, mathematics, natural sciences (i.e., physics, chemistry, and biology), social sciences, and English proficiency. Subject exams in history, philosophy, and geography have also been administered in some years.

tion institution, so it is taken by over 95 percent of all seniors in any given year. Second, schools care about their position in the annual ranking that the government announces based on the average performance of their students. Therefore, schools usually prepare their students during their senior year using material made available by the exam authority, the Colombian Institute for the Assessment of Education (ICFES). Students can also study on their own using such preparation material, or they can enroll at private institutions that provide instruction to take the exam.

During their senior year in high school, students consider whether or not they will apply to college. College application requires that students choose a major right away, so they seek admission to a specific college-program rather than to college itself. College programs usually set cutoffs based on the overall score on the high school exit exam and, depending on the field, some programs may even set minimum scores on the subject exams.[5]

2.2 Teacher Labor Market and the Merit-based Hiring Reform

Becoming a teacher in Colombia requires either holding a college degree or a pedagogy diploma granted by a Normal school.[6] Most of the supply of teachers at public and private schools is provided by college graduates from education majors,[7] who are more frequently women, individuals who grew up in families of low socioeconomic status, and students with low performance on the high school exit exam.

In Colombia, teaching positions at public schools are attractive for monetary and non-monetary reasons. The entry-level salary offered to college graduates with little or no experience is about ten percent higher than the average earnings of graduates with an education degree who find formal employment elsewhere, and it is only two

[5] Admissions to some of the most competitive public universities require taking also an admission exam designed by each school. Nonetheless, students must also submit their high school exit exam scores as part of their application.

[6] Normal schools (Escuelas Normales Superiores, in Spanish) are high schools where students take pedagogy classes as part of their curriculum. Graduates from these schools can take an additional year of classes to obtain a pedagogy diploma certifying that they are prepared to teach children in preschool and elementary grades.

[7] Between 2004 and 2019, the share of teachers with an education college degree ranged from 45 to 52 percent at public schools, and from 48 to 59 percent in private ones.

percent lower than the average of new graduates from the field of business and accounting.[8] These positions additionally offer high job stability, annual bonuses, and coverage in a social and health insurance system especially designed for public school teachers.[9]

The Colombian government regulates most aspects of a public school teacher's career including remuneration, hiring, promotions, and retirement. Before 2002, the rules and procedures concerning these aspects were defined in Decree 2277 of 1979. Under this regulation, the process of hiring new teachers was decentralized, and started with an annual assessment by city mayors to determine the number of vacancies in schools within their jurisdiction. However, besides some education and experience requirements to appoint teachers, there were no standardized criteria on how local and regional authorities were to screen candidates.[10] This lack of clarity led to speculations that the allocation of vacancies was susceptible to political interests (Bustamante, 1996; Duarte, 2001, 2003). Salaries were also regulated and followed a fourteen-level career ladder, where each step was linked to specific years of experience and education. For instance, teachers with no experience but holding an education college degree were assigned to level seven at the time of hiring. Promotions were tied to additional years of experience, completion of pedagogy courses, and graduate education attainment.[11] Public school teachers' jobs were highly stable, and only severe misconduct could prevent a teacher from working until retirement. Teacher and student performance played no role in promotions or tenure.

The hiring and employment conditions of public school teachers were reformed by Decree 1278 in 2002, which introduced a merit-based system aiming to improve

[8] Recent college graduates employed at public schools had a monthly wage of 629 US dollars in 2010. Appendix Table 1 displays the average monthly earnings of graduates from different fields employed in the formal sector.

[9] Representative survey data suggest that public school teachers are more satisfied in their jobs, work fewer hours a week, and have higher salaries than other teachers (see Appendix Table 2).

[10] In 1989, the government enacted Decree 1706, establishing that all teachers must be hired through a public call to fill vacancies, although no details were given on how local authorities were to screen candidates.

[11] For example, teachers must have taught for three years to be promoted from level seven to eight on the career ladder. To further progress to level nine, teachers were required three additional years and a certain number of pedagogy courses.

the quality of public education. The reform linked hiring, tenure, and promotion to a teacher evaluation process. Under this regulation, a new hiring process starts when the government has determined the number of teaching vacancies available nationwide and announces a *public call* to fill such positions. Applicants must take a written exam measuring two sets of skills: (i) knowledge of the subject that each candidate wants to teach, and (ii) teaching aptitude (including pedagogy and subject-specific knowledge). A minimum score of 60 out of 100 in both components of the exam is necessary to continue in the process.[12] Only the applicants that perform well in the written exam are cleared to be interviewed individually.[13] Finally, candidates obtain a weighted score based on the written exams and the interview results. To fill teaching position vacancies, the government first computes a general ranking of approved candidates according to their individual score. Then, in a public audience, top-ranked applicants choose their most preferred position or school among those available. This allocation process continues in descending order until all remaining vacancies have been considered by teachers lower in the ranking.[14]

Public teachers hired in the post-reform period are assessed by means of a trial period, annual evaluations, and written exams to obtain promotions. Novice teachers are not immediately granted a permanent contract. Instead, school principals must submit a report evaluating these teachers' performance after at least four months in the position.[15] However, teachers rarely fail their trial period evaluation (Garcia et al., 2014; Forero and Saavedra, 2019). Past the trial period, teachers are evaluated annually by their principals, and their continued employment is conditioned on not fail-

[12] A psychometric test is also included along with the knowledge and aptitude written exams, but candidates are not required to pass a minimum score. However, the scores from the psychometric test are considered for computing each candidate's overall performance in the hiring process.

[13] A third party, commonly a university, is hired by the government to conduct the interviews and verify that each candidate holds the education degree and the experience required for the teaching position. Candidates get higher scores in this stage of the hiring process based on their interview performance, experience, and education.

[14] The process can finish either because there are no more vacancies available or because all the eligible teachers have been already matched. It is possible that certain vacancies remained unfilled after the public audience.

[15] Principals collect information on a teacher's performance in academic aspects (such as knowledge of the teaching subject, class planning, didactic strategies, and evaluation methods), school administrative duties, and the teacher's involvement with students' families and environment. To approve the trial period, teachers must obtain a minimum score of 60 out of 100.

ing two consecutive evaluations. However, this mechanism is ineffective at firing low-performing teachers since annual evaluations are assumed as a means to provide feedback rather than a system to monitor performance.[16] Finally, the reform tied promotions to performance on a written exam evaluating teachers' knowledge.[17]

The reform also increased salaries of public school teachers. College graduates with no prior teaching experience earn entry-level wages 12 percent higher post-reform, a gap that can account for an earnings premium of 34 percent after 15 years of experience. Appendix Figure 1 plots the wage-experience profiles for college graduates hired pre- and post-reform.[18] Such an increase in wages was intended to attract a higher-quality pool of teacher candidates to fill the vacancies at public schools.[19]

Six public calls to fill vacancies nationwide have taken place since the reform was enacted in 2002 (see Figure 1a).[20] The first was carried out in 2004 when the government announced that 44,596 teachers were needed. The second and third calls were made from 2005 to 2006 (with 21,868 vacancies) and from 2006 to 2007 (with 12,788 vacancies). A fourth call to fill 23,524 vacancies was announced in 2009, and approved candidates started filling these positions in 2010. The fifth call began in 2012, but took the government until late 2015 to start appointing candidates for the 17,941 vacancies that were initially announced.[21] More recently, in 2021, the government announced a new public call to fill approximately 29,000 teaching vacancies. The public calls have been oversubscribed; applicants in each hiring process have more than doubled the

[16] Principals evaluate teachers based on (i) primary functional abilities, such as teaching and handling administrative duties, and (ii) behavioral skills, such as leadership, communication, interpersonal relations, and teamwork abilities. Teachers require a score of at least 60 out of 100 to approve.

[17] In 2014, the exam was replaced by the evaluation of a class recording. In both cases, teachers must have three additional years of experience (after their being hired or their last promotion) and obtain a score above 80 out of 100 to be promoted.

[18] For this figure, we assume that that teachers hired post-reform are promoted every five years.

[19] Teachers hired pre-reform can participate in each new hiring process on equal conditions as any other applicant. However, only a few decide to do so, given that a large share of them have many years of experience and are at the top of the wage ladder (Ome, 2012, 2013).

[20] The government has also made special smaller calls to fill vacancies in distant areas. Two of them, in 2006 and 2012, to fill positions at a small number of public schools that provide education to ethnic communities (i.e., indigenous and afro-descendants). The most recent call was made in 2018 to hire candidates willing to teach in rural schools from municipalities that have suffered the consequences of the armed conflict in the country.

[21] Appendix Figure 2 shows the entry dates of successful applicants who start their four-month trial period after being hired.

number of vacancies, suggesting that teaching positions at public schools are attractive and competitive (see Figure 1b). In addition, due to the large-scale implementation of the reform, nearly half of all public school teachers were hired under the new regulation by 2015 (see Figure 2).

2.3 Temporary Teachers

The decree 1278 of 2002 introduced a nationwide public call for eligible candidates instead of periodic annual hiring processes. However, since the lists of eligible candidates were only updated every few years, the system was unable to quickly fill teacher vacancies that arose due to retirements, departures, or unfilled positions. To address the short-term demand for teachers in schools and ensure continued education, temporary teachers were employed until the next wave of teachers selected from the public call could replace them. Between 2007 and 2015, temporary teachers accounted for a significant proportion of the teaching staff, comprising 12 to 20 percent of all positions. Despite their contracts not specifying a fixed term, around two-thirds of temporary teachers remained in their positions for at least two years until they were replaced by eligible candidates selected through the most recent centralized public call. Temporary teachers were more prevalent in remote and low-income areas, where the merit-based system has been less effective at filling vacancies due to lower demand for such positions (Garcia et al., 2014; Forero and Saavedra, 2019; Bonilla-Mejía et al., 2018).

3 Data

We use Colombian administrative data from three main sources. First, we rely on census data of public school teachers from 2007 to 2015, collected through the human resources system of the Colombian Ministry of Education, known as *Humano*.[22] The

[22]Principals from all public schools submit teachers' information twice a year using Humano, as mandated by Resolution 166 of the Ministry of Education. This census-like administrative data is recorded in cross-section files known as Anexo 3A.

data includes unique identifiers of about 400,000 teachers as well as the exact date when they were hired. This allows us to retrospectively create a longitudinal dataset with information on teachers' experience and career development over time.

Second, we use test score data of all students enrolled at public and private high schools. The data corresponds to over nine million students who took the high school exit exam in the second semester of each year, between 2000 and 2019.[23] The exam, known as *Saber 11*, assesses the knowledge of nearly all senior students in different subjects such as mathematics, reading comprehension, English proficiency, social sciences, and natural sciences. We standardize the subject- and global-test scores within cohorts.[24] In addition to performance on the exam, these data include students' demographic characteristics such as gender, age, and household socioeconomic stratum that proxies family income.[25] The information is administered by the exam authority, ICFES, and is available online.[26]

Third, we use census-like administrative records of students enrolled in college between 1998 and 2016. The Ministry of Education collects this data through a system known as *Spadies*, which is used to monitor higher education dropout and graduation rates across time. The data corresponds to over five million students, who can be uniquely identified in the data. The information includes the year and semester when students enrolled in a college program, an indicator variable if they have graduated, and the date of graduation. It also records the student's percentile scores in the high school exit exam and socioeconomic information at the time of admission.

We merge the college records to (i) the census of public school teachers and (ii) the test score data of high school students. The link between the teacher census and the college records allows us to use the percentile score of teachers in the high school exit

[23] Most schools in the country administer the exam during the second semester of the year. Only the most elite private schools –and a negligible portion of the public schools– administer the exam in the first semester.

[24] Each cohort corresponds to students who took the exam on the same date.

[25] Residential properties in Colombia are assigned a "socioeconomic" index (or stratum) from one to six depending on the neighborhood were the property is located. A higher index indicates that the neighborhood has more access to amenities and public services. The index proxies family income.

[26] We obtained access from the Ministry of Education to restricted data that includes the identifiers of students who took the exam between 2002 and 2015. This allows us to merge a large subset of the test score data to college administrative records.

exam as a proxy of skills for teachers hired before and after 2005, when the Colombian government conducted the first public call to fill vacancies. The link with the high school exit exam allows us to observe college enrollment and graduation for multiple cohorts of students who took the exam before and after the reform of the teacher hiring system.[27]

Table 1 presents summary statistics of our sample of students, separated by those who attended public and private high schools. Both in public and private schools, students are 18 years old on average, and slightly more than half are women. Students in public schools come from families with poorer socioeconomic backgrounds, as shown by their mother's education, socioeconomic stratum, and family size indicators. In addition, a higher share of private school students attend a full-day schedule, and are located in the country's urban zones and main cities.

4 Effects on Teachers' Skills and Experience

The introduction of a merit-based hiring system increased the average skills of public school teachers but also decreased their average experience. Figure 3 shows evidence of a sharp increase on the performance of teachers hired after 2005 on the exit exam scores of the teachers when they were in high school, our measure of cognitive skills. Figure 3a displays the mean percentile score of teachers hired in each quarter from 1995 to 2015. A discontinuous increase in the performance of newly hired teachers is observed in 2005, the first year the Colombian government implemented the new hiring reform. Teachers hired after 2004 had test scores that were 17 percentile points higher than previously hired teachers. Figure 3b plots the inter-quartile range and the average high school exit exam score of active teachers between 2002 and 2015 separat-

[27] College records can only be linked to test scores data of seniors who took the high school exit exam between 2002 and 2015. Using the information when students take the exam and the year-semester when they start college, we compute enrollment rates for different time windows to make cross-cohort comparisons: immediate, one-year, and two-year enrollment. We use a six-year time window for college graduation rates.

ing those that were hired pre- and post-reform.[28] The 25^{th} percentile score of teachers hired post-reform is similar to the median score of teachers hired pre-reform, suggesting that the reform changed the pool of teachers hired by selecting higher-skilled individuals.

As a result of modifications made to the pool of teachers, there was a decrease in the amount of teaching experience possessed by public school educators. Figure 4a illustrates how the teaching experience at public schools went from a bimodal distribution to a distribution with three modes post-reform. The new mode corresponds to teachers with less than five years of experience, who became a significant share of the teaching staff. Similarly, Figure 4b shows that the share of teachers with less than five years of teaching experience quickly increased from just less than 10 percent in 2002 to 30 percent by 2008. As the first wave of new teachers gained experience, the fraction of inexperienced teachers in Colombia fell to 20 percent were it remain fairly stable over the later part of our sample period.

4.1 The Merit-Based Reform and Teacher Turnover

The merit-based reform decreased public teachers' level of experience by unexpectedly increasing teacher turnover. A constant amount of turnover is expected every year as teachers age, retire, or switch occupations. Any hiring scheme for teachers deals with this set of issues as novice teachers have to replace those who leave the profession. The Colombian 2002 reform, however, created a mechanism in which novice teachers replaced not only those who retired or left the profession (i.e., the *expected* turnover) but also those teachers who were temporary but had already accumulated several years of teaching experience (i.e., the *unexpected* turnover).

In the decade prior to the reform, a very small number of permanent teachers where hired (see Appendix Figure 6), and vacancies were largely filled with temporary teachers. In early 2003, before the first public call for teachers, over 50,000 teach-

[28] We only observe teachers working between 2007 and 2015, but we impute the mean and interquartile range retrospectively using teachers' hiring dates.

ers were temporary teachers from the old hiring system – or 17 percent of all teachers. Three public calls occurred between 2004 and 2008, and nearly 80,000 new permanent teachers were hired (see Figure 1a). By 2008, three years after the first public call of teachers, over 45,000 of these temporary teachers had been replaced by novice teachers. This resulted in a large turnover of the stock of teachers post-reform and in an increase of 20 percent in the fraction of teachers with less than five years of experience. In nearly 17 percent of classrooms, novice teachers replaced temporary teachers who were considerably older and had already accumulated multiple years of experience (see Appendix Figures 5c and 5d).

This pattern in teacher turnover continued after the first public call. Temporary teachers were replaced after every public call, and public schools continue to lose a pool of experienced teachers who were replaced by an entrant wave of novice teachers (see Appendix Figure 5a and 5b). Appendix Figure 4 shows that the number of temporary teachers increased between public calls, and then dropped when the next wave of new novice teachers was hired. This helps explaining why the share of teachers with less than five years of experience has remained at around 20 percent after 2010. Even though the reform was intended to attract and select more skilled teachers, it also promoted more frequent teacher turnover, exposing students to teachers with less experience.

The reform seem to have only affected average teachers experience and skills, whereas it had no effect on other characteristics of the pool of public teachers. We present the evolution of other characteristics, before and after the reform, in Appendix Figure 7. We do not observe any changes in the share of female teachers, the percentage of teachers with a college degree, or in the average age when teachers are hired. In addition, while the reform increased teacher turnover, Appendix Figure 3 shows that there was little change in the number of teachers in Colombia during this time period. However, since student population were declining, student-teacher ratios were monotonically declining during this time period.

The merit-based reform affected fundamental inputs for students learning, such as

teachers' skills and experience, while leaving other teacher characteristics unaffected. This setting suggest that the reform could have had effects on students learning as novice teachers can have a negative effect on students' academic achievement (Rivkin, Hanushek and Kain, 2005; Hanushek and Rivkin, 2006; Araujo et al., 2016), while the effect of higher-skilled teachers on student learning can be ambiguous (Araujo et al., 2016; Estrada, 2019; Cruz-Aguayo, Hincapie and Rodriguez, 2020). We address the effect of the reform on students in the following section.

5 Effects on Student Academic Achievement

In this section, we analyze the effects of the reform of the merit-based hiring system on student outcomes. We first describe our empirical strategy and the assumptions we rely on to interpret our results as the causal effect of the reform. Subsequently, we present and interpret the results obtained from this empirical approach.

5.1 Empirical Strategy

Our empirical strategy identifies the effects on student outcomes of the new merit-based teacher hiring system. We exploit the fact that the new teacher hiring system was implemented only for public schools and did not directly affect private schools. This setting allows us to identify the causal effect of the policy using students enrolled at private schools as a counterfactual group. Given that students at public and private schools are initially different, we employ a difference-in-differences strategy that eliminates pre-existing differences. Formally, we estimate:

$$Y_{ist} = \mu_t + \mu_s + \sum_{\tau \neq 2004}^{T} \delta_\tau \times \mathbb{1}[\tau = t] \times \text{Public}_s + X_i'\gamma + \varepsilon_{ist}, \tag{1}$$

where Y_{ist} represents the outcome of student i, who graduates from high-school s in year t. Our main outcomes of interest are the student's overall score in the high school exit exam and the likelihood of college enrollment and graduation. The variable Public_s is an indicator variable for whether school s is a public school. The param-

eters of interest are δ_τ, $\tau \in \{2000, ..., 2003, 2005, ...\}$, which represent dynamic event study effects of the merit-based hiring system. We control for between school and year variation by including school fixed effects, μ_s, and year fixed effects, μ_t. Additionally, we condition on a vector of individual characteristics, X_i, that includes the student's age, gender, a socioeconomic stratum proxy for family income, and an indicator if the student takes classes in the morning, afternoon, at night, or on weekends. Our most saturated specification also includes municipality linear trends. Standard errors are clustered at the school-year level.

Our model estimates will be consistent if the trends in academic outcomes between students from private and public schools would have remained parallel in the absence of the merit-based hiring reform in 2005. This strategy does not apply any staggered adoption or continuous treatment. Therefore, our parameters can be interpreted as causal under a classical parallel trends assumption in absence of other policy changes that may have affected public schools simultaneously, confounding the reform's effect. While the counterfactual parallel trend assumption cannot be observed, the dynamic effects estimated in our event study strategy allows us to test for parallel trends prior to the reform and provide some evidence for the validity of this assumption.

Unlike previous work studying the effects of the reform, we focus on estimating its aggregate unconditional impacts, stemming from changes in teacher composition at public schools. Our strategy differs from Brutti and Sánchez-Torres (2022), who exploit school-subject variation in the share of newly hired teachers post-reform. Their estimator accounts for potential changes in the characteristics of the teaching staff by controlling for teachers' average age, experience, and education level. However, some of these characteristics were affected by the inflow of new teachers hired through the centralized system, such as experience. Also, they impose the strong assumption that vacancies across time are orthogonal to unobserved factors related to student learning. This assumption may be broken since endogoneity may arise from the fact that successful candidates participating in each merit-based hiring process are allowed to

fill a vacancy at their preferred school.[29]

5.2 Results

We present two main sets of findings on the unintended consequences of the merit-based teacher hiring system reform in Colombia. First, we document negative effects of the reform on test scores. Second, we document the reform's negative effect on the likelihood of enrolling and graduating college.

Effect on test scores – We start by estimating the reform's effect on students' overall performance on the high school exit exam. We define overall performance as the average score on the five subject exams: reading comprehension, mathematics, natural sciences, social sciences, and English proficiency. Figure 5 displays the dynamic effects of the reform on overall performance in the exam.

Five main observations are relevant to point out. First, the results are very similar regardless of the specification used. Second, the gap in test scores between public and private students remains stable during the pre-treatment period (from 2000 to 2004) and appears to follow a parallel trend. This supports the validity of our identification strategy. Third, the post-reform period estimates indicate that public school students obtained lower scores soon after the reform was put in place. Negative effects on test scores of public school students start to appear in 2005 as the first new teachers were being hired. This negative effect continued to grow until 2008 in which the test scores of students in public schools were 0.12 standard deviations lower than private school students relative to before the reform was implemented. Fourth, the negative effect appears to stabilize to around -0.10 standard deviations from 2007 to 2013 (with the exception of 2011). Lastly, after 2013 the negative effect of the reform appears to diminish and converges to about -0.05 standard deviations, or about half of the effect five years after the reform.

We present static difference-in-differences point estimates on overall and subject

[29] Ome (2012, 2013) follows a similar strategy to Brutti and Sánchez-Torres (2022), but instead of using within-school-subject variation, this author exploits within-school variation. Both approaches share the similar concerns.

test scores in Table 2. On average, the overall performance of students at public schools compared to those at private schools decreased by about 8.2 percent of a standard deviation after the merit-based teacher hiring system was implemented. This effect is equivalent to the negative impact of being taught by a first-year teacher (Staiger and Rockoff, 2010) and to a one standard deviation decrease in teacher quality (Chetty, Friedman and Rockoff, 2014; Petek and Pope, 2023). These results are mostly driven by large negative effects on mathematics and English proficiency (ranging from 14 to 16 percent of a standard deviation), although negative effects are also estimated for all other subjects (ranging from 2.6 to 6.6 percent of a standard deviation).[30]

Effect on college outcomes – In Figure 6, we present the dynamic effects of the reform on college enrollment and graduation. A negative effect that persists over time is observed for both outcomes. The result on enrollment captures the impact on going directly from high school to college, given that our outcome only records a value of one for students who enroll in college within the next six months after completing high school.[31] The dynamic effect we observe for college enrollment follow a similar pattern to what we find for test scores although the largest negative effect occurs somewhat later in 2012. The negative effect of the reform on college enrollment starts to occur in 2006 and continues to grow until 2012 in which public school students are 5 percentage points less likely to enroll in high school. This negative effect then starts to converge back to zero and by 2015 only reduces college enrollment by 3 percentage points.

The reform also negatively impacted students' likelihood of graduating from college, as shown by Figure 6. The initial negative impact of the reform on college graduation is delayed until the cohort of students who graduated high school in 2006 has gone through college. The negative impact found on the cohort of students graduating high school in 2006 continues to grow over time and by the 2009 cohort (the last cohort we are able calculate 6-year graduation rates) the reform decreased the likelihood of a public school student graduating from college by over 2 percentage points.

[30] In Appendix Figure 8 we present dynamic estimates of the effect by subject exam scores.

[31] In Appendix Figure 9, we present results using wider time windows to define college enrollment – specifically, one-year and two-year enrollment rates. Similar results are obtained for these alternative definitions of enrollment.

Since many tertiary institutions in the country offer admission based on the applicant's performance on the high school exit exam, the effect on college enrollment and graduation may be largely driven by the negative impact of the reform on test scores. While we are unable to differentiate other ways the reform may have impacted student's outcomes, the reform, in particular high levels of exposure to novice teachers, may have also impacted students college outcomes through non-test score channels.

Table 3 summarize the difference-in-differences results on college outcomes by estimating static point estimates. As can be seen in the first six columns, for both immediate college enrollment rates and 2-year post high school college enrollment rates, the reform decreased enrollment by just over three percentage points. For college enrollment within 6-months, this estimated negative effect is equivalent to a 20 percent decline in enrollment. With the base rate of college enrolled doubling from 16 percent to 32 percent when using a 2-year enrollment window, the negative effect on college enrollment shrinks to 10 percent. Similarly for college graduation, the reform decreased the likelihood of a public school student graduating from college by 0.9 percentage points or 10 percent.

5.3 Teaching Experience and Student Outcomes

The 2002 reform negatively affected students' learning, even though it selected teachers with higher cognitive skilled. This negative effect of the reform may be driven by the increased exposed to less experienced teachers. We provide two pieces of evidence that link the reduction in students' learning with increased student exposure to novice teachers after the reform.

First, we observe that the dynamic effects of the reform on students' test scores closely mirror the change in the fraction of teachers with less than five years of experience (compare Figure 4b and 5). Between 2004 and 2008, public schools received a large influx of novice teachers, going from 10 to 30 percent of all teachers. As such, a significant share of students started being taught by teachers with little to no ex-

perience.[32] During this same time period the test scores of public school students relative to private school students declined by a little over 0.10 standard deviations. As the fraction of novice teachers remained fairly stable between 2007 and 2010, the estimated negative effect of the reform remained fairly stable at around -0.10 standard deviations. Lastly, as the fraction of novice teachers fell half way back to pre-reform levels after 2013 (20 percent instead of 10 percent), we also see that the negative effect of the reform is approximately cut in half to -0.05 standard deviations. The results on college enrollment also mirror the fraction of novice teachers in public schools during this time time period, as evidenced by Figure 6a. The mirrored patterns between the fraction of novice teachers and the dynamic effects of the reform suggest that teaching experience may be playing a prominent role in explaining the negative effects of the reform.

Second, we find that the effect on students' learning was stronger among public schools that were more exposed to novice teachers after the reform. To do this, we calculate the faction of teachers in each school with less than 5 years of experience in 2008. We then interact the faction of novice teachers with an indicator for being in a public school and an indicator for the year being after the reform along with the same controls in Equation (1). Table 4 reports the coefficient on the interaction between the public school and after reform indicators and the coefficient on the triple interaction between public school, after the reform, and the fraction of novice teachers. For the fully saturated model in the third column of each panel, the coefficient on the interaction between the public school and after reform indicators is only slightly smaller than the effect found in our main specification in Table 2 with a negative effect on overall test scores of 0.075 and 0.082 standard deviations, respectively. The coefficient of the triple interaction reported in the first row shows that the effects are 40 percent larger for schools with a high fraction of novice teachers in 2008. While these results give the heterogeneous treatment effect of the reform on student test scores for schools with

[32]Students were also exposed to an increase in teacher turnover. However, as Staiger and Rockoff (2010) point out, the primary cost of teacher turnover on student achievement stems from students being taught by novice teachers and not from firing and hiring new teachers.

a high or low fraction of novice teachers in 2008, these results should be taken with caution since the variation in exposure to novice teacher is not exogenous variation. The fraction of novice teachers in 2008 could partly be an endogenous response to the reform. In addition, the fraction of novice teachers in 2008 could also be correlated with other school, student, or location characteristics. This could potentially explain why the heterogeneous effects found in the first row are highly sensitive to the set of controls included.

In addition to test scores outcomes, Table 5 reports the heterogeneous effects of the reform on college outcome by the fraction of novice teachers in 2008. Somewhat smaller than our main results, we find that the reform reduced immediate college enrollment, 2-year enrollment, and graduation by 0.026, 0.027, and 0.007 percentage points, respectively, for schools with few novice teacher in 2008. However, the effect on college enrollment was twice as large for students in public schools with a high fraction of novice teachers and three times larger for college graduation. While the same caveats remain for these results, they suggest a larger effect for students in schools that were exposed to more novice teachers after the reform. Dynamic effects estimated for schools with a high, medium, and low fraction of novice teachers in 2008 are shown in Appendix Figure 10.

These two pieces of evidence suggest that student learning could have been affected by students' increased exposure to teachers with lower levels of experience and to a larger degree in schools with greater levels of teacher turnover. Poorly designed policy that affect teacher retention and turnover may decrease student learning despite the selection of a higher cognitively skilled pool of public teachers.

6 Discussion and Policy Recommendations

The negative effect of the merit-based hiring system on students' academic outcomes informs policymakers in two main areas. First, the results provides direct insight into important design decisions for teacher hiring systems. Second, the results

provide broader implications for many educations policies that influence teacher retention and turnover.

During the last two decades, a large number countries, particularly in Latin America, have introduced merit-based policies regulating the process for hiring teachers similar to the reform in Colombia (Cruz-Aguayo, Hincapie and Rodriguez, 2020; Elacqua et al., 2018). These systems typically used a centralized hiring system in which public school vacancies are allocated among candidates based on certain criteria such as passing a standardized exam. Typically, these policies heavily weight a few ex-ante pieces of information about teachers – such as teachers' own cognitive ability as measured by test scores – at the expense of other ex-ante pieces of information – such has prior teaching experience. As such, these policies typically have their intended effect of increasing the average pre-college test scores, and likely the cognitive skills, of newly hired teachers. However, these policies also lower less weight teacher characteristics such as experience. Given that novice teachers are less effective instructors (Hanushek, 1971; Rivkin, Hanushek and Kain, 2005), and teachers' skills can only explain a small fraction of the variation in teacher quality (Hanushek and Rivkin, 2006; Araujo et al., 2016; Cruz-Aguayo, Ibarrarán and Schady, 2017), these policy more heavily weight less important teacher characteristics at the expense of more important characteristics.[33] As such, not only will these new hiring policies likely fail to improve teacher quality and, consequently, student outcomes, these policies are likely to negatively impact teacher quality and student outcomes. Therefore, education reforms and hiring systems should carefully evaluate the ex-ante information they use when designing their policies. For new teachers, especially those with no prior teaching experience, policies may want to altogether avoid ex-ante information and focus on using ex-post measures to improve teacher quality, such as only offering permanent contract based on teachers' effectiveness.

[33]Previous literature in this area has shown that identifying candidates that will become high-quality teachers is difficult (Rockoff et al., 2011). Although some screening systems offer potential positive results (Goldhaber, Grout and Huntington-Klein, 2017; Jacob et al., 2018; Estrada, 2019), the evidence is still limited. In addition, in Latin America, estimates of the effects of merit-based screening systems by Estrada (2019) and Brutti and Sánchez-Torres (2022) contrast with findings by Cruz-Aguayo, Ibarrarán and Schady (2017), Ome (2013) and our own evidence.

In addition to better understanding what ex-ante information policymakers should emphasize when designing hiring polices, our results also inform broader education policies that influence teacher retention and turnover. Our results show that the large shock of replacing experienced teachers with novice teachers (even novice teachers with much higher cognitive skills) in the first few years of the reform negatively impacted students. This large shock on teacher experience arose from the fact that Colombia first filled vacancies with inexperienced temporary teachers for several years then replaced them with a public call for permanent teachers. Instead of filling a vacancy with a teacher who would likely teach for several decades, a vacancy was first filled with a temporary teacher and then refilled with a permanent teacher. Instead of each vacancy requiring the training of only one new teacher (and students experiencing learning losses from only one teacher in the early part of the learning curve), each vacancy required two teachers going through this process. This occurred as a large shock in first few years of the reform and is the likely reason for the initial large negative effects of the reform. After the initial shock, this underlying double turnover for each vacancy continued at a lower, but stead rate. Each vacancy required a temporary teacher to fill the position for one to five years and then a permanent teacher fills the position. Unfortunately, students are then more likely to be taught by a teacher on the early part of the learning curve. This underlying double filling of vacancies potentially explains why after the initial large negative effect of the reform, the estimated dynamic effect does not fully converge back to zero (see Figure 5). These results imply that an important way of improving students outcomes is by keeping teachers in the profession for extended periods of time and therefore reducing the number students who are taught by novice teachers in any given year. While teacher turnover at the school or district level may play an important role, these results shine a light on the importance of teacher turnover at the level of the teaching profession.

7 Conclusion

Teachers are the most relevant factor for human capital development in education systems (Chetty, Friedman and Rockoff, 2014; Hanushek and Rivkin, 2006; Rivkin, Hanushek and Kain, 2005). As such, education authorities across districts and countries implement policies to improve teacher quality and, consequently, student outcomes. For legal and political economy reasons, policy changes to improve teacher quality typically focus on new hires rather than current instructors. However, identifying effective teachers can be a complex and challenging task, mainly because value-added measures of teacher quality are uncorrelated to teachers' observable characteristics such as education level, licensure, IQ scores, and scores rating the performance of teacher candidates from a screening and hiring process (Araujo et al., 2016).

In this paper, we study the aggregate effects of a large-scale reform that introduced a centralized merit-based system to hire new public school teachers in Colombia. The evidence we obtain from administrative data on teachers shows that two relevant attributes of the teacher's body were affected by the reform: cognitive skills and teaching experience. While cognitive skills increased sharply among teachers hired post-reform implementation (17 percentile point increase in teachers' own test scores), the reform increased the share of teachers with little to no experience by a large amount (from 10 percent to 30 percent at its peak).

To estimate the reform's effect on students' outcomes, we employ a difference-in-differences strategy leveraging the fact that the reform did not directly affect private school students. Our results show that students' test scores decreased by about 8.2 percent of a standard deviation after the implementation of the new hiring policy at public schools. We also find that the probability of enrolling and graduating from college dropped by 20 and 10 percent, respectively. These negative effects are in line with the evidence provided by the literature that suggests that i) teacher quality is not correlated with teachers' test scores or scores rating information gathered before teachers are hired (Cruz-Aguayo, Hincapie and Rodriguez, 2020), and ii) teacher quality is typically lower during the first five years of teaching (Araujo et al., 2016; Hanushek and

Woessmann, 2011).

Despite concerted effort, increased spending, and being well intended, the merit-based teacher hiring reform reduced students' academic outcomes instead of improving them. The likely reason for this was that policymakers designed a reform which heavily weighted one proxy for teacher quality – teachers' own cognitive ability as measured by test scores – at the expense of another proxy for teacher quality – teachers' level of experience. After the reform, most policymakers believe the reform had been a large success since "higher quality teachers" (as measured by teacher test score performance) had been hired and where teaching in schools. Unfortunately, a misunderstanding of which ex-ante teacher information is most important for predicting teacher quality led to the reform actually negatively impacted students. These results suggest that future education reforms and hiring systems should carefully evaluate what ex-ante information they use when designing their policies.

References

Araujo, Maria Daniela. 2022. "Measuring the effect of competitive teacher recruitment on student achievement: Evidence from Ecuador." Unpublished Manuscript.

Araujo, Maria Daniela, Yyannú Cruz-Aguayo, Pablo Ibarrarán, and Norbert Schady. 2020. "Does Test-Based Teacher Recruitment Work in the Developing World? Experimental Evidence from Ecuador." IZA Institute of Labor Economics Discussion Paper 13830.

Araujo, M. Caridad, Pedro Carneiro, Yyannú Cruz-Aguayo, and Norbert Schady. 2016. "Teacher Quality and Learning Outcomes in Kindergarten." *The Quarterly Journal of Economics*, 131(3): 1415–1453.

Bassi, Marina, Matías Busso, and Juan Sebastián Muñoz. 2015. "Enrollment, Graduation, and Dropout Rates in Latin America: Is the Glass Half Empty or Half Full?" *Economía*, 16(1): 113–156.

Bautista, Marcela. 2009. "La Profesionalización Docente en Colombia." *Revista Colombiana De Sociología*, 32(2): 111–132.

Blanden, Jo, Matthias Doeple, and Jan Stuhler. 2022. "Educational Inequality." In *Handbook of the Economics of Education*. Vol. Volume 6, , ed. John Jerrim, Sandra McNally, Anders Bjorklund, Jonas Radl, Martin Hallsten and Christopher Rauh. Elsevier.

Bonilla-Mejía, Leonardo, Erika Londoño-Ortega, Lina Cardona-Sosa, and Luis Daniel Trujillo-Escalante. 2018. "¿Quiénes son los docentes en Colombia? Características generales y brechas regionales." Banco de la Republica de Colombia Documentos de Trabajo sobre Economía Regional y Urbana 276.

Brutti, Zelda, and Fabio Sánchez-Torres. 2022. "Turning around teacher quality in Latin America: Renewed confidence and lessons from Colombia." *Economic Analysis and Policy*, 73: 62–93.

Bustamante, Guillermo. 1996. "El Concurso para Ingresar a la Carrera Docente ¿Una Evaluación Más o Una Oportunidad para Pensar la Educación." *Educación y Cultura*, 39: 24–30.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104(9): 2633–79.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *The Quarterly Journal of Economics*, 126(4): 1593–1660.

Cruz-Aguayo, Yyannú, Diana Hincapie, and Chaterine Rodriguez. 2020. *Testing our teachers. Keys to a successful teacher evaluation system.* Interamerican Development Bank.

Cruz-Aguayo, Yyannú, Pablo Ibarrarán, and Norbert Schady. 2017. "Do tests applied to teachers predict their effectiveness?" *Economics Letters*, 159: 108–111.

Duarte, Jesús. 2001. "Política y Educación: Tentaciones Particularistas en la Educación Latinoamericana." In *Economía Política de las Reformas Educativas en América Latina*. 117–142. Santiago de Chile: Centro de Investigación y Desarrollo de la Educación.

Duarte, Jesús. 2003. *Educacion Pública y Clientelismo en Colombia.* Medellin: Universidad de Antioquia.

Elacqua, Gregory, Diana Hincapie, Emiliana Vegas, and Mariana Alfonso. 2018. *Profesión: Profesor en América Latina ¿Por qué se perdió el prestigio docente y cómo recuperarlo?* Interamerican Development Bank.

Estrada, Ricardo. 2019. "Rules versus Discretion in Public Service: Teacher Hiring in Mexico." *Journal of Labor Economics*, 37(2): 545–579.

Forero, David, and Victor Saavedra. 2019. *Los 10 Pasos para Hacer de Colombia la Mejor Educada de América Latina.* Bogotá: Fedesarrollo.

Garcia, Sandra, Dario Maldonado, Guillermo Perry, Catherine Rodriguez, and Juan Esteban Saavedra. 2014. *Tras la Excelencia Docente: Cómo Mejorar la Calidad de la Educacion para Todos los Colombianos.* Bogotá: Fundación Compartir.

Goldhaber, Dan, Cyrus Grout, and Nick Huntington-Klein. 2017. "Screen Twice, Cut Once: Assessing the Predictive Validity of Applicant Selection Tools." *Education Finance and Policy*, 12(2): 197–223.

Hanushek, Eric. 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." *The American Economic Review*, 61(2): 280–288.

Hanushek, Eric A. 2011. "The economic value of higher teacher quality." *Economics of Education Review*, 30(3): 466–479.

Hanushek, Eric A., and Ludger Woessmann. 2011. "Chapter 2 - The Economics of International Differences in Educational Achievement." In *Handbook of the Economics of Education*. Vol. 3, , ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 89–200. Elsevier.

Hanushek, Eric A., and Steven G. Rivkin. 2006. "Chapter 18 Teacher Quality." In *Handbook of the Economics of Education*. Vol. 2, , ed. E. Hanushek and F. Welch, 1051–1078. Elsevier.

Hanushek, Eric A., and Steven G. Rivkin. 2012. "The Distribution of Teacher Quality and Implications for Policy." *Annual Review of Economics*, 4(1): 131–157.

Harris, Douglas N., William K. Ingle, and Stacey A. Rutledge. 2014. "How Teacher Evaluation Methods Matter for Accountability: A Comparative Analysis of Teacher Effectiveness Ratings by Principals and Teacher Value-Added Measures." *American Educational Research Journal*, 51(1): 73–112.

Jackson, C. Kirabo. 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes." *Journal of Political Economy*, 126(5): 2072–2107.

Jackson, C. Kirabo, Jonah E. Rockoff, and Douglas O. Staiger. 2014. "Teacher Effects and Teacher-Related Policies." *Annual Review of Economics*, 6(1): 801–825.

Jacob, Brian A., Jonah E. Rockoff, Eric S. Taylor, Benjamin Lindy, and Rachel Rosen. 2018. "Teacher applicant hiring and teacher performance: Evidence from DC public schools." *Journal of Public Economics*, 166: 81–97.

Kane, Thomas, and Douglas Staiger. 2005. "Using Imperfect Information to Identify Effective Teachers." Unpublished Manuscript.

Las 2 Orillas. 2014. "Los colegios de la elite: los más caros pero no los mejores." *Las 2 Orillas*.

Ome, Alejandro. 2012. "The Effects of Meritocracy for Teachers in Colombia." Fedesarrollo Informes de Investigación 010260.

Ome, Alejandro. 2013. "El estatuto de profesionalización docente: una primera evaluación." Fedesarrollo Cuadernos de Fedesarrollo 011553.

Petek, Nathan, and Nolan G Pope. 2023. "The Multidimensional Impact of Teachers on Students." *Journal of Political Economy*, forthcoming.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica*, 73(2): 417–458.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review*, 94(2): 247–252.

Rockoff, Jonah E., Brian A. Jacob, Thomas J. Kane, and Douglas O. Staiger. 2011. "Can You Recognize an Effective Teacher When You Recruit One?" *Education Finance and Policy*, 6(1): 43–74.

Rose, Evan K, Jonathan T Schellenberg, and Yotam Shem-Tov. 2022. "The Effects of Teacher Quality on Adult Criminal Justice Contact." National Bureau of Economic Research Working Paper 30274.

Staiger, Douglas O., and Jonah E. Rockoff. 2010. "Searching for Effective Teachers with Imperfect Information." *Journal of Economic Perspectives*, 24(3): 97–118.

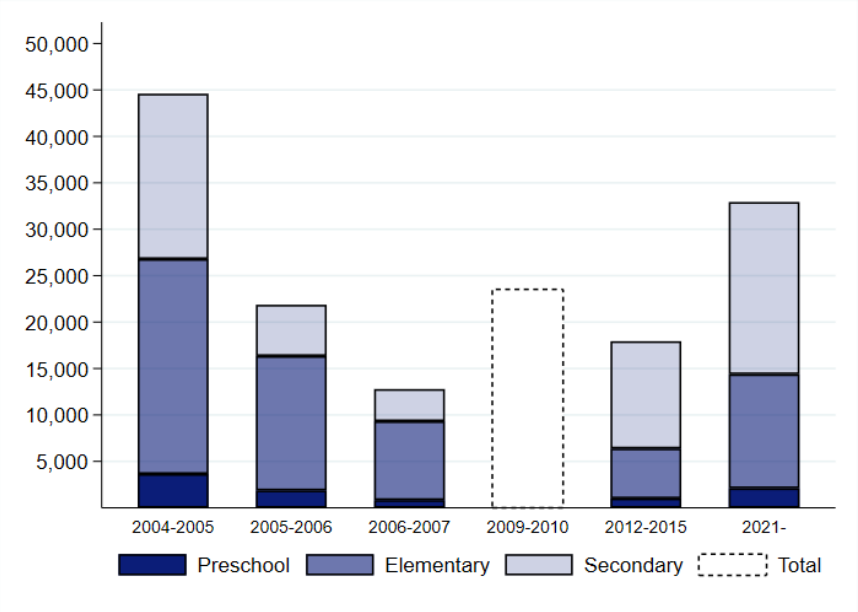
Velasquez, Maria, Oscar Arcos, Adriana Rodriguez, and Fabian Gonzalez. 2010. *Docentes y Directivos Docentes. Acerca de su Condicion y Perfiles sobre Resultados de Convocatorias de Concursos de Meritos.* CID, Universidad Nacional de Colombia.

World Bank. 2018. "World Development Report 2018: Learning to Realize Education's Promise." World Bank, Washington, DC.

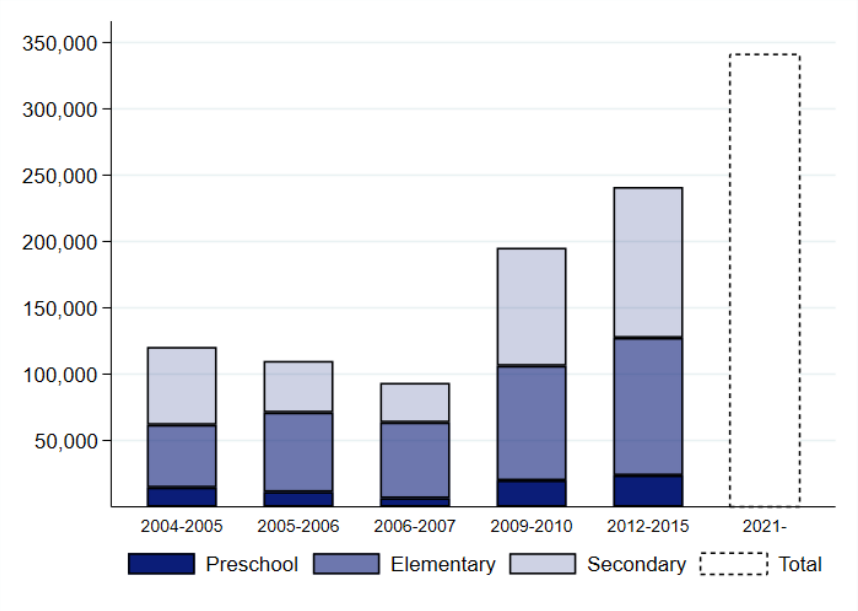
Figures and Tables

Figure 1: Vacancies and Applicants by Merit-based Hiring Process

(a) Teaching Vacancies

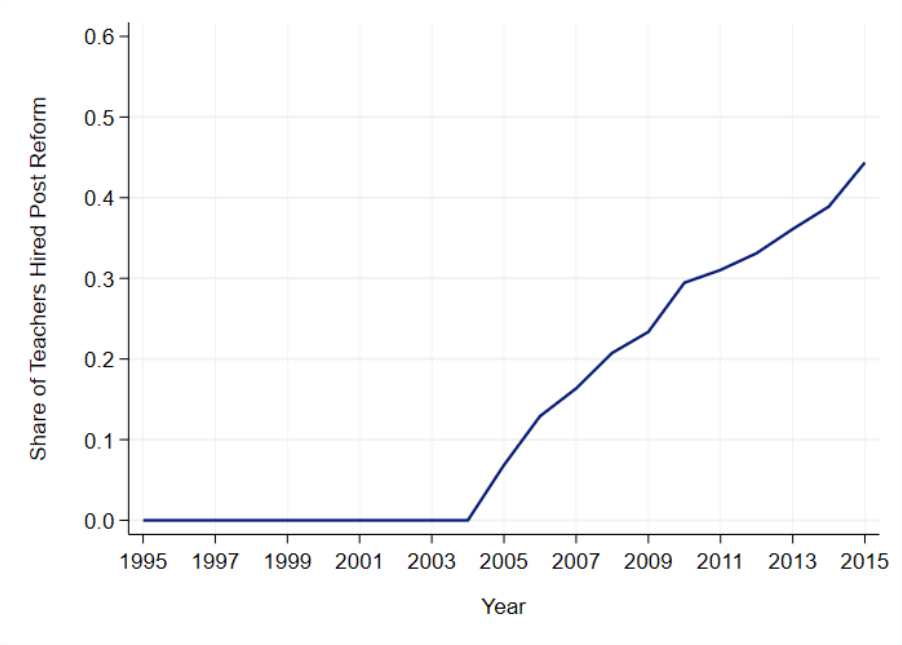


(b) Applicants to Teaching Positions



Notes. Panels 1a and 1b plot, respectively, the number of vacancies and applicants by merit competition across all nationwide hiring processes between 2004 and 2021. Information on vacancies and applicants was gathered from different sources, including the Colombian Ministry of Education, the National Commission for the Civil Service, and Velasquez et al. (2010). Information by teaching level was unavailable for vacancies announced in the 2009-2010 hiring process and for applicants in the most recent process, announced in 2021. Applicants' information corresponds to individuals who took the entry exam assessing teaching aptitude and subject-specific knowledge.

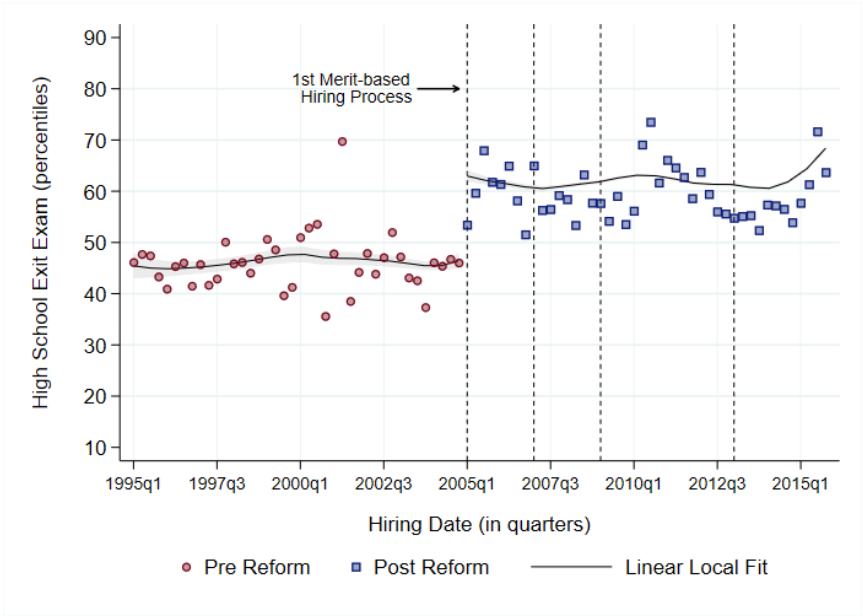
Figure 2: Share of Teachers Hired Post Reform Implementation



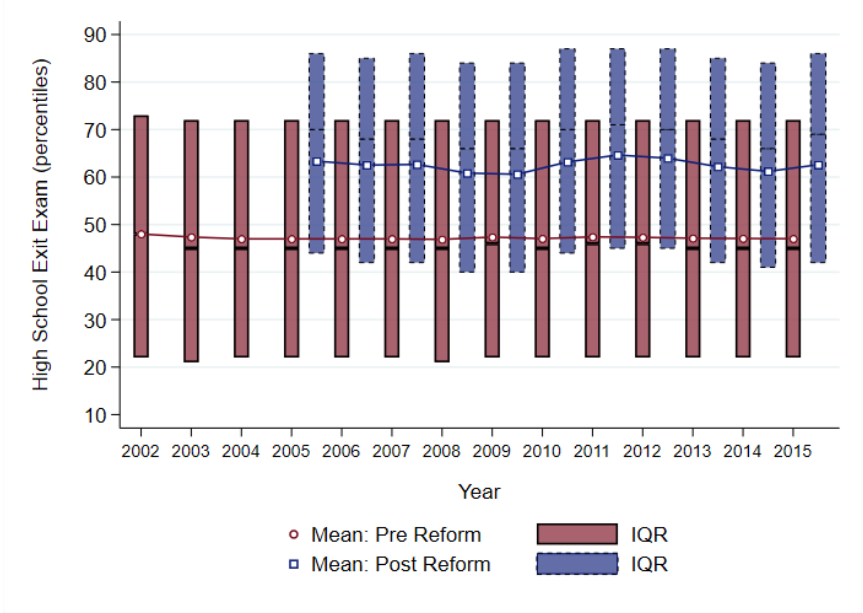
Notes. The solid line represents the annual share of teachers hired after the reform was implemented in 2005.

Figure 3: Pre-college Test Scores of Public School Teachers

(a) Test Scores by Teacher's Hiring Date



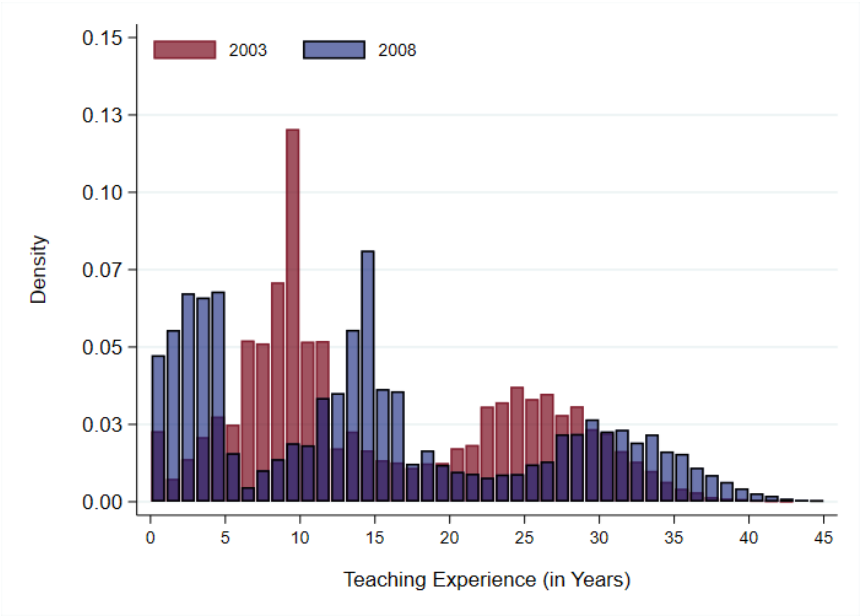
(b) Distribution Pre and Post Merit System



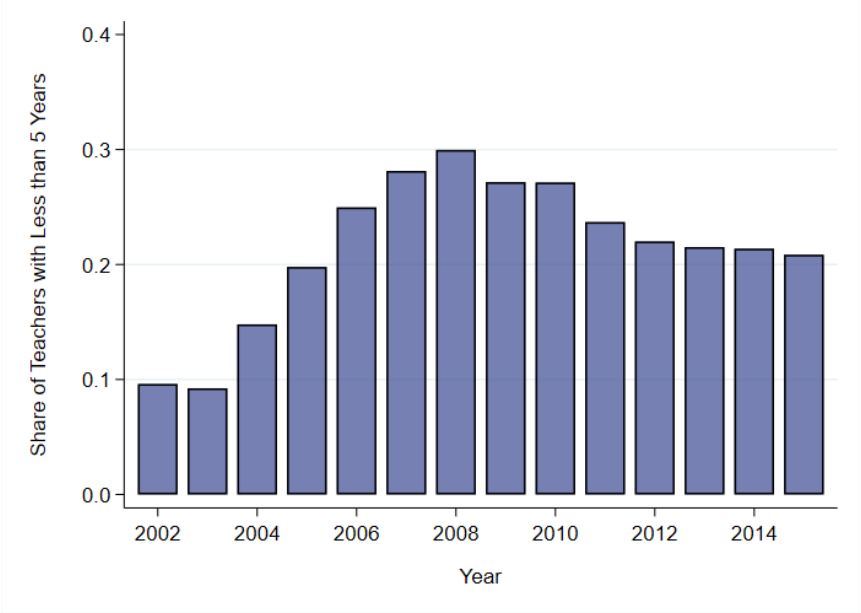
Notes. Panel 3a plots the average percentile in the high school exit exam of teachers hired in each quarter between 1995 and 2015. Solid lines represent local linear regressions fitted using individual-level data of teachers hired before and after 2005. Confidence intervals at the 95% level are displayed around each non-parametric regression. Panel 3b plots the annual interquartile range (IQR), median, and mean performance in the high school exit exam of public school teachers hired before and after 2005.

Figure 4: Teaching Experience Pre and Post Reform

(a) Experience Distribution Pre and Post Reform

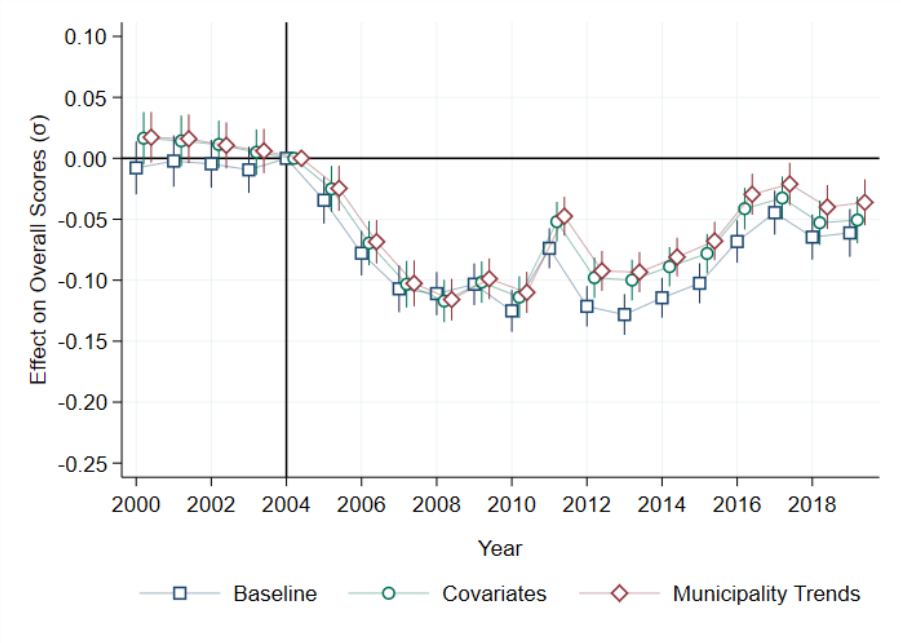


(b) Share of Teachers with Less than Five Years of Experience



Notes. Teaching experience is computed based on each teacher's hiring date. Panel 4a shows the density of experience among teachers working in 2003 and 2008. Panel 4b plots the share of teachers with less than five years of experience working at public schools in any given year between 2002 and 2015. Results displayed between 2002 and 2006 are computed retrospectively using the 2007 teacher census and each teacher's earliest hiring date.

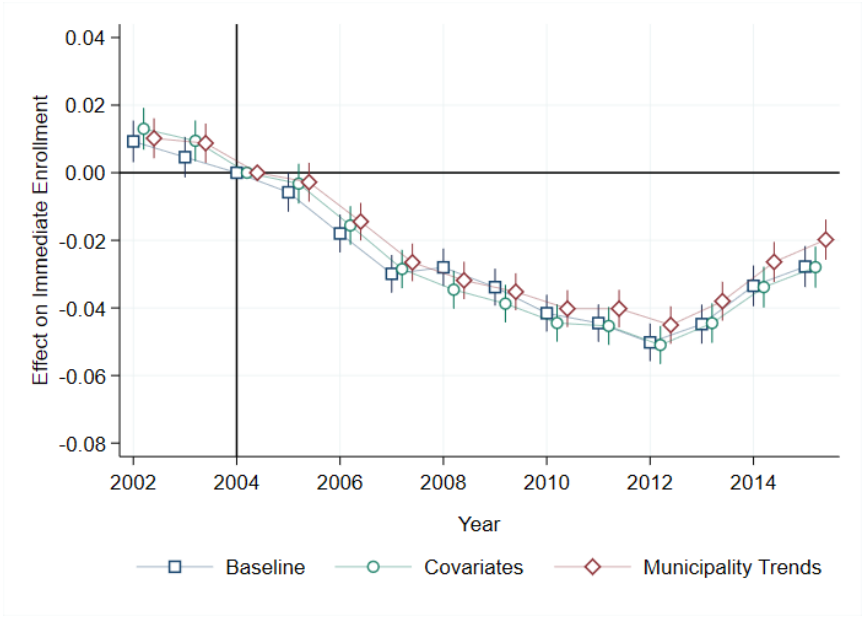
Figure 5: Dynamic Effects of a Merit-based Teacher Hiring Policy on Students' Test Scores



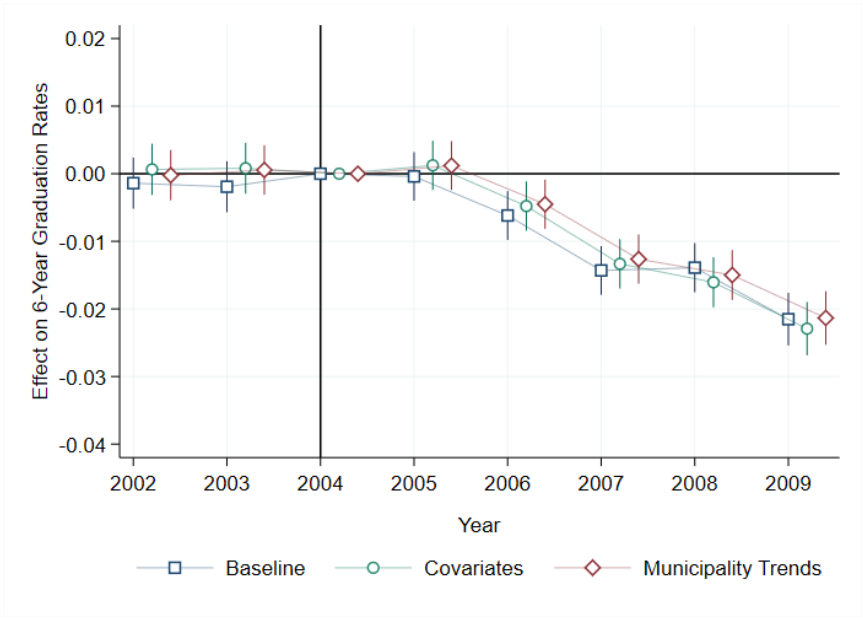
Notes. Ordinary Least Squares estimates of the dynamic effects, δ_τ, of equation 1. The outcome variable is the overall score on the high school exit exam. Overall scores are computed as the average performance in five subject exams: reading comprehension, mathematics, natural sciences, social sciences, and English proficiency. Scores are standardized within each student's cohort. The baseline specification includes school and year fixed effects. The specification with covariates additionally controls for the student's age, gender, socioeconomic stratum, and schooling time (i.e., whole day, morning, afternoon, night, or weekends). The full specification includes municipality linear trends in addition to all other covariates. 95% confidence intervals are displayed around plotted coefficients and are computed using standard errors clustered at the school $\times$ year level.

Figure 6: Dynamic Effects of a Merit-based Teacher Hiring Policy on Students' College Outcomes

(a) Immediate College Enrollment



(b) College Graduation



Notes. Ordinary Least Squares estimates of the dynamic effects, δ_τ, of equation 1. The outcome variable in Panel 6a indicates whether a student enrolls in a college program within six months after graduating high school. In Panel 6b, the outcome variable indicates whether a student graduates from a college program within six years after completing high school. The baseline specification includes school and year fixed effects. The specification with covariates additionally controls for the student's age, gender, socioeconomic stratum, and schooling time (i.e., whole day, morning, afternoon, night, or weekends). The full specification includes municipality linear trends in addition to all other covariates. 95% confidence intervals are displayed around plotted coefficients and are computed using standard errors clustered at the school $\times$ year level.

Table 1: Students' Summary Statistics, 2000-2019 (Pool)

	Public Schools		Private Schools	
	Mean	S.D.	Mean	S.D.
	(1)	(2)	(3)	(4)
Student's Characteristics :				
Age	18.09	3.28	18.36	4.29
Female	0.55	0.50	0.52	0.50
Working	0.10	0.30	0.12	0.32
Family Background :				
Socioeconomic Stratum	1.73	0.77	2.66	1.07
Family Size :				
1 or 2	0.05	0.22	0.07	0.25
3 or 4	0.41	0.49	0.52	0.50
5 or 6	0.39	0.49	0.33	0.47
7 or more	0.16	0.36	0.09	0.28
Mother's Education :				
None or Any Preschool	0.05	0.21	0.04	0.18
Any Elementary	0.40	0.49	0.20	0.40
Any High School	0.42	0.49	0.38	0.49
Any College	0.13	0.34	0.38	0.49
School's Characteristics :				
Urban	0.86	0.35	0.96	0.19
Main City	0.35	0.48	0.64	0.48
Schooling Time :				
Morning	0.55	0.50	0.33	0.47
Afternoon	0.21	0.41	0.07	0.25
Whole day	0.14	0.35	0.44	0.50
Weekends or Night	0.10	0.29	0.16	0.36
Observations	6,627,860		2,322,799	

Notes. Summary statistics pooling students who took the high school exit exam between 2000 and 2019. Socioeconomic stratum is a categorical variable that classifies households based on the physical conditions of the house and the neighborhood where they live in. Households in stratum 1 are the poorest, while households in stratum 6 are the richest. Utility subsidies are allocated based on a household's stratum. Mother's education corresponds to the highest level attended, whether it was completed or not. Main city indicates whether a student lives in one of the thirteen major cities in the country. Information on mother's education, family size, and whether or not a student works, is not available for cohorts between 2004 and 2007.

Table 2: Effect of a Merit-based Teacher Hiring Policy on Students' Test Scores

	Dependent Variable : Test Scores (σ)								
Panel A :									
	Overall			Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Public $\times$ $\mathbb{1}[t \geq 2005]$	-0.085***	-0.085***	-0.082***	-0.198***	-0.183***	-0.148***	-0.029***	-0.034***	-0.029***
	[0.004]	[0.004]	[0.004]	[0.005]	[0.005]	[0.005]	[0.004]	[0.004]	[0.004]
R-squared	0.354	0.386	0.389	0.190	0.221	0.225	0.210	0.231	0.233
Panel B :									
	English			Natural Sciences			Social Sciences		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Public $\times$ $\mathbb{1}[t \geq 2005]$	-0.221***	-0.214***	-0.165***	-0.043***	-0.040***	-0.026***	-0.069***	-0.069***	-0.066***
	[0.005]	[0.005]	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
R-squared	0.325	0.339	0.343	0.252	0.286	0.289	0.209	0.227	0.229
Observations	8,950,659	8,950,659	8,950,659	8,950,659	8,950,659	8,950,659	8,950,659	8,950,659	8,950,659
Covariates		Yes	Yes		Yes	Yes		Yes	Yes
Municipality Trends			Yes			Yes			Yes

Notes. Ordinary Least Squares estimates of the effect of a merit-based teacher hiring policy on test scores, based on the following equation: $Y_{ist} = \mu_t + \mu_s + \delta\,(\text{Public}_s \times \mathbb{1}[t \geq 2005]) + X_i'\gamma + \varepsilon_{ist}$. Public_s indicates whether a student is enrolled in a public high school. Outcome variables are displayed at the top of each column and correspond to overall performance on the high school exit exam and test scores in all evaluated subjects. Overall scores are computed as the student's average in five subject exams: reading, mathematics, natural sciences, social sciences, and English. For students taking the exam between 2000 and 2013, the natural sciences score is computed as the average of physics, chemistry, and biology. Starting in 2014, the exam authority only provides a general score – instead of independent subject scores – in natural sciences. Social sciences scores are computed as the average of history and geography between 2000 to 2005. Starting in 2014, the social science exam includes civic competencies questions in addition to history and geography questions. The mathematics exam includes quantitative reasoning competencies starting in 2014. Test scores are standardized within each student's cohort. All regressions include school and year fixed effects. Specifications with covariates control for the student's age, gender, socioeconomic stratum, and schooling time (i.e., whole day, morning, afternoon, night, or weekends). Standard errors are displayed in square brackets and are clustered at the school $\times$ year level.

Table 3: Effect of a Merit-based Teacher Hiring Policy on Students' College Outcomes

	Dependent Variable :								
	College Enrollment						College Graduation		
	Immediate			2-year					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Public $\times$ $\mathbb{1}[t \geq 2005]$	-0.036***	-0.039***	-0.033***	-0.031***	-0.040***	-0.032***	-0.010***	-0.011***	-0.009***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]
R-squared	0.148	0.157	0.158	0.210	0.232	0.233	0.084	0.088	0.089
Outcome Mean	0.16	0.16	0.16	0.32	0.32	0.32	0.10	0.10	0.10
Observations	6,223,132	6,223,132	6,223,132	5,162,588	5,162,588	5,162,588	3,069,537	3,069,537	3,069,537
Covariates		Yes	Yes		Yes	Yes		Yes	Yes
Municipality Trends			Yes			Yes			Yes

Notes. Ordinary Least Squares estimates of the effect of a merit-based teacher hiring policy on college outcomes, based on the following equation: $Y_{ist} = \mu_t + \mu_s + \delta\,(\text{Public}_s \times \mathbb{1}[t \geq 2005]) + X_i'\gamma + \varepsilon_{ist}$. Public_s indicates whether a student is enrolled in a public high school. Outcome variables are displayed at the top of each column and correspond to college enrollment and graduation indicators with different time windows. Immediate enrollment indicates whether a student enrolls in a college program within six months of graduating high school. 2-year enrollment indicates whether a student enrolls in college within the next two years. College graduation indicates whether a student graduates from college in the following six years after completing high school. All regressions include school and year fixed effects. Specifications with covariates control for the student's age, gender, socioeconomic stratum, and schooling time (i.e., whole day, morning, afternoon, night, or weekends). Results in columns (1) to (3) use information of cohorts 2002 to 2015, columns (4) to (6) use cohorts 2002 to 2013, and columns (7) to (9) use cohorts 2002 to 2009. Standard errors are displayed in square brackets and are clustered at the school $\times$ year level.

Table 4: Heterogeneous Effects of a Merit-based Hiring Policy on Test Scores

	Dependent Variable : Test Scores (σ)								
Panel A :									
	Overall			Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Public $\times$ $\mathbb{1}[t \geq 2005] \times$ Frac. Novice	-0.098***	-0.054***	-0.030**	-0.170***	-0.131***	-0.120***	-0.057***	-0.021*	0.008
	[0.013]	[0.013]	[0.013]	[0.014]	[0.014]	[0.014]	[0.012]	[0.011]	[0.012]
Public $\times$ $\mathbb{1}[t \geq 2005]$	-0.059***	-0.069***	-0.075***	-0.158***	-0.152***	-0.122***	-0.012***	-0.026***	-0.030***
	[0.005]	[0.005]	[0.005]	[0.006]	[0.006]	[0.006]	[0.005]	[0.005]	[0.005]
R-squared	0.358	0.391	0.393	0.192	0.222	0.227	0.213	0.233	0.236
Panel B :									
	English			Natural Sciences			Social Sciences		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Public $\times \mathbb{1}[t \geq 2005] \times$ Frac. Novice	-0.142***	-0.111***	-0.109***	-0.045***	-0.003	0.011	-0.079***	-0.043***	-0.007
	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.013]	[0.012]	[0.012]	[0.012]
Public $\times$ $\mathbb{1}[t \geq 2005]$	-0.188***	-0.188***	-0.142***	-0.030***	-0.037***	-0.027***	-0.047***	-0.056***	-0.063***
	[0.006]	[0.006]	[0.006]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
R-squared	0.325	0.339	0.343	0.252	0.286	0.289	0.209	0.227	0.229
Observations	8,283,963	8,283,963	8,283,963	8,283,963	8,283,963	8,283,963	8,283,963	8,283,963	8,283,963
Covariates		Yes	Yes		Yes	Yes		Yes	Yes
Municipality Trends			Yes			Yes			Yes

Notes. Ordinary Least Squares estimates of the following equation: $Y_{ist} = \mu_t + \mu_s + \beta\,(\text{Public}_s \times \mathbb{1}[\tau \geq 2005] \times \text{Frac. Novice}_s) + \delta\,(\text{Public}_s \times \mathbb{1}[t \geq 2005]) + X_i'\gamma + \varepsilon_{ist}$. Public_s indicates whether a student is enrolled in a public high school. Frac. Novice_s represents the time-invariant fraction of teachers hired within the last five years by 2008. Outcome variables are displayed at the top of each column and correspond to overall performance on the high school exit exam and test scores in all evaluated subjects. Overall scores are computed as the student's average in five subject exams: reading, mathematics, natural sciences, social sciences, and English. Test scores are standardized within each student's cohort. All regressions include school and year fixed effects. Specifications with covariates control for the student's age, gender, socioeconomic stratum, and schooling time (i.e., whole day, morning, afternoon, night, or weekends). Standard errors are displayed in square brackets and are clustered at the school $\times$ year level.

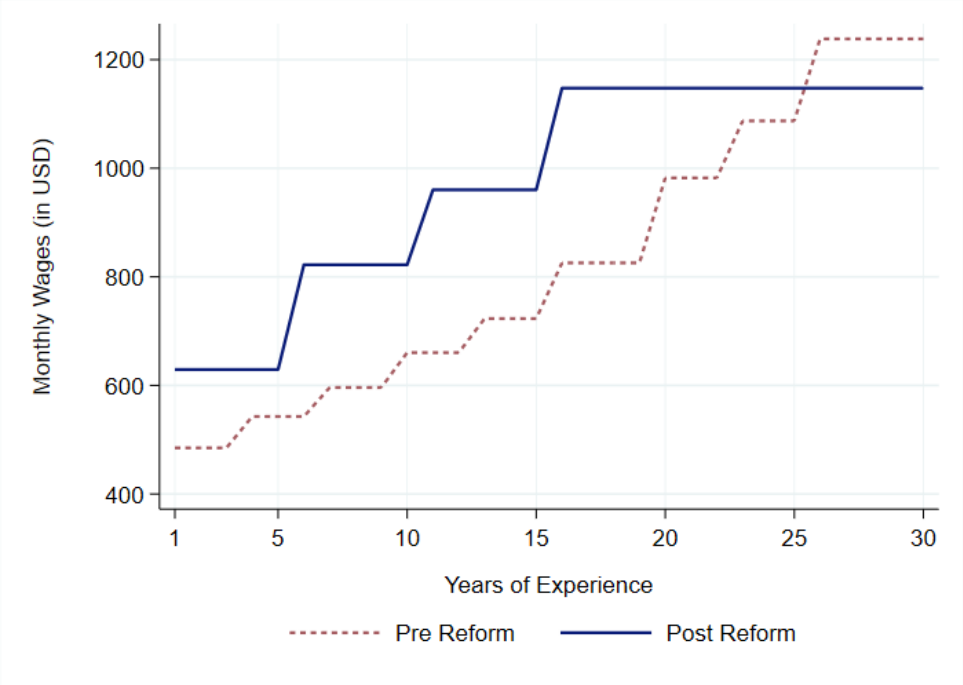
Table 5: Heterogeneous Effects of a Merit-based Hiring Policy on College Outcomes

	Dependent Variable :								
	College Enrollment						College Graduation		
	Immediate			2-year					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Public $\times \mathbb{1}[t \geq 2005] \times$ Frac. Novice	-0.040***	-0.036***	-0.030***	-0.031***	-0.025***	-0.030***	-0.000	-0.000	-0.014***
	[0.004]	[0.004]	[0.004]	[0.006]	[0.006]	[0.006]	[0.003]	[0.003]	[0.003]
Public $\times \mathbb{1}[t \geq 2005]$	-0.027***	-0.031***	-0.026***	-0.026***	-0.037***	-0.027***	-0.011***	-0.012***	-0.007***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]
R-squared	0.150	0.158	0.159	0.210	0.231	0.232	0.086	0.089	0.090
Observations	5,793,448	5,793,448	5,793,448	4,815,563	4,815,563	4,815,563	2,872,578	2,872,578	2,872,578
Covariates		Yes	Yes		Yes	Yes		Yes	Yes
Municipality Trends			Yes			Yes			Yes

Notes. Ordinary Least Squares estimates of the following equation: $Y_{ist} = \mu_t + \mu_s + \beta\,(\text{Public}_s \times \mathbb{1}[\tau \geq 2005] \times \text{Frac. Novice}_s) + \delta\,(\text{Public}_s \times \mathbb{1}[t \geq 2005]) + X_i'\gamma + \varepsilon_{ist}$. Public_s indicates whether a student is enrolled in a public high school. Frac. Novice_s represents the time-invariant fraction of teachers hired within the last five years by 2008. Outcome variables are displayed at the top of each column and correspond to college enrollment and graduation indicators with different time windows. Immediate enrollment indicates whether a student enrolls in a college program within six months of graduating high school. 2-year enrollment indicates whether a student enrolls in college within the next two years. College graduation indicates whether a student graduates from college in the following six years after completing high school. All regressions include school and year fixed effects. Specifications with covariates control for the student's age, gender, socioeconomic stratum, and schooling time (i.e., whole day, morning, afternoon, night, or weekends). Results in columns (1) to (3) use information of cohorts 2002 to 2015, columns (4) to (6) use cohorts 2002 to 2013, and columns (7) to (9) use cohorts 2002 to 2009. Standard errors are displayed in square brackets and are clustered at the school $\times$ year level.

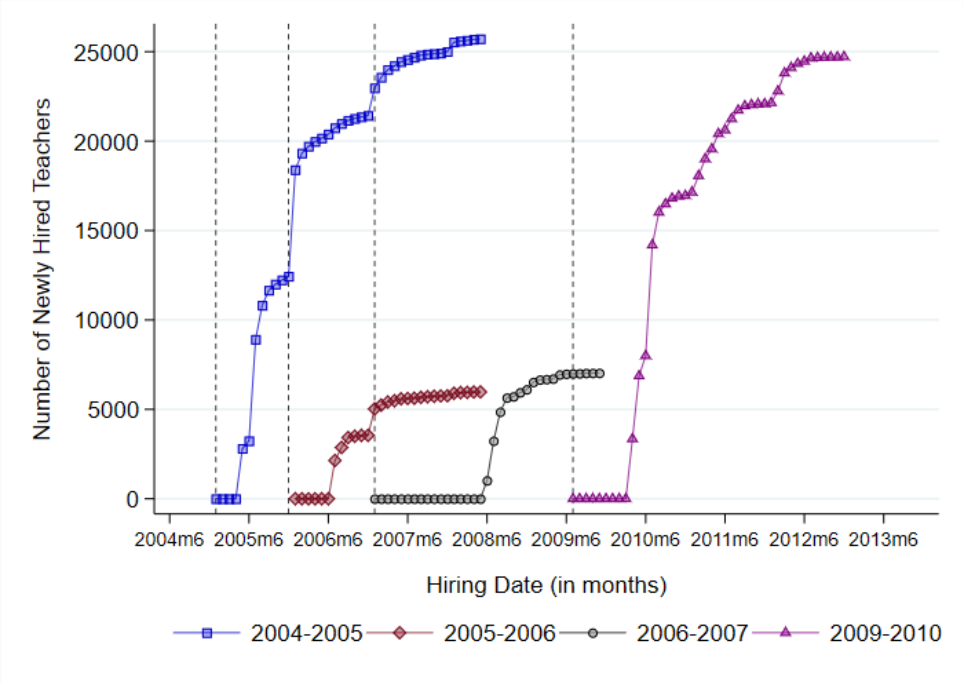
A Additional Figures and Tables

Appendix Figure 1: Wage-Experience Profiles Pre and Post Reform



Notes. Wage-experience profiles are computed based on public school teachers' salaries in 2010, found in Decrees 1367 and 1369. The profile of teachers hired post-reform assumes promotions every five years. The daily average of the exchange rate in 2010, 1 $USD = 1898.7 $COP, is used to present salaries in US dollars.

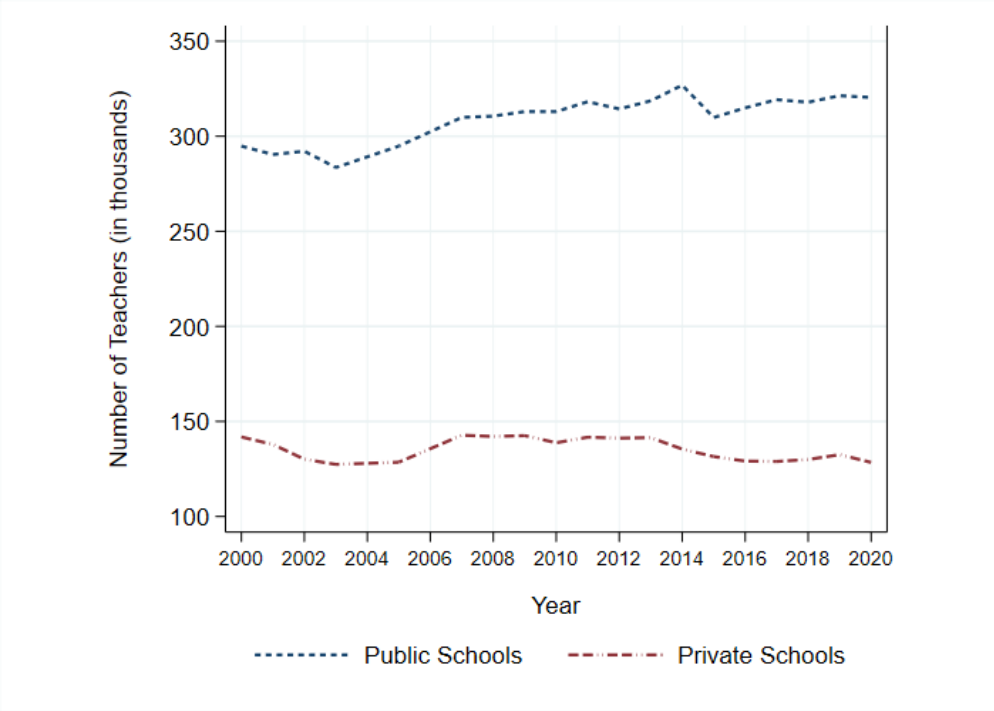
Appendix Figure 2: New Hires Across Time by Merit-based Screening Process



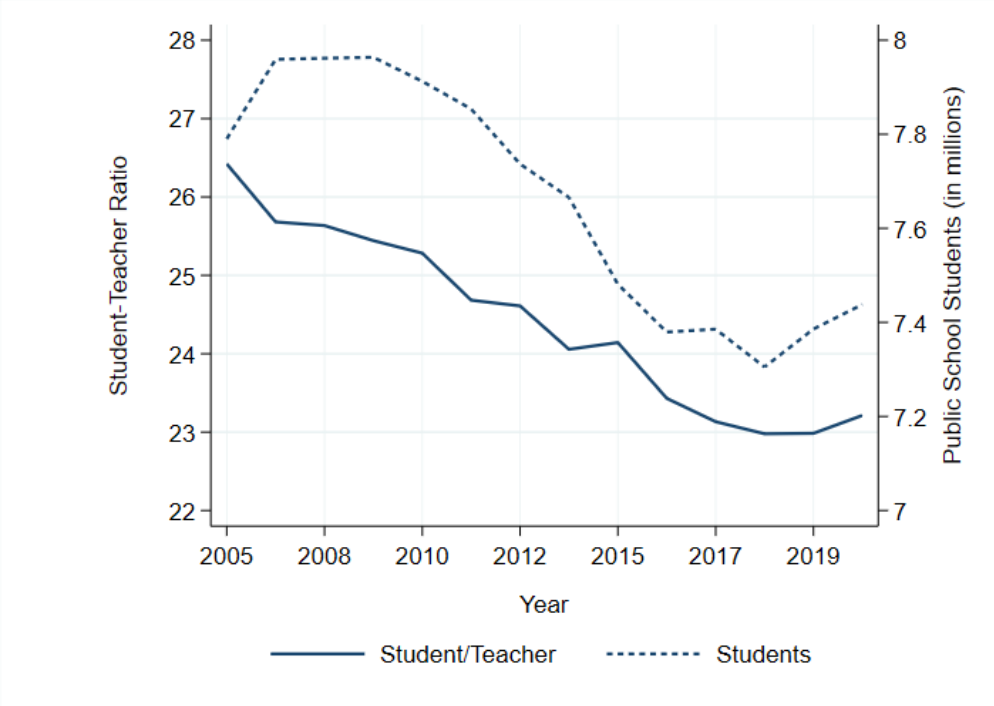
Notes. This figure plots the (cumulative) number of teachers hired across time in each merit-based hiring process between 2004 and 2014. Vertical dashed lines represent the month when individuals hired took the entry exam used by the Colombian government to screen applicants.

Appendix Figure 3: Teachers by Type of School and Student-Teacher Ratio

(a) Number of Teachers by Type of School



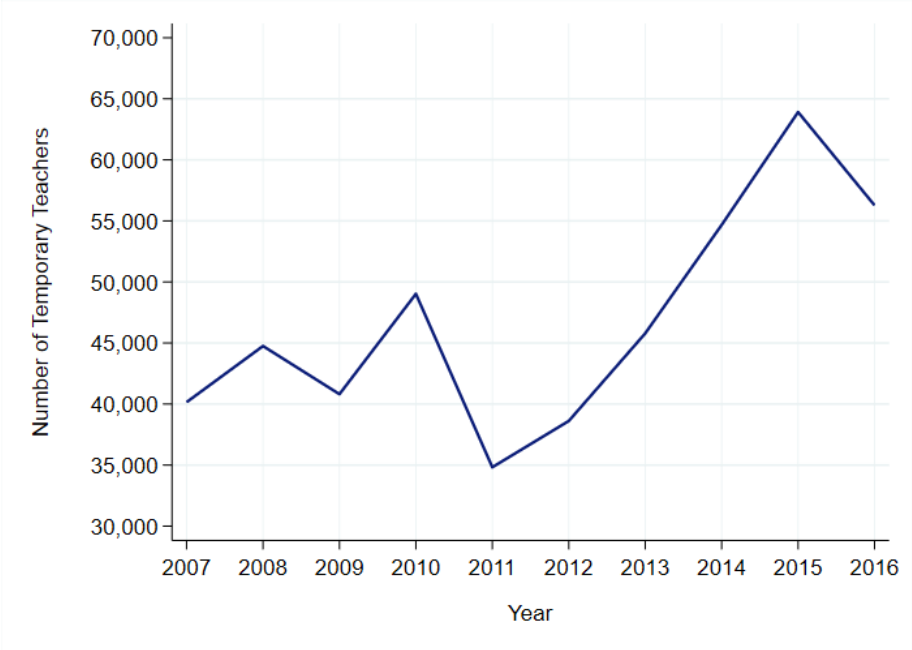
(b) Student-Teacher Ratio at Public Schools



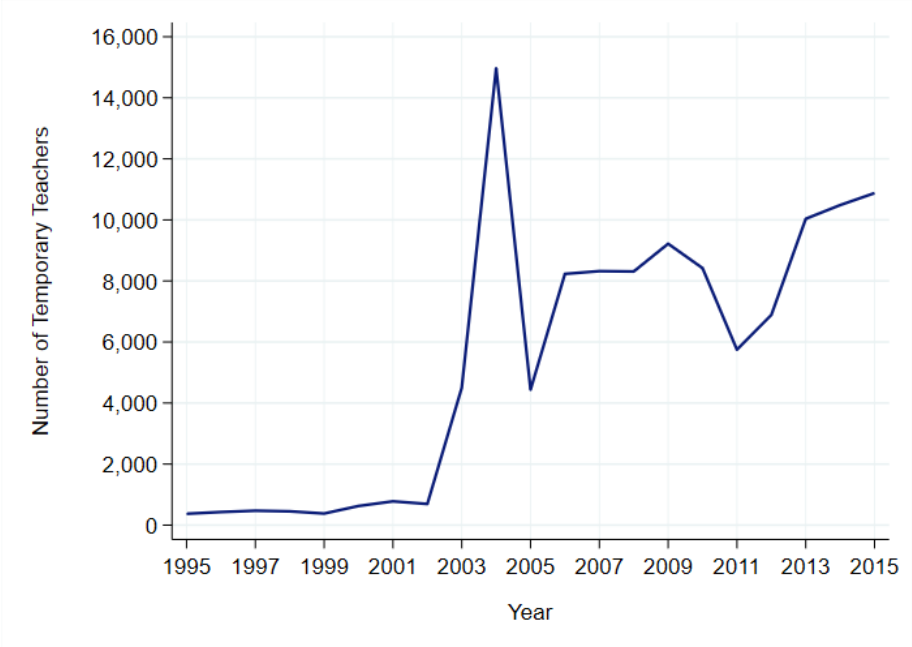
Notes. Panel 3a displays the number of teachers working in public and private schools between 2000 and 2020. Data from 2000 to 2006 is based on information from the Ministry of Education found in Bautista (2009). Information from 2007 to 2020 is based on reports publicly available from the National Administrative Department of Statistics (DANE). Panel 3b displays the student-teacher ratio and the number of students enrolled at public schools between 2005 and 2019. Enrollment information is based on reports publicly available by DANE.

Appendix Figure 4: Stock and Flow of Temporary Teachers

(a) Stock: Annual Number of Temporary Teachers



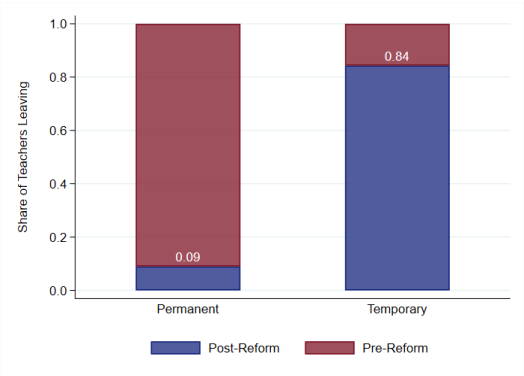
(b) Flow: Number of New System Temporary Teachers Hired Annually



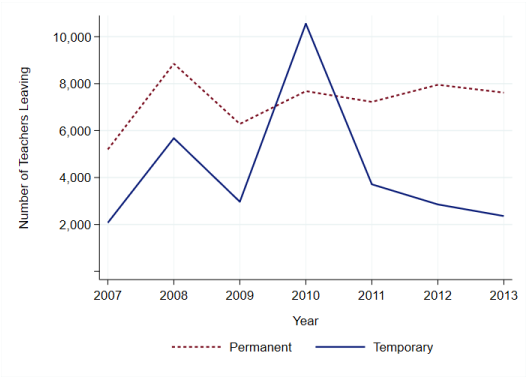
Notes. Panel 4a displays the number of temporary teachers working at public schools in any given year between 2007 and 2016. Panel 4b displays the number of newly hired temporary teachers in each year between 1995 and 2015, based on the pool of teachers working at public schools from 2007 to 2015.

Appendix Figure 5: Teachers Leaving Public School Positions Over Time by Type of Contract

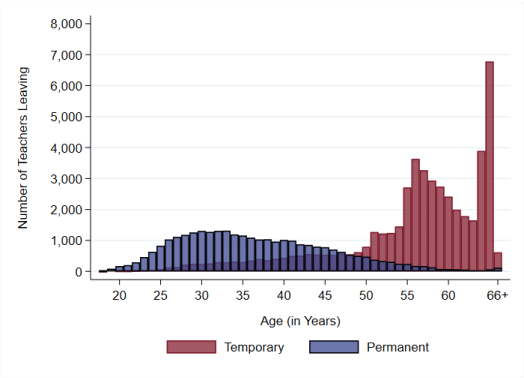
(a) Share of Teachers Hired Pre and Post Reform



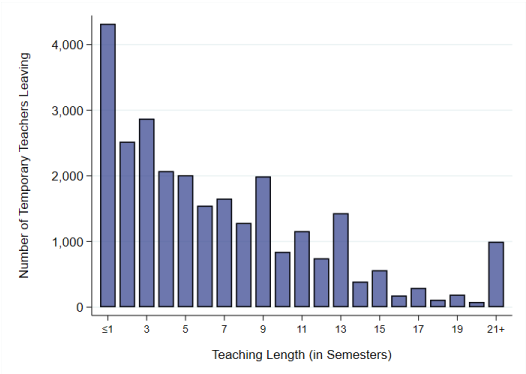
(b) Number of Teachers Leaving Across Time



(c) Distribution of Teachers by Age



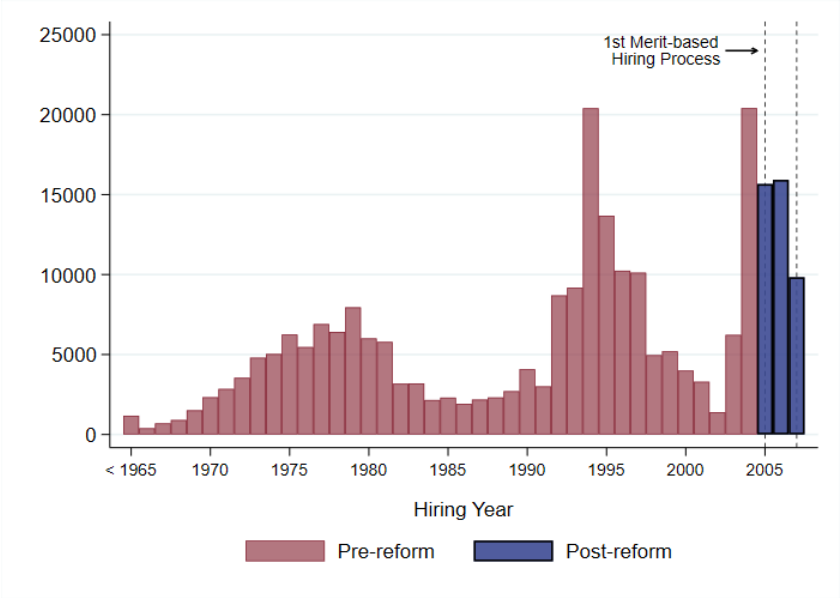
(d) Temporary Teachers' Work Experience at Moment of Leaving



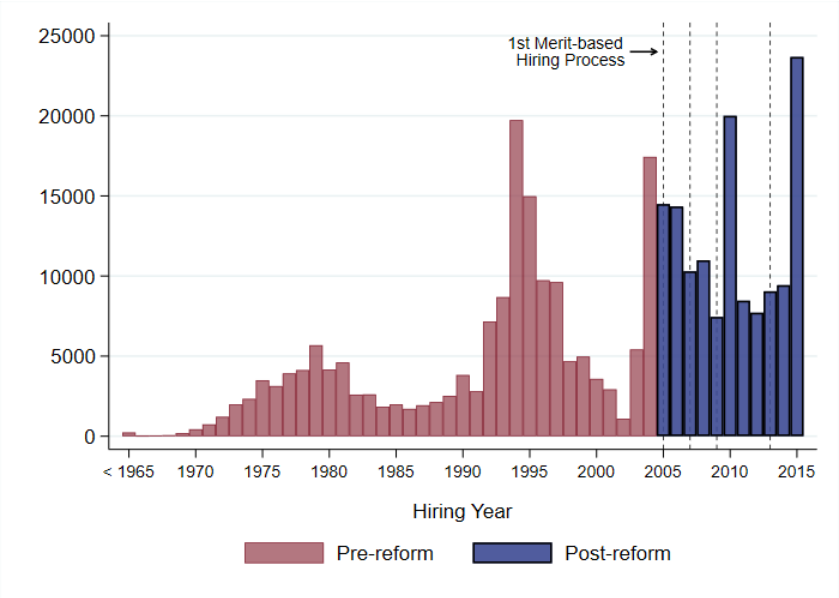
Notes. This figure plots information on individuals who stop working as public school teachers between 2007 and 2013. We assume a teacher leaves or stops working when this is not observed during two consecutive years in the teacher census data. Panel 5a displays the share of teachers hired pre- or post-reform by type of contract (i.e., temporary or permanent). Panel 5b presents the number of teachers leaving across time by type of contract. Panel 5c plots the number of teachers leaving by age and type of contract. Panel 5d plots the distribution of the time that temporary teacher work at public schools before leaving (in semesters).

Appendix Figure 6: Stock of Teachers by Hiring Date

(a) 2007 Cross-section



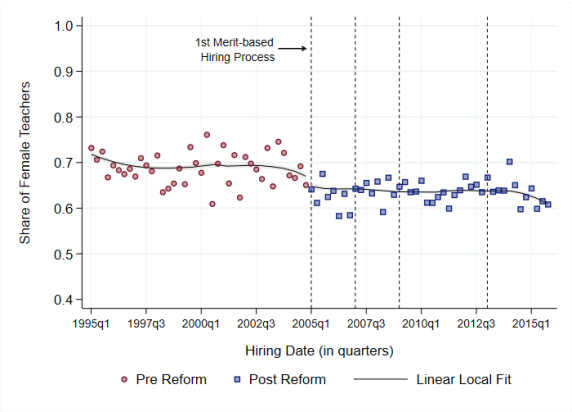
(b) 2015 Cross-section



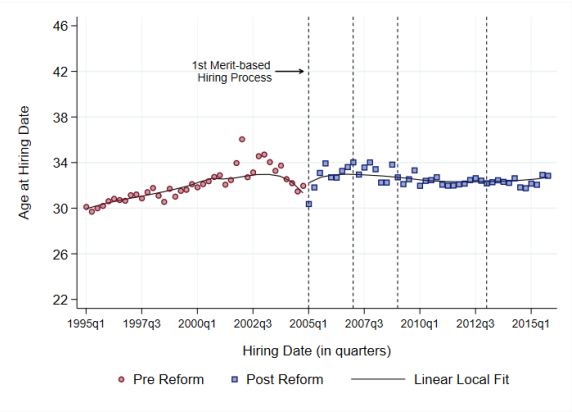
Notes. Panel 6a plots the number of teachers hired annually, based on the stock of teachers working in 2007. Panel 6b, on the other hand, uses the information of teachers working in 2015. Dashed vertical lines represent the years when a new merit-based hiring process starts.

Appendix Figure 7: Effects of the Merit-Based Hiring Policy on Teachers' Characteristics

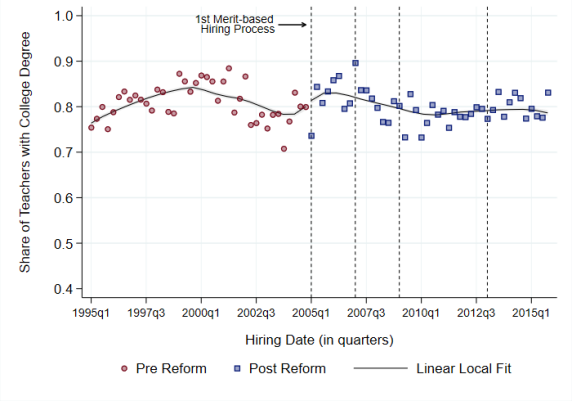
(a) Gender



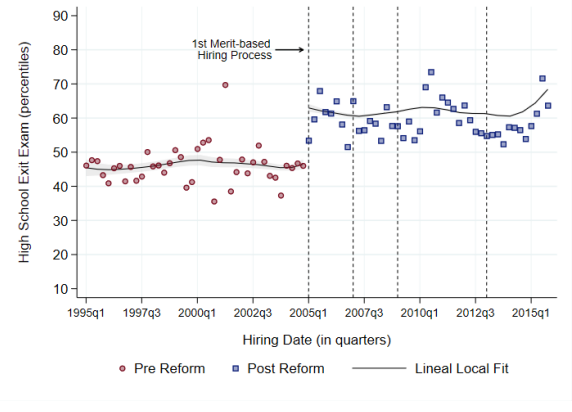
(b) Age at Hiring Date



(c) College Degree

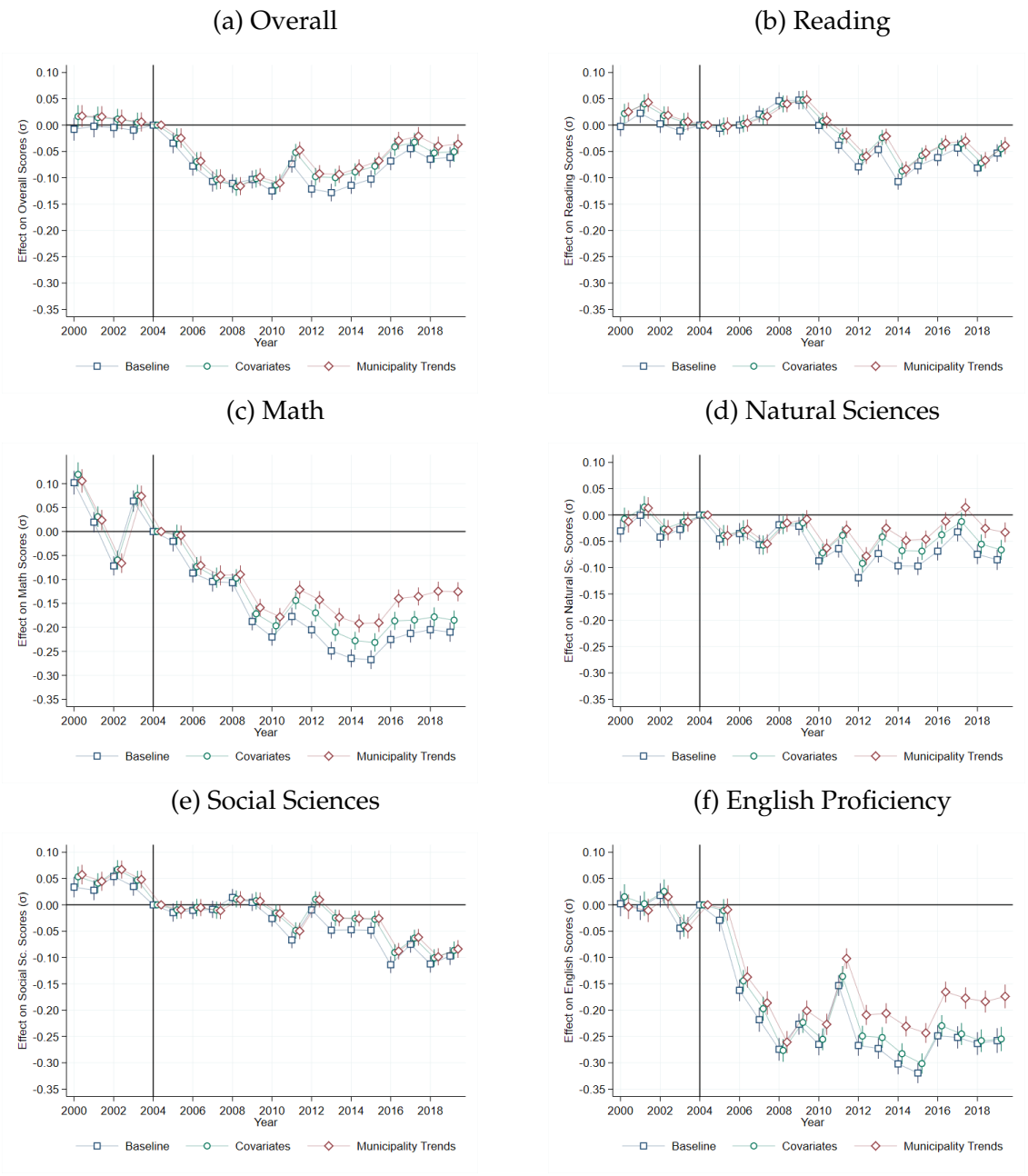


(d) Test Scores



Notes. Plotted dots in Panel 7a represent the share of female teachers hired in the same quarter between 1995 and 2015. Panel 7b plots the average age (at hiring) of teachers hired in the same quarter. Panel 7c plots the share of teachers holding a college degree. Panel 7d plots the average percentile in the high school exit exam of teachers in the same quarter. Solid lines represent local linear regressions fitted using individual-level data. 95% confidence intervals are displayed around each non-parametric regression. Dashed vertical lines represent the quarter when a new hiring process starts.

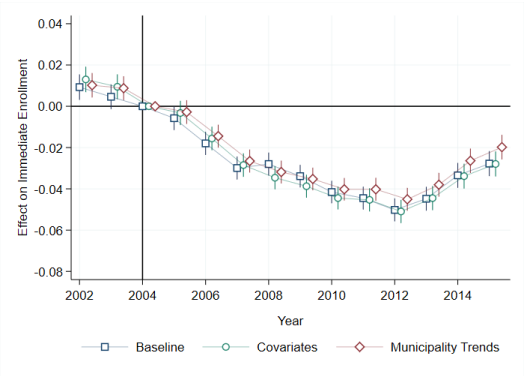
Appendix Figure 8: Dynamic Effects of a Merit-Based Teacher Hiring Policy on Students' Test Scores



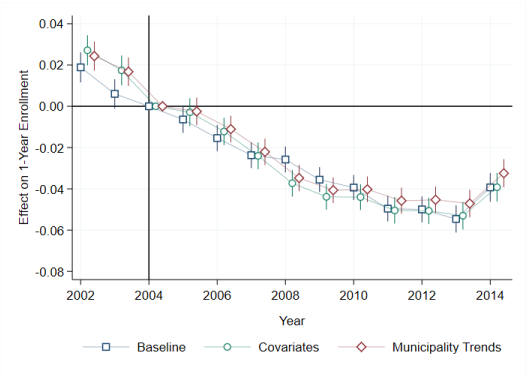
Notes. Ordinary Least Squares estimates of the dynamic effects, δ_τ, of equation 1. Outcome variables correspond to overall performance on the high school exit exam and test scores in all evaluated subjects. Overall scores are computed as the average performance in five subject exams: reading comprehension, mathematics, natural sciences, social sciences, and English proficiency. Scores are standardized within each student's cohort. The baseline specification includes school and year fixed effects. The specification with covariates additionally controls for the student's age, gender, socioeconomic stratum, and schooling time (i.e., whole day, morning, afternoon, night, or weekends). The full specification includes municipality linear trends in addition to all other covariates. 95% confidence intervals are displayed around plotted coefficients and are computed using standard errors clustered at the school $\times$ year level.

Appendix Figure 9: Dynamic Effects of a Merit-Based Teacher Hiring Policy on Students' College Enrollment Rates

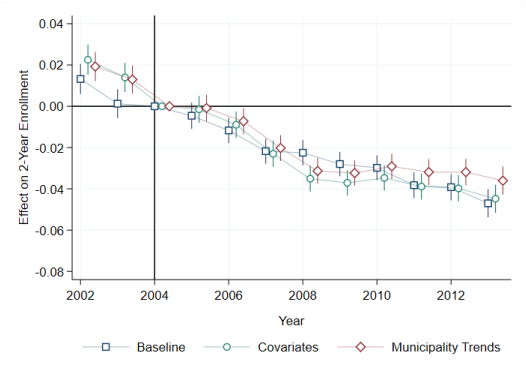
(a) Immediate Enrollment



(b) Enrollment within One Year



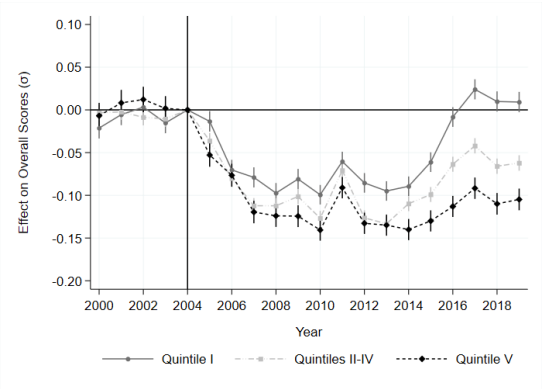
(c) Enrollment within Two Years



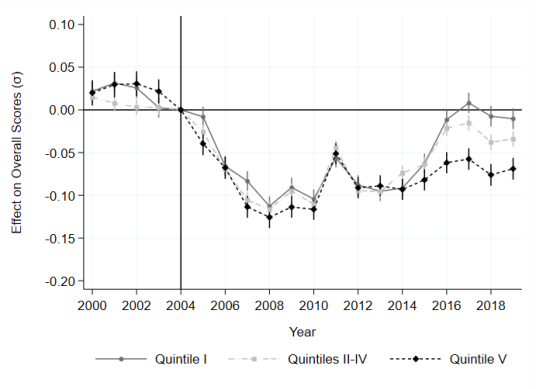
Notes. Ordinary Least Squares estimates of the dynamic effects, δ_τ, of equation 1. Outcome variables are indicators for whether a student enrolls in college within six months (immediate), one year, or two years, after taking the high school exit exam. The baseline specification includes school and year fixed effects. The specification with covariates additionally controls for the student's age, gender, socioeconomic stratum, and schooling time (i.e., whole day, morning, afternoon, night, or weekends). The full specification includes municipality linear trends in addition to all other covariates. 95% confidence intervals are displayed around plotted coefficients and are computed using standard errors clustered at the school $\times$ year level.

Appendix Figure 10: Dynamic Heterogeneous Effects of a Merit-Based Teacher Hiring Policy on Students' Outcomes

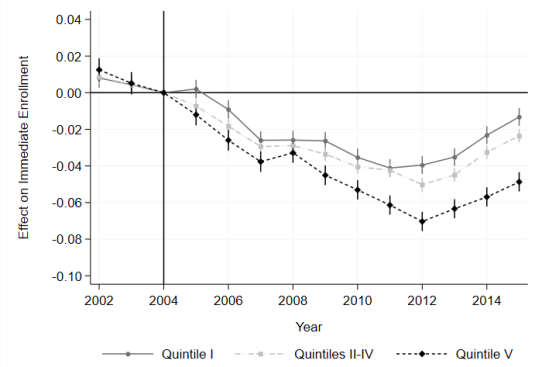
(a) Baseline Model: Overall Scores



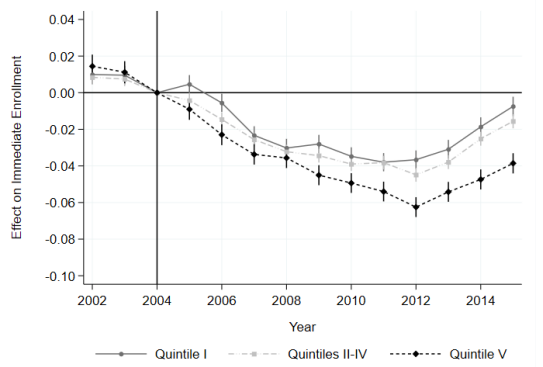
(b) Full Model: Overall Scores



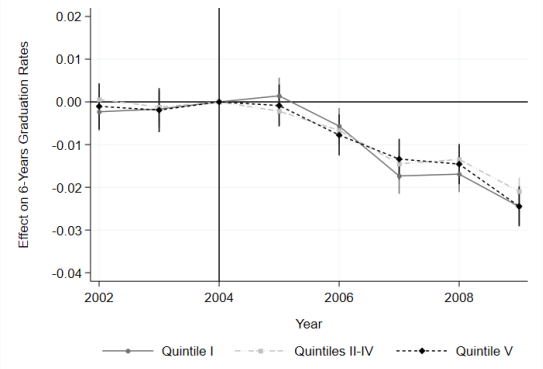
(c) Baseline Model: College Enrollment



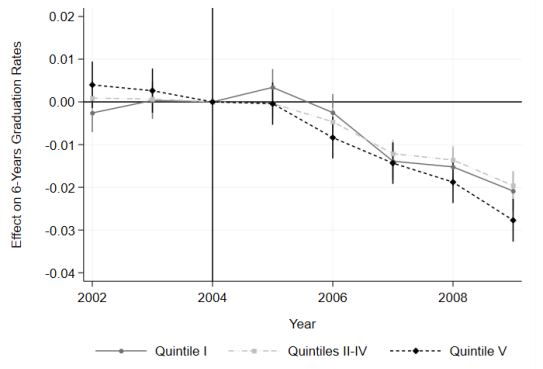
(d) Full Model: College Enrollment



(e) Baseline Model: College Graduation



(f) Full Model: College Graduation



Notes. Ordinary Least Squares estimates of the dynamic effects, δ_τ^I, δ_τ^{II-IV}, and δ_τ^V, of equation: $Y_{ist} = \mu_t + \mu_s + \sum_\tau(\delta_\tau^I \times \text{Novice}_s^I + \delta_\tau^{II-IV} \times \text{Novice}_s^{II-IV} + \delta_\tau^V \times \text{Novice}_s^V) \times \mathbb{1}[\tau = t] \times \text{Public}_s + X_i'\gamma + \varepsilon_{ist}$. Three mutually exclusive groups are defined based on quintiles of the fraction of novice teachers at a student's school by 2008: i) Quintile I, ii) Quintiles II to IV, and iii) Quintile V. Students in Quintile I are enrolled at schools with the lowest fraction of novice teachers. We define novice teachers as teachers hired within the last five years. Novice_s^I is an indicator for whether a student is classified in Quintile I, while Novice_s^V indicates if the student is in Quintile V. Overall scores from the high school exit exam are standardized within each student's cohort. Immediate enrollment indicates whether a student enrolls in college within the next six months after graduating high school. College graduation indicates whether a student graduates from college in the next six years after completing high school. The baseline specification includes school and year fixed effects. The full specification controls for age, gender, socioeconomic stratum, schooling time (i.e., whole day, morning, afternoon, night, or weekends), and municipality linear trends. 95% confidence intervals are displayed around plotted coefficients and were computed using standard errors clustered at the school $\times$ year level.

Appendix Table 1: Statistics of College Students and Graduates by Field of Study

	Field of Study :							
	Education	Agricultural Sciences	Business & Accounting	Social Sciences	Health	Engineering	Economics	Math & Natural Sc.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A : Students								
Female	0.62	0.42	0.60	0.57	0.73	0.32	0.56	0.53
	(0.49)	(0.49)	(0.49)	(0.49)	(0.44)	(0.47)	(0.50)	(0.50)
Age at Enrollment	20.44	19.86	20.65	19.90	19.19	19.45	19.28	18.82
	(3.24)	(2.89)	(3.34)	(2.96)	(2.63)	(2.85)	(2.63)	(2.31)
Low Income	0.69	0.61	0.59	0.54	0.54	0.57	0.46	0.59
	(0.46)	(0.49)	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)
High School Exam	57.61	58.08	58.52	61.03	64.92	68.22	69.02	75.29
	(27.98)	(27.73)	(26.71)	(27.79)	(28.16)	(26.78)	(25.80)	(24.38)
Mother's Education :								
Secondary	0.46	0.39	0.44	0.39	0.39	0.42	0.40	0.40
	(0.50)	(0.49)	(0.50)	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)
College	0.21	0.30	0.27	0.39	0.43	0.35	0.43	0.38
	(0.41)	(0.46)	(0.44)	(0.49)	(0.50)	(0.48)	(0.49)	(0.49)
Graduation Rate	0.49	0.40	0.47	0.55	0.57	0.41	0.52	0.46
	(0.50)	(0.49)	(0.50)	(0.50)	(0.49)	(0.49)	(0.50)	(0.50)
Panel B : Graduates								
Age at Graduation	26.15	25.40	25.42	24.79	24.52	24.74	24.35	24.71
	(3.48)	(3.11)	(3.58)	(3.28)	(2.85)	(2.95)	(2.87)	(2.60)
Earnings Aft. Grad. (t) :								
t = 1	571.26	533.20	642.06	662.77	918.39	754.02	709.09	767.81
	(290.38)	(341.32)	(437.35)	(411.11)	(634.69)	(505.98)	(482.20)	(485.02)
t = 2	627.70	597.22	726.13	749.37	951.17	876.19	822.65	878.12
	(324.48)	(423.32)	(506.54)	(469.12)	(658.81)	(593.30)	(568.30)	(562.31)
t = 3	688.82	665.24	818.11	839.44	1000.66	1005.01	960.59	1004.71
	(355.04)	(473.16)	(583.25)	(533.29)	(671.28)	(682.34)	(660.31)	(659.81)
t = 4	747.13	737.57	922.01	942.51	1070.19	1143.92	1097.58	1133.05
	(383.07)	(513.17)	(664.41)	(610.61)	(698.07)	(768.50)	(753.72)	(748.66)

Notes. Statistics in Panel A correspond to the pool of students who enrolled in college between 2002 and 2015, based on information from *Spadies* data. Low income is computed using an indicator variable equal to one if the student's family is classified in the two lowest socioeconomic strata. Households in Colombia are classified into one of six strata based on the physical conditions of the house and the neighborhood where they live. Families in stratum 1 are the poorest, while families in 6 are the richest. Statistics in Panel B correspond to the pool of students graduating from college between 2007 and 2014, based on the data from the Ministry of Education's *Observatorio Laboral para La Educación* (OLE). Earnings are computed using social security records of all workers in the formal sector. Earnings are deflated and expressed in US dollars of 2010 using the daily average of the exchange rate that year, 1 $USD = 1898.7 $COP.

Appendix Table 2: Descriptive Statistics of Public School Teachers

	Public Teachers	Other Teachers	P-value
	(1)	(2)	(3)
Monthly Wages (in 2010 USD)	896.28	752.33	0.000
Hourly Wages (in 2010 USD)	6.59	4.57	0.000
Weekly Hours	30.34	38.88	0.000
Age	46.33	42.04	0.000
Years of Education	17.00	16.29	0.000
Female	0.64	0.64	0.693
Found job in open call	0.56	0.34	0.000
Tenure (Months)	199.62	123.82	0.000
Is part of a union	0.62	0.25	0.000
Satisfied with current contract	0.98	0.86	0.000

Notes. Statistics in this table are computed using the Colombian household survey (*Gran Encuesta Integrada de Hogares*, GEIH) between 2008 and 2018, publicly available from DANE. Monthly and hourly wages are deflated and expressed in US dollars of 2010. We identified teachers as preschool, elementary, and secondary education workers, based on 4-digit industry codes. Among these workers, we identified public school teachers as those who: (i) contribute to the special pension fund for public school teachers and (ii) work less than 40 hours a week as mandated by the law for all public school teachers. Other teachers correspond to the rest of the workers in the same industry.

Appendix Table 3: Effect of Merit-based Teacher Hiring Policy on Students' College Outcomes

	College Enrollment						College Graduation		
	Immediate			2-year					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Public $\times$ 1($t \geq 2005$)	-0.026***	-0.028***	-0.021***	-0.020***	-0.028***	-0.022***	-0.010***	-0.011***	-0.009***
	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]
R-squared	0.150	0.156	0.157	0.204	0.220	0.222	0.084	0.088	0.089
Observations	3,069,537	3,069,537	3,069,537	3,069,537	3,069,537	3,069,537	3,069,537	3,069,537	3,069,537
Covariates		Yes	Yes		Yes	Yes		Yes	Yes
Municipality Trends			Yes			Yes			Yes

Notes. Ordinary Least Squares estimates of the effect of a merit-based teacher hiring policy on college outcomes, based on the following equation: $Y_{ist} = \mu_t + \mu_s + \delta\,(\text{Public}_s \times \mathbb{1}[t \geq 2005]) + X_i'\gamma + \varepsilon_{ist}$. Public_s indicates whether a student is enrolled in a public high school. Outcome variables are displayed at the top of each column and correspond to college enrollment and graduation indicators with different time windows. Immediate enrollment indicates whether a student enrolls in a college program within six months of graduating high school. 2-year enrollment indicates whether a student enrolls in college within the next two years. College graduation indicates whether a student graduates from college in the following six years after completing high school. All regressions include school and year fixed effects. Specifications with covariates control for the student's age, gender, socioeconomic stratum, and schooling time (i.e., whole day, morning, afternoon, night, or weekends). All results are based on information of students from cohorts 2002 to 2009. Standard errors are displayed in square brackets and clustered at the school $\times$ year level.