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*

March 29, 2023

Abstract

Teacher quality is a key factor in improving student academic achievement. As such, educational policymakers strive to design systems to hire the most effective teachers. This paper examines the effects of a national policy reform in Colombia that established a merit-based teacher-hiring system intended to enhance teacher quality and improve student learning. Implemented in 2005 for all public schools, the policy ties teacher-hiring decisions to candidates' performance on an exam evaluating subject-specific knowledge and teaching aptitude. We find that though the policy sharply increased pre-college test scores of teachers, it also decreased the overall stock of teacher experience and led to sharp decreases in students' exam performance and educational attainment. Using a difference-in-differences strategy to compare the outcomes of students from public and private schools over two decades, we show that the hiring reform decreased students' performance on high school exit exams by 8 percent of a standard deviation, and reduced the likelihood that students enroll in and graduate from college by more than 10 percent. The results underscore that relying exclusively on specific ex ante measures of teacher quality to screen candidates may unintentionally reduce students' learning gains.

Keywords: Teachers, teaching experience, teacher screening, Colombia, 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). Policymakers often implement sweeping education reforms in an effort to improve educational outcomes and close the gaps with high-performing areas. Because teacher quality has been shown to be a main determinant of students' 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).[1]

The success of these teacher-hiring systems depends on whether the information used to screen candidates accurately predicts teacher quality. However, as past work has shown, many of the observable characteristics of those training to become teachers fail to predict their future effectiveness in educating students (Hanushek and Rivkin, 2006; Rockoff et al., 2011). Therefore, schools may struggle to establish selection criteria to discern the potential of teacher candidates, enhance teacher quality, and ultimately improve students' learning outcomes (Kane and Staiger, 2005; Harris, Ingle and Rutledge, 2014). Furthermore, hiring systems that heavily weight specific indicators – such as licensing requirements, educational attainment, or performance on standardized exams – can backfire; they can lead to decisions that ignore or downplay other 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.

In this paper, we examine the impacts on the attributes of teachers and the achievements of students from a nationwide reform that sought to enhance teacher quality

[1]Other countries using such systems include Germany, Belgium, Austria, and Cyprus (Robalino et al., 2007).

and improve student learning in the public schools of Colombia. Implemented in 2005, the reform introduced a centralized, merit-based teacher-hiring system that put a premium on the scores candidates had achieved on a national, standardized exam taken to enter the teacher career. In addition, the reform sought to attract a high-quality pool of teacher candidates by raising salaries. The centralized hiring policy replaced a decentralized one that had been criticized for being subject to political influence. Within 10 years, as a result of the government's large-scale implementation of the reform, nearly half of all public school teachers had been hired under the new regulation.

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 in terms of incoming teachers' characteristics such as pre-college test scores, education levels, 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 use a difference-in-differences strategy to compare students in public and private schools before and after the teacher-selection policy was implemented.

We find that teachers hired under the new system have substantially higher pre-college test scores (i.e., a measure of cognitive skills) than teachers hired under the previous system. While the education level, gender, and age of teachers at the point of hire remained stable, teachers' measures of cognitive skills increased. After the reform was implemented, incoming teachers' test scores rose by 17 percentile points.

However, the reform also led schools to replace many temporary teachers (i.e., teachers without a long-term contract) who had several years of teaching experience in public schools. Within two years of implementation of the reform, public schools had replaced nearly 50,000 temporary teachers – 17 percent of all teachers – who had

been hired prior to the policy change. In addition, after the first two years of the reform, districts continued to annually replace more than 4,000 temporary teachers – representing 37 percent of the teachers who left the profession each year. As a result, the newly hired teachers were significantly less experienced than those employed before the reform was passed. Indeed, four years after the reform was implemented, the share of teachers with fewer than five years of experience tripled, rising from 10 percent to 30 percent.

Due to these changes in teacher composition, the reform had large negative impacts on students' learning and educational attainment. We find that the introduction of the teacher-hiring system decreased the overall performance of public school students on high school exit exams. In the 15 years following the reforms, students' average scores fell by 8.2 percent of a standard deviation – roughly equivalent of the size of the decrease that has been documented to occur from a one-standard-deviation decrease in teacher quality (Chetty, Friedman and Rockoff, 2014; Petek and Pope, 2023). Though the overall negative effect on students' performance is largely driven by large negative impacts on the mathematics and English subject exams, negative effects are also evident for exam scores in all other subjects: reading, natural sciences, and social sciences. Moreover, the reform also had large, negative impacts on students' college enrollment and college graduation rates. The new teacher-hiring system decreased college enrollment by 3.3 percentage points, equivalent to a 21 percent drop in the likelihood of a public school student attending college after high school. These large declines stem in part from the important role that the high school exit exams play in college admissions; when making admissions decisions for different programs, most colleges in the country define cutoffs or assign significant weight to student performance on the high school exit exam. The reform also led college graduation rates (of public high school graduates) to fall. Among those students who had attended public schools after the reform was implemented, there was a 0.9 percentage point (or 9 percent) decline in the likelihood of graduating from college.

The negative effects of the reform seem to be driven by an increase of students'

exposure to teachers with less teaching experience in public schools. We interact the treatment indicator with the baseline share of teachers with fewer than five years of experience (i.e., *novice* teachers hereafter), and find that the negative effects are 40 percent larger (in absolute value) for students in public schools with a higher fraction of novice teachers. We also find a strong negative correlation between the fraction of novice teachers and the dynamic effects of the reform. Jointly, these results suggest that a larger exposure to inexperienced teachers post-reform can help explain students' learning losses.

Even though the reform's explicit intent was to improve teacher quality and, ultimately, student learning, the change negatively impacted students' educational outcomes. Our findings underscore the need for caution in devising and implementing policies that too heavily rely on screening and hiring prospective teachers by using ex ante measures of quality. While the Colombian reform increased the test scores of new teachers, such scores are not necessarily correlated with effectiveness in improving student learning – as our findings underscore. Indeed, the exclusive reliance on the candidates' test score result outweighed all other criteria, including the value that experience might play in an applicant's success in the role of educator. Thus, the reform also inadvertently reduced the stock of teaching experience, which can affect students' academic achievement (Staiger and Rockoff, 2010; Araujo et al., 2016).

Our findings offer a cautionary tale about the negative ramifications that can result from well-meaning efforts to improve teacher quality. In light of the results of our study, we argue that teacher-hiring policies should carefully evaluate what ex ante information they use for decisions on which teacher applicants to hire. Given how difficult it is to predict teacher quality, reforms that aim to improve the caliber of educators may want to avoid selecting teachers based exclusively on ex ante information. Indeed, our findings suggest that a more reliable way to enhance the quality of teachers may be through retaining and laying off teachers based on ex post measures of teacher quality. In addition, our results indicate that the efficiency of any reforms to educational hiring practices will be affected by other forces they set in motion; in this

case, for example, the policy change led to increased teacher turnover. The districts replaced the temporary teachers upon whom the schools had heavily relied with novice teachers who met the new requirements, leading to an unusual increase in teacher turnover.

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 candidates' background information (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 on subject-specific knowledge and teaching aptitude exams. The evidence provided by these papers is mixed, documenting either positive effects 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 result in worse student outcomes when they lead school administrators to downplay the role of other important factors – such as experience – in the teaching production function.

This paper is similar to the work of Ome (2012, 2013) and Brutti and Sánchez-Torres (2022), who study the same Colombian reform that introduced a centralized merit-based system to hire new teachers. The estimation strategy of these papers exploits the variation from the policy change by examining impacts of *within-school* and *within-subject* employment of teachers hired under both the new and old systems. Using this approach Ome (2012, 2013) finds negligible effects on student test scores, and Brutti and Sánchez-Torres (2022) finds evidence of a small positive impact. However, these results are potentially biased due to strong identification assumptions: the share of

teachers hired under the reform is orthogonal to unobserved factors within schools or school subject. By contrast, we provide evidence regarding the aggregate effects on the education market that stem from the reform. We also rely on weaker identification assumptions, and we provide evidence on validity. In addition, their main results condition on teacher experience, an attribute that was affected by the reform.

Our paper also relates to empirical work regarding the effects of teacher quality as measured by teacher value-added measures. Research in this area has consistently found evidence that teacher quality explains a significant fraction of the variation in students' academic performance, education attainment, and adulthood outcomes, such as 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 scant predictive power that teachers' observable characteristics have on the future effectiveness of teachers to foster student learning down the line; such observable characteristics include education level of the teacher candidates and the certifications held (Rivkin, Hanushek and Kain, 2005; Hanushek and Rivkin, 2006; Jackson, Rockoff and Staiger, 2014; Araujo et al., 2016). The evidence indicates that, while teachers' education does not explain students' learning in most school systems, teachers are typically 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).

Recent research has sought to analyze the relationship between teacher quality and characteristics that are difficult for education authorities to observe. Using an experimental setting in Ecuador and extensive information on teachers, Araujo et al. (2016) find that children assigned to teachers with fewer than three years of experience learn less than the average student. They also document how other characteristics of teachers – including skills, measured by IQ scores and the Big Five personality traits, and tenure – are not correlated with student performance. Additional evidence from

Ecuador also suggests that students assigned to teachers who performed better on 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: Section 2 provides the context for our work by describing the Colombian education system, the teacher labor market, and the reform of the teacher hiring policy. Section 3 presents the data sources and descriptive statistics. Section 4 provides evidence of the reform's effect on different attributes of the teaching staff. Section 5 presents main results; we introduce the research design and the identification assumptions we use to interpret our results as causal effects of the merit-based hiring system; and we then present evidence of the reform's effect on students' academic outcomes. Section 6 discusses how our findings align with previous work on teacher quality and provides 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, which was approved in 2002 and implemented in 2005. We focus on the government's effort to attract higher-skilled teacher applicants, the details and changes introduced by the new centralized hiring system, and the role of temporary teachers in the context of the teacher hiring reform.

2.1 The Colombian Education System

School enrollment rates have grown dramatically in Colombia over the past several decades. By 2010, Colombia's elementary education had reached near-universal enrollment, and secondary education enrollment had risen from 35 percent to 77 percent over the previous two 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. On standardized international exams, Colombian students' 75th percentile score lies well below the 25th percentile score for students in member countries of the Organisation of Economic Co-operation and Development (OECD). This low performance has been persistent over time and has pushed Colombia to rank low among countries that participate in the Programme for International Student Assessment (PISA) exams.[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 consists of vocational programs of two and three years, or bachelor's degree 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 two reasons. First, the exam is mandatory to enroll in any higher education institution; in any given year, 95 percent of all seniors take the exam. Second,

[2]Colombia ranked 47 out of 58 countries that took the 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, with Colombia among the lowest-ranked countries.

[3]In 2014, the annual tuition among private secondary schools 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.

exam results serve as the basis of an annual ranking of all high schools; this ranking, which shows how each school's students performed, is published by the national government. Because schools' reputations are affected by these rankings, they 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, and they can pay to enroll at private institutions that provide instruction to take the exam.

During their senior year in high school, students consider whether they will apply to college. Because students are admitted to a specific college program, the application process requires them to select a major area of study. 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 The 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 *Escuelas Normales Superiores*.[6] A large share of teachers at public and private schools are college graduates from education majors.[7] Students who are education majors are more likely to be women, more likely to be from low-income families, and more likely to have received low scores on the 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 10 percent higher than the average earnings of graduates with an education degree who find formal employment elsewhere, and it is only 2

[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] *Escuelas Normales Superiores* ("normal schools," in English) 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 salary for new graduates with degrees in business and accounting.[8] In addition, these teaching positions offer a great deal of employment stability, annual bonuses, and coverage in a social and health insurance system specifically 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 to each school district. The process began with an annual assessment by city mayors to determine the number of vacancies in schools within their jurisdiction. There were no standardized criteria on how local and regional authorities were to screen candidates.[10] Some minimal education and experience standards were required to be met to be appointed as a teacher. Entry exams were used, but they varied by region and had no clear evaluation standards.[11] The lack of clarity led to speculation that the allocation of vacancies was susceptible to manipulation by political interests (Bustamante, 1996; Duarte, 2001, 2003).

Salaries were also regulated. Progression was determined by a 14-level career ladder, with each step linked to a specific number of years of experience and a certain level of education. For instance, teachers who lacked professional experience but held an education college degree were assigned to a career level seven at the time of hiring. Promotions were tied to additional years of experience, completion of pedagogy courses, and graduate education attainment.[12] Public school teachers' jobs were highly stable, and only severe misconduct was likely to prevent a teacher from work-

[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 of employment 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] The exams were cancelled often due to implementation issues (Tiempo, 1996).

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

ing continuously until he or she reached retirement age. Teacher and student performance played no role in promotions or tenure.

In 2002, the hiring and employment conditions of public school teachers were reformed by Decree 1278, which introduced a merit-based, centralized system. The aim was to improve the quality of public education. The reform linked hiring, tenure, and promotion to a teacher-evaluation process. Under this regulation, the government determines the number of teaching vacancies available nationwide and then announces a *public call* to fill such positions.

Applicants must take a written exam measuring (i) knowledge of the subject that each candidate wants to teach, and (ii) teaching aptitude (including pedagogy and subject-specific knowledge). Applicants must earn a score of at least 60 (out of 100) in both components of the exam is necessary to continue in the process.[13] Only the applicants that perform well in the written exam may move on to the individual interviews in the next stage of the process.[14] Finally, candidates obtain a weighted score based on the written exams and the interview results. The written exam, however, heavily outweighs the other components of the score. It corresponds to 70 percent of the final score, whereas 20 percent is given to the curriculum (including experience) and a remaining 10 percent to the interview.

To fill vacancies, the government first computes a general ranking of approved candidates according to their individual scores. Then, in a public hearing, 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 teacher applicants who are lower in the rankings.[15]

Public teachers hired in the post-reform period are assessed by means of a trial pe-

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

[14] 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 has the experience required for the teaching position. In this stage of the hiring process, candidates are given scores based on their interview performance, experience, and education.

[15] 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 hearing.

riod, annual evaluations, and written exams, which are used to determine promotions. On paper, new 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.[16] However, in practice, teachers rarely fail their trial period evaluation (Garcia et al., 2014; Forero and Saavedra, 2019). After the trial period, teachers undergo annual evaluations, which are conducted by their principals. Their continued employment is conditional on not failing two consecutive evaluations. 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.[17] Finally, the reform tied promotions to performance on a written exam evaluating teachers' knowledge.[18]

The reform also increased the salaries of public school teachers. After the reform, college graduates with no prior teaching experience were hired at entry-level wages that were 12 percent higher than they had been. The increase accounts for an earnings premium of 34 percent after 15 years of experience.[19] Such an increase in wages was intended to attract a higher-quality pool of teacher candidates to fill the vacancies at public schools.[20]

Six public calls to fill vacancies nationwide have taken place since the reform was enacted in 2002 (see Figure 1a).[21] The first call was carried out in 2004 when the gov-

[16]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 the environment. To pass the trial period, teachers must obtain a minimum score of 60 out of 100.

[17]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 be approved.

[18]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 receiving their last promotion), and they obtain a score above 80 out of 100 to be promoted.

[19]Appendix Figure 1 plots the wage-experience profiles for college graduates hired before and after the reform. For this figure, we assume that teachers hired post-reform were promoted every five years.

[20]Teachers hired prior to the reform could participate in the new hiring process in any year, facing the same conditions as any other applicant. However, only a few decided to do so, given that a large share of them had many years of experience, were at the top of the wage ladder, and switching to the new hiring system would not increase their salaries (Ome, 2012, 2013).

[21]The government has also made special smaller calls to fill vacancies in distant areas. Two of these, conducted in 2006 and 2012, were used to fill positions at a small number of public schools that provide education to ethnic communities. The most recent call was made in 2018 to hire candidates willing to teach in rural schools or in areas that have suffered the consequences of the armed conflict in the

ernment announced that 44,596 teachers were needed. Appendix Figure 2 shows that it took more than a year for these teachers to be hired. (This implies that instruction under the new reform began in 2005.) 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; however, the government only began appointing candidates in late 2015 to fill the 17,941 vacancies that had initially been announced.[22] More recently, in 2021, the government announced a new public call to fill approximately 29,000 teaching vacancies.

The public calls have been oversubscribed. Each call has attracted a number of applicants that is more than double the number of vacancies, suggesting that teaching positions at public schools are attractive and competitive (see Figure 1b). The large-scale implementation of the reform has reshaped the country's educational workforce; by 2015, nearly half of all public school teachers in Colombia had been hired under the provisions of the new regulation (see Figure 2).

2.3 Temporary Teachers

Under the new hiring system, the lists of eligible candidates are only updated every few years; as a result, the new system has been unable to quickly fill teacher vacancies that arise annually due to retirements or departures. Some positions remain unfilled. To address the short-term demand for teachers in schools and ensure that continued education took place, districts employ temporary teachers who are hired to work until the next wave of teachers selected from the public call can replace them. These temporary teachers are also required to hold a college degree or a pedagogy diploma but, at the time of being hired, they need not have passed the teaching screening subject or pedagogy exam.

Between 2007 and 2015, temporary teachers accounted for a significant proportion

country.

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

of the teaching staff, comprising 12 to 20 percent of all positions. The contracts of temporary teachers do not specify a fixed term of employment; in practice, 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 are 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. The first source provides census data about public school teachers over the period from 2007 to 2015. These data come from the human resources system of the Colombian Ministry of Education.[23] The data include unique identifiers of about 400,000 teachers as well as the exact date when each teacher was hired. This allows us to retrospectively create a longitudinal data set with information on teachers' experience and career development over time.

The second data source we use provides test-score data of all students enrolled at public and private high schools. The data correspond to over 9 million students who took the high school exit exam in the second semester of each year, between 2000 and 2019.[24] 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 scores on subject-specific and overall tests within cohorts.[25] These data include test scores and students'

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

[24]Students in most schools in the country take the exam during the second semester of the year. Only students in the most elite private schools – and a negligible portion of the students in public schools– take the exam in the first semester.

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

demographic characteristics such as gender, age, and the household socioeconomic stratum that serves as a proxy for family income.[26] The information is administered by the exam authority (ICFES), and is available online.[27]

Our third data source provides census-like administrative records of students enrolled in college between 1998 and 2016. The Ministry of Education collects these data through a system known as *Spadies*, which is used to monitor education dropout and graduation rates across time. The data correspond to more than 5 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 exam as a proxy of skills for teachers hired before and after 2005 – when the reform effectively started to change the composition of teachers in Colombian schools. 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.[28]

Table 1 presents summary statistics of our sample of students, separated by those who attended public and private high schools. Students in the sample, both from public and private schools, are 18 years old on average, and slightly more than half

[26] Residential properties in Colombia are assigned a socioeconomic index level (or stratum) from one to six, depending on the neighborhood where the property is located. A higher index indicates that the neighborhood has more access to amenities and public services. The index stratum serves as a proxy for family income.

[27] We obtained access from the Ministry of Education to restricted data that include 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.

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

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 live 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 that in the wake of the change in the hiring system, there was a sharp increase in the high school exam scores of those who were hired to be teachers; these scores serve as 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 hiring reform. Teachers hired after 2004 had test scores that were 17 percentile points higher than those of 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 separating those that were hired before and after the reform.[29] 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 level of teaching experience at public schools changed.[30] Under the new system, teachers with fewer than five years of experience became a significant share of the teaching staff. Similarly, Figure 4b shows that the share of novice teachers (i.e., those with fewer than five years of teaching experience) quickly increased from

[29] We only observe teachers working between 2007 and 2015, but we impute the mean and inter-quartile range retrospectively using the dates when teachers were hired.

[30] Our measure of experience corresponds to teaching experience in the public sector.

just under 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 where it remained fairly stable over the later part of our sample period.

4.1 The Merit-based Reform and Teacher Turnover

The merit-based reform led to a decrease in the average public teacher's level of experience in the public sector by unexpectedly increasing teacher turnover. A constant amount of turnover is expected every year as teachers retire or switch occupations. Any hiring scheme for teachers deals with this set of issues, with novice teachers arriving to replace those who leave the profession. The Colombian 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, only a small number of permanent teachers was hired, which led to an increase in the stock of temporary teachers working in public schools.[31] Because the total number of public teachers remained constant in the period under analysis,[32] vacated positions were largely filled with temporary teachers. These hires led to a larger stock of temporary teachers. By 2004, just before the implementation of the reform, there were 55,000 temporary teachers, representing 17 percent of all teachers (Jerez, 2004). At that time, many of these temporary teachers had more than five years of experience working in public schools.

After the reform was implemented, three public calls occurred between 2004 and 2008. From these three public calls, approximately 80,000 new permanent teachers were hired.[33] By 2007, nearly 50,000 (or 91 percent out of the 55,000) of the tempo-

[31] Appendix Figure 3 shows a large drop in the number of teachers hired between 1996 and the year prior to the reform. Because the demand for teachers likely remained relatively constant we infer that between 2002, the year of the approval of the decree, and 2005, the actual year of implementation, a large number of temporary teachers were hired. These temporary teachers had to have a degree in education or pedagogy, similar to permanent teachers.

[32] Appendix Figure 4 plots a relatively constant stock of teachers throughout the analyzed period.

[33] See the number of vacancies available between 2004 and 2008 in Figure 1a.

rary teachers active in 2004 had been replaced by novice teachers (Jerez, 2004).[34,35] This resulted in a large turnover of the stock of teachers post-reform and in an increase of 20 percentage points in the fraction of teachers with less than five years of experience.[36] Consequently, after the first three public calls, in nearly 17 percent of classrooms, novice teachers with less experience replaced more experienced temporary teachers.

This pattern in teacher turnover continued after the first public calls. 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. After 2007, a third of all teachers leaving the profession each year were temporary teachers. These temporary teachers were considerably younger than permanent teachers, but they had already accumulated multiple years of teaching experience.[37]

The new merit-based system could have been designed to implement annual calls for teachers so that to continuously filled vacancies with permanent teachers. Instead, the system first filled vacancies with temporary instructors who were then replaced with permanent teachers in later public calls. The number of temporary teachers increased between public calls and then dropped when the next wave of new novice teachers was hired.[38] This explains why the share of teachers with fewer 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.

[34] Appendix Figure 5 shows the drop in the number of temporary teachers hired between 2004 (pre-reform) and 2007 (post-reform).

[35] It is reassuring to find that the number of teachers that left the public system matches the 80,000 vacancies filled with the first three public calls. The number of teachers that left can be computed as the number of temporary teachers who were recorded as working in 2004 but who had left by 2007 (50,000) plus the number of permanent teachers who retired (32,000). Appendix Figure 6a shows that approximately 8,000 permanent teachers retire each year. This would mean that about 32,000 teachers would be expected to have retired over this four-year period. The total number of teachers who left is then around 82,000, combining permanent and temporary teachers. This number is very close to the number of vacancies posted in the by 2007 in first three public calls in Figure 1a.

[36] See Figure 4b.

[37] See Appendix Figure 6a for the share of teachers who leave the profession and are temporary, Appendix Figure 6b for the age distribution, and Appendix Figure 6c for their accumulated years of experience.

[38] See Appendix Figure 5.

The reform appears to have only affected average teachers' experience and skills; it does not seem to have had any effect on other characteristics of the pool of public teachers. We do not observe any changes in the share of female teachers, the percentage of teachers with a college degree, or the average age at which teachers were hired.[39] In addition, while the reform increased teacher turnover, there was little change in the number of teachers in Colombia during this time period.[40] However, because the student population was declining, student-teacher ratios were monotonically decreasing during this time period.[41]

The merit-based reform affected fundamental inputs for students' learning, such as teachers' skills (captured by test score measures) and experience, while leaving other teacher characteristics unaffected. This suggests that the reform could have affected students' learning through the conduit of educators' level of experience. Novice teachers can be less effective at improving students' academic achievement (Rivkin, Hanushek and Kain, 2005; Hanushek and Rivkin, 2006; Araujo et al., 2016). By contrast, the effect of teachers' better academic credentials – as measured by the performance in teachers' test scores– 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. We then present and interpret the results obtained from this empirical approach.

[39] We present the evolution of teachers' characteristics before and after the reform in Appendix Figure 7.

[40] See Appendix Figure 4.

[41] See Appendix Figure 8 for an evolution of the student-teacher ratio.

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 distinction 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 college graduation. The variable Public_s is an indicator variable for whether school s is a public school. The parameters 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 for whether the student takes classes in the morning, in the 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 are 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 classic parallel-trends assumption in the absence of other policy changes that could have simultaneously affected public (or private) schools and

confounded the reform's effect. While the counterfactual parallel-trends assumption cannot be observed, the dynamic effects estimated in our event study strategy allow us to test for parallel trends prior to the reform and to provide some evidence for the validity of this assumption.

Consistency also requires that the reform did not affect private schools. While ultimately it is not possible to directly test this assumption, we provide two pieces of evidence consistent with it. First, the total number of teachers in private schools remained constant throughout the period of analysis; this suggests that there was not an influx of teachers previously working in the public sector being hired by private schools.[42] Second, as we show in Section 5.2, students' test scores in private schools remained fairly constant in the period under analysis.

Unlike previous work studying the effects of the reform, this paper focuses on estimating the aggregate, unconditional impacts that stem from changes in teacher composition at public schools as a result of the reform. Our strategy differs from that of 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 key characteristics were affected by the inflow of new teachers hired through the centralized system; this is certainly the case for the matter of experience. Also, their work imposes the strong assumption that vacancies across time are orthogonal to unobserved factors related to student learning. This assumption may be violated since successful candidates participating in each merit-based hiring process are allowed to fill a vacancy at their preferred school.[43]

[42] See Appendix Figure 4.

[43] 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 similar limitations.

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 students' test scores. Second, we document the reform's negative effect on the likelihood that students enroll in and graduate from 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.

We underscore four main observations of relevance to our findings. First, the gap in test scores between public and private students is close to zero and stable during the pre-treatment period (from 2000 to 2004). Test scores of students in private and public schools appear to follow a parallel trend. This supports the validity of our identification strategy. Second, 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 started to appear in 2005 when the first new teachers were being hired. This negative effect continued to grow until 2008, when the test scores of students in public schools were 0.12 of a standard deviation lower than those of private school students. Third, the negative effect appears to stabilize in 2008; with public school students scoring 0.10 of a standard deviation lower than private school students in the period from 2008 to 2013 (with the exception of 2011). After 2013 the negative effect of the reform appears to diminish. It settles at a point at which the scores of public school students are about 0.05 of a standard deviation lower than those of private school students. The level of the impact of the reform at that point is about roughly half of the effect that surfaced five years after the reform. Fourth, the results are very similar regardless of the specification used.[44]

[44] Appendix Figure 9 shows test scores in private and public schools for the period from 2000 to 2013 – a period in which exams where the most comparable. After that, comparisons across years are more difficult to do because of changes to the exam made by ICFES. While scores at private schools remained

We present static difference-in-differences point estimates for the scores on overall exam and on subject-specific tests 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 (ranging from 14 to 16 percent of a standard deviation) in mathematics and English proficiency. At the same time, however, estimates for all other subjects show negative effects (ranging from 2.6 to 6.6 percent of a standard deviation).[45]

Effect on college outcomes – Figure 6 shows the dynamic effects of the reform on students' college enrollment and college graduation. For both outcomes, we observe a negative effect that persists over time. The result on enrollment captures the impact of transitioning directly from high school to college – given that our outcome only records a value of one for students who enrolled in college within the next six months immediately after completing high school.[46] The dynamic effect we observe for college enrollment follows a similar pattern to the one that emerges for test scores, although the largest negative effect occurs somewhat later, in 2012. The negative effect on college enrollment begins in 2006 and continues to grow until 2012, when public school students are 5 percentage points less likely to enroll in college. This negative effect then converges back to zero, and by 2015 the measured negative effect of the reform on college enrollment is approximately 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 observed only after the cohort of students who took the high-school-exit exam

relatively constant, scores in public schools declined.

[45] In Appendix Figure 10 we present dynamic estimates of the effect by subject-specific exam scores.

[46] In Appendix Figure 11, we present results using wider time windows to define college enrollment – specifically, one-year and two-year enrollment rates. Results are similar for these alternative definitions of enrollment.

in 2006 has attended college. The negative impact on the cohorts of students we observe continued to grow over time. By the 2009 cohort (the last cohort for which we are able to calculate six-year graduation rates), results show that the reform had decreased the likelihood of a public school student graduating from college by over 2 percentage points. Many higher education institutions in the country offer admission based on the applicant's performance on the high school exit exam. Thus, the effect on college enrollment and graduation may be partly driven by the negative impact of the reform on students' high school test scores.

Table 3 summarizes the difference-in-differences results of the reform on college outcomes. As can be seen in the first six columns, for both immediate college enrollment rates and for college enrollment rates measured two years after high school graduation, the reform decreased enrollment by over 3 percentage points on average. For college enrollment, this estimated negative effect is equivalent to a 20 percent decrease in enrollment after six months; and 10 percent after two years of high school graduation. 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.[47] we present complementary results for these estimations in a constant sample of individuals across outcomes.

5.3 Teaching Experience and Student Outcomes

The reform negatively affected students' learning, even though it led school districts to hire teachers who ostensibly had higher measured cognitive skills (as measured by their own high school exit exam scores). This contradictory impact is likely driven by the increased exposure of students to teachers with lower levels of experience working in public schools. We provide two pieces of evidence consistent with this hypothesis.

First, we observe that the dynamic effects of the reform on students' test scores closely mirror the change in the fraction of teachers with fewer than five years of ex-

[47] In Appendix Table 3

perience (see Figures 4b and 5). Between 2004 and 2008, public schools received a large influx of novice teachers. Indeed, novice teachers, who represented 10 percent of the teachers in public schools prior to the reforms, represented 30 percent of all teachers by 2008. As such, a significant share of students were taught by teachers with little to no experience.[48] During this same time period the test scores of public school students relative to those of private school students declined by a little over 0.10 of a standard deviation. 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 of a standard deviation. As the fraction of novice teachers fell, so did the negative impact. The presence of novice teachers fell from 30 percent in 2008 to 20 percent in 2013 (though not back to the 10 percent level observed prior to the reform). The negative effect of the reform on students' scores follows a similar pattern. As the share of novice teachers falls, so does the negative impact as measured by exam scores; the impact shifts from -0.10 of a standard deviation in 2011 to -0.05 of a standard deviation in 2013. Over the same period, the results on college enrollment mirror this same pattern and effect size, as evidenced by Figure 6a. The mirrored patterns and size of the patterns that emerge between the fraction of novice teachers in the system and the dynamic effects of the reform on student achievement and the pursuit of higher education suggest that teaching experience likely plays a prominent role in explaining the negative effects of the reform on students' academic outcomes.

Second, we find that the negative effect on students' learning was larger at public schools that were more exposed to novice teachers after the reform. We reach this conclusion by proceeding as follows: we calculate the baseline fraction of teachers with fewer than five years of experience in 2007 in each school.[49] Then we interact that fraction with an indicator variable equal to one if the student attended a public school and an indicator variable equal to one if the year in which the test scores are observed

[48]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 the effect of novice teachers instructing students, not from firing and hiring new teachers per se.

[49]To compute the baseline share of novice teachers, we compute the percentage of teachers hired during the five years prior to 2007 by using the 2007 and 2008 cross-sections of the teachers' census.

corresponds to the post-reform period (the specification also includes the same set of controls of Equation (1)). Table 4 reports the coefficients of the interaction between the public school and the post-reform indicators and the coefficients of the triple interaction between those two indicator variables and the fraction of novice teachers. It is important to highlight that the variation in exposure to novice teachers is not necessarily exogenous. The fraction of novice teachers in 2007 could partly reflect an endogenous response to the reform, and it could also be correlated with other school, student, or location characteristics. For these reason, the following set of results should be interpreted with caution.

For the fully saturated model in the third column of each panel, the coefficient on the interaction between the public school and post-reform indicators is only slightly smaller than the effect found in our main specification; with a negative effect on overall test scores of 0.075 (Table 4) and 0.082 (Table 2) of a standard deviation. The coefficient of the triple interaction reported in the first row shows that the negative effects are larger in schools with higher fraction of novice teachers in 2008. Table 5 also reports the heterogeneous effects of the reform on college outcomes. For example, in schools that had no novice teachers in 2007, the reform reduced immediate college enrollment by 2.6 percentage points, reduced college enrollment two years later by 2.7 percentage points, and reduced college graduation rates by 0.7 of a percentage point. By contrast, for students in schools with only novice teachers, the negative effects were of more than twice the size in all three outcomes. Dynamic effects estimated for schools with a high, medium, and low fraction of novice teachers in 2007 are shown in Appendix Figure 12.

These two pieces of evidence suggest that student learning could have been affected by students' increased exposure to teachers with lower levels of experience – with such exposure especially pronounced in schools with greater levels of teacher turnover. Thus, policies that affect teacher retention and turnover rates may decrease student learning, even if such policies stem from efforts to select educators from a higher cognitively skilled pool of applicants.

6 Discussion and Policy Recommendations

Our results provide direct insights for educational policies – both those that determine how teachers are hired and those that influence teacher retention and turnover.

During the last two decades, many countries, particularly in Latin America, have introduced nationally standardized, merit-based policies to regulate the process for hiring teachers; many of these efforts are similar to the reform undertaken in Colombia (Cruz-Aguayo, Hincapie and Rodriguez, 2020; Elacqua et al., 2018). These systems typically use a centralized hiring system in which public school vacancies are allocated among candidates based on certain criteria, such as passing a standardized exam. These criteria can heavily weight a few ex ante pieces of information – such as teachers' own cognitive skills, subject knowledge, and/or teaching ability as measured by standardized test scores – at the expense of other information – such as prior teaching experience or actual performance in a classroom. Therefore, merit-based policies may succeed in increasing teachers' average pre-college test scores, and their cognitive skills, but also reduce the weight of other non-targeted teacher characteristics (such as experience) in the selection process.

Policies that more heavily weight some teacher characteristics at the expense of other important characteristics might fail to improve students' academic achievement. A shift to criteria that in which ex ante information trumps experience can backfire both because novice teachers are less effective instructors (Hanushek, 1971; Rivkin, Hanushek and Kain, 2005) and because 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).[50] 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, policymak-

[50] 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 the evidence we provide in this paper.

ers may want to reduce the emphasis on a small number of ex ante measures. Policymakers should consider broadening the screening strategies and potentially focusing more on ex post measures to improve teacher quality. For example, districts could offer permanent contracts based on measures of teachers' effectiveness (e.g., through value-added measures based on students' outcomes).

Our results also inform broader educational policies that influence teacher retention and turnover. We show that the large shock that occurred by replacing experienced teachers with novice teachers negatively impacted students during the first few years of the reform, even though these novice teachers had higher measured cognitive skills. This change in teacher experience was the result of processes that heavily relied on filling vacancies with less experienced teachers, rather than by filling positions in ways that were designed to cultivate stability and help teachers improve and build on experience in positions that were likely to allow for permanency in employment and career advances. As a result, the vacancies were first filled with inexperienced temporary teachers and then refilled with permanent teachers from the next public call. That meant that filling each teaching vacancy required the training of not just one new teacher but two – leading students to learn less on average, from not just one but two different instructors. This occurred as a large shock in the first few years of the reform, and it 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 steady rate. Each vacancy required a temporary teacher to fill the position for one to five years and then a permanent teacher was hired to fill the position. As a result of such practices, students were more likely to have been taught by a teacher who was in the early phase of the learning curve in their educational careers.[51]

Our results imply that an important way of improving student outcomes is to keep teachers in the profession for extended periods of time, therefore reducing the number of students who are taught by novice teachers in any given year. While teacher

[51]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).

turnover at the school or district level may play an important role, these results shine a light on the importance of teacher turnover within the profession more widely.

7 Conclusion

Teachers are the most relevant factor for human-capital development in education systems. As such, education authorities across districts and countries implement policies to improve teacher quality and, in turn, student outcomes. Policy changes to improve teacher quality typically focus on new hires rather than on current instructors. However, identifying effective teachers ex ante can be a complex and challenging task, mainly because value-added measures of teacher quality are not correlated with observable characteristics such as education level, licensure, IQ scores, and scores rating the performance of teacher candidates from screening and hiring processes.

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. Our findings show that even though the reform led the system to hire more teachers with higher cognitive skills, as measured by the teachers' own scores on standardized exams, the reform also led to poorer outcomes for students –as measured by their high school exit exams, and rates of college enrollment and college graduation. Cognitive skills increased sharply among teachers hired in the wake of the reform, with teachers' test scores increasing by 17 percentile points. At the same time, the share of teachers with little to no experience also sharply increased, from 10 percent of the teaching staff to 30 percent at its peak. Meanwhile, students' test scores decreased by about 8.2 percent of a standard deviation. The probability that a student enrolled in college dropped by 20 percent, and the probability that a student graduated from college dropped by 10 percent. The negative effects on student achievement and educational progression that we document are in line with the evidence provided by the literature. Such literature suggests that: i) teacher quality is not correlated with teachers' test scores or with scores rating information gathered before teachers are hired; and ii) teacher quality is

typically lower during the first five years of teaching.

Despite concerted effort, increased spending, and the best of intentions, the merit-based teacher hiring reform reduced students' academic outcomes. The likely reason for this was the new selection system 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. Our results suggest that future education reforms and hiring systems should carefully evaluate what ex ante information they use when designing their policies and perhaps, whenever possible, combine that information with ex post value-added measures to make retention and promotion decisions.

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.

Jerez, Angela. 2004. "El Dilema de los Maestros Provisionales." *El Tiempo*.

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.

Robalino, Magaly, Anton Körner, Murillo Torrecilla, and Francisco Javier. 2007. *Evaluación del Desempeño y Carrera Profesional Docente. Un Estudio Comparado entre 50 Países de América y Europa.* Santiago de Chile: UNESCO.

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.

Tiempo, Redacción El. 1996. "Anulado Concurso para Docentes." *El Tiempo*.

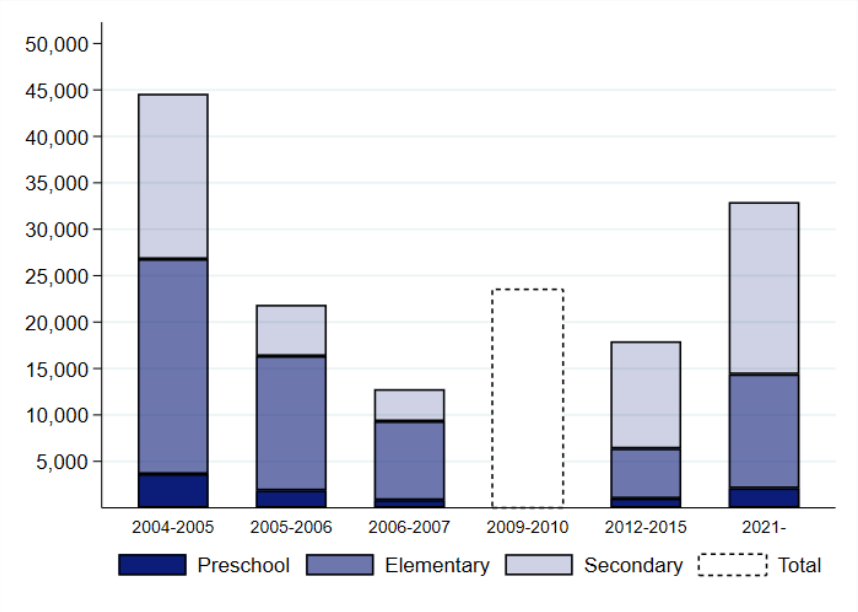
Velasquez, Maria, Oscar Arcos, Adriana Rodriguez, and Fabian Gonzalez. 2010. *Docentes y Directivos Docentes. Acerca de su Condicion y Perfiles sobre Resultados de Convocatorias de Concursos de Meritos.* CID, Universidad Nacional de Colombia.

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

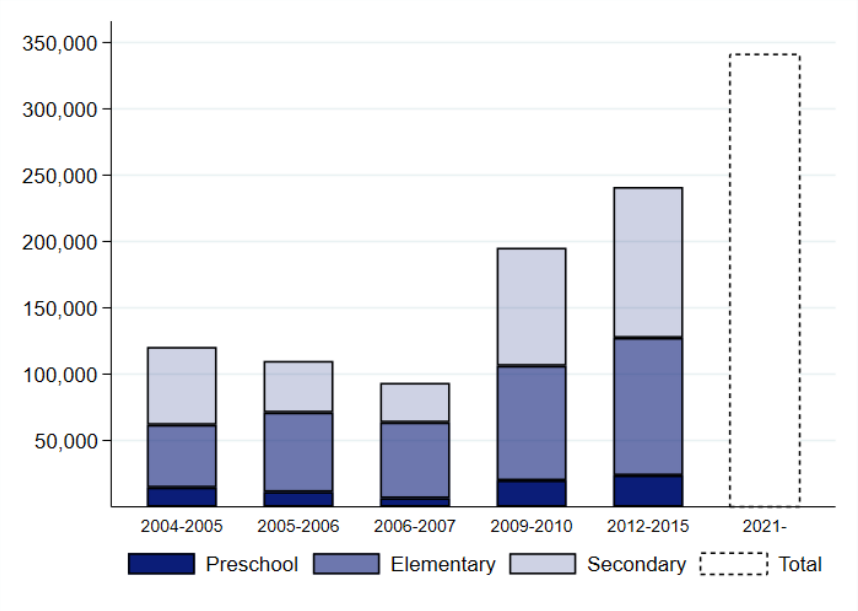
Figures and Tables

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

(a) Teaching Vacancies

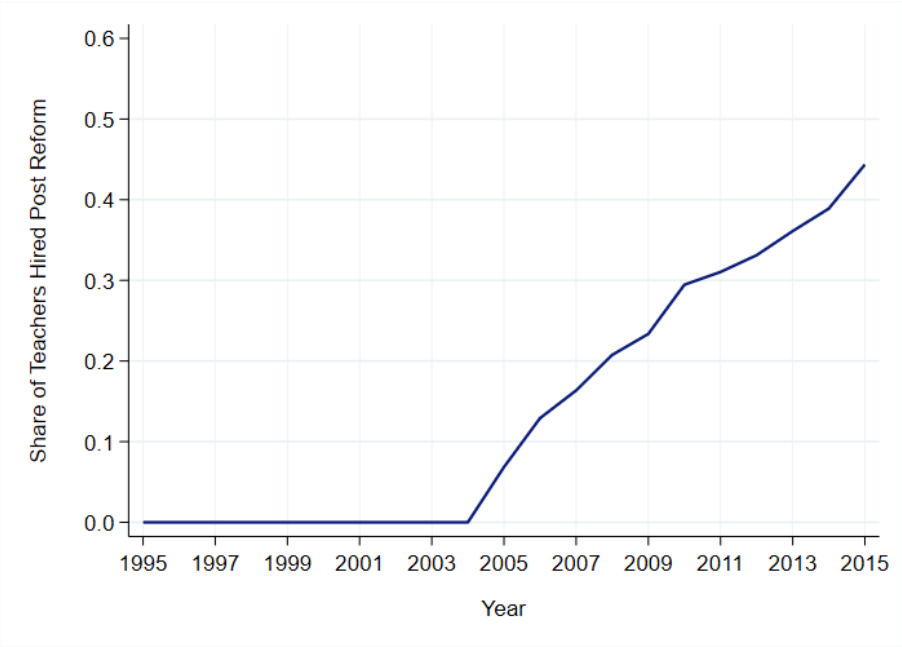


(b) Applicants to Teaching Positions



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.

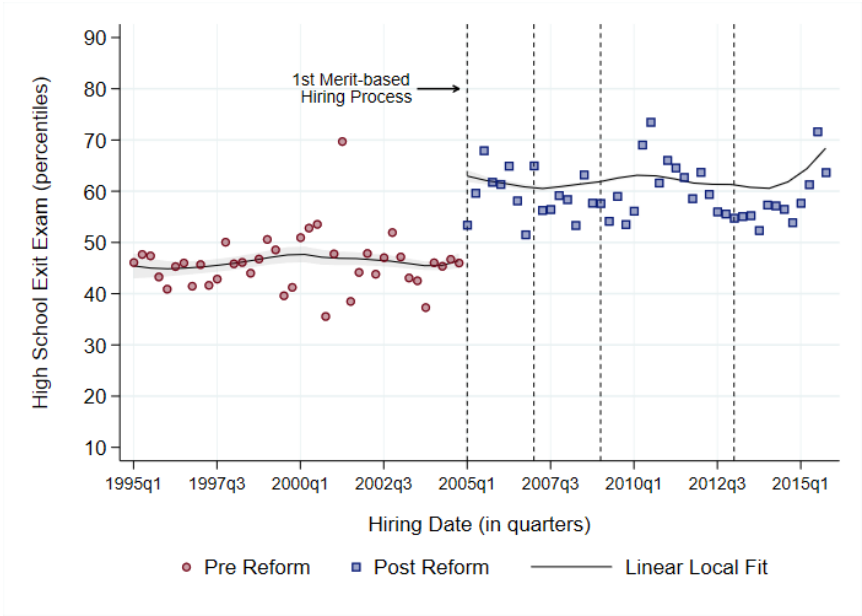
Figure 2: Share of Teachers Hired After the Reform's Implementation



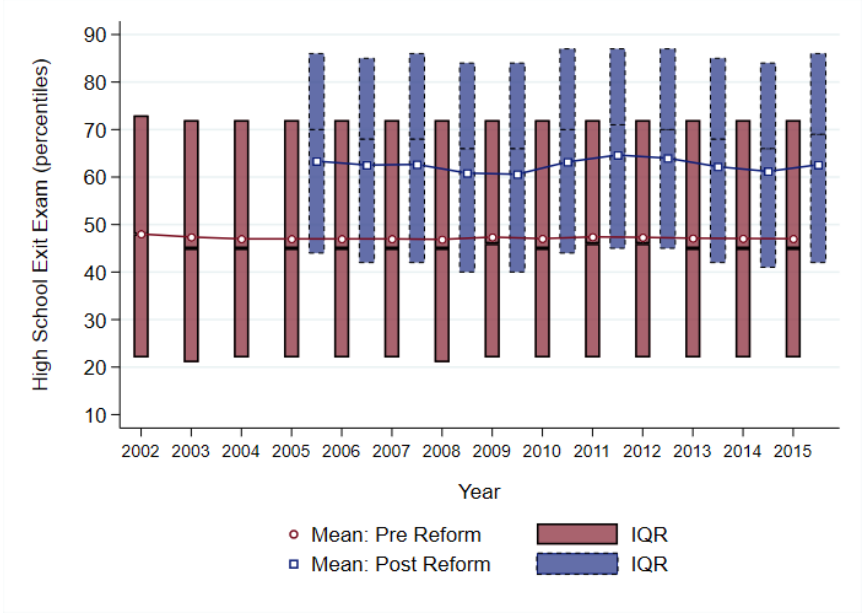
Notes. The solid line represents the annual share of teachers hired after the reform was implemented in 2005.

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

(a) Test Scores by Teacher's Hiring Date



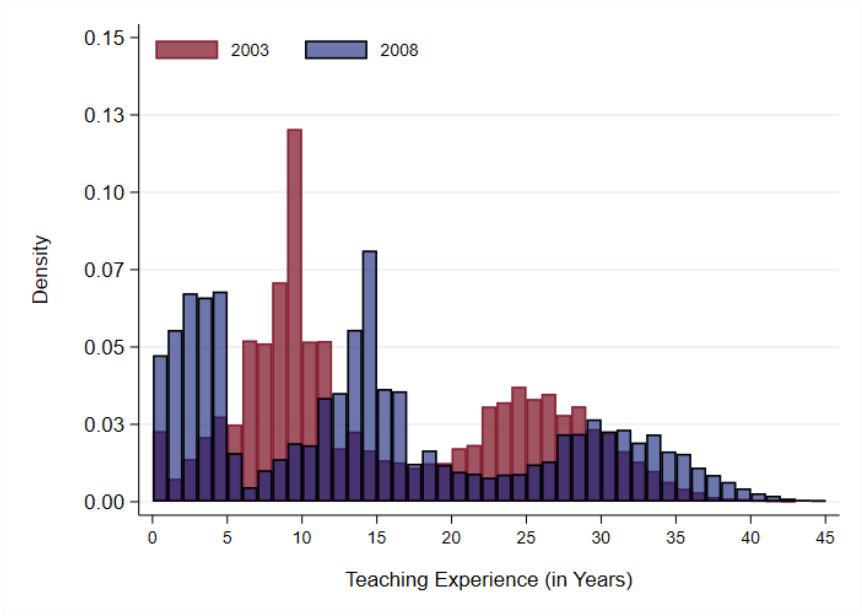
(b) Distribution Pre- and Post- Merit System



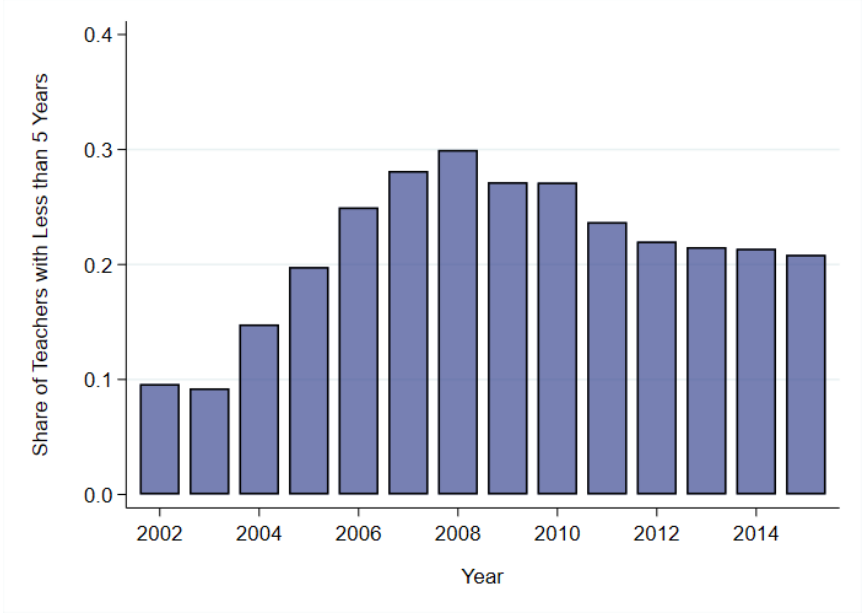
Notes. Panel 3a plots the average percentile in the high school exit exam of teachers hired in each quarter between 1995 and 2015. Solid lines represent local linear regressions fitted using individual-level data of teachers hired before and after 2005. Confidence intervals at the 95% level are displayed 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 Before and After the Reform

(a) Experience Distribution Pre and Post Reform

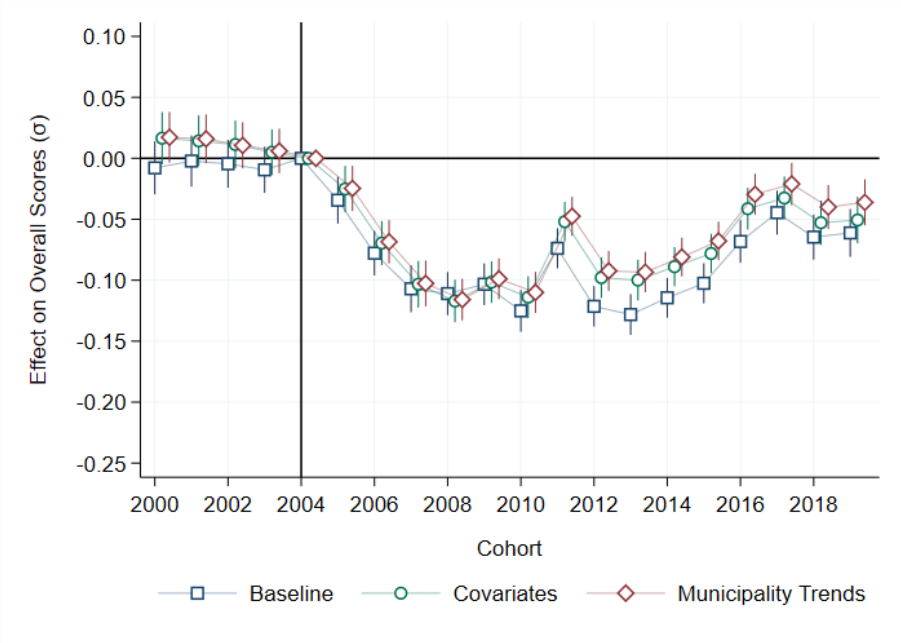


(b) Share of Teachers with Less than Five Years of Experience



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 fewer 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.

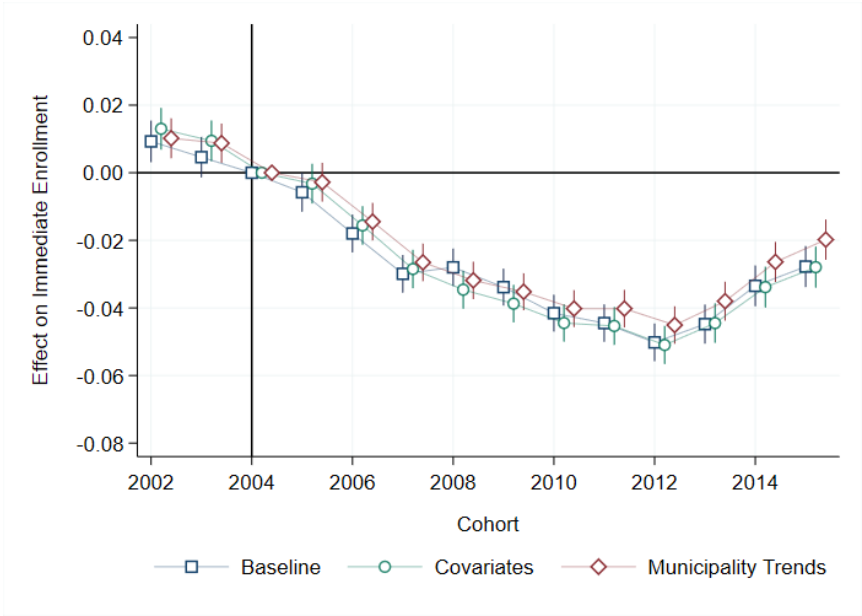
Figure 5: Dynamic Effects of a Merit-based Teacher-hiring Policy on Students' Test Scores



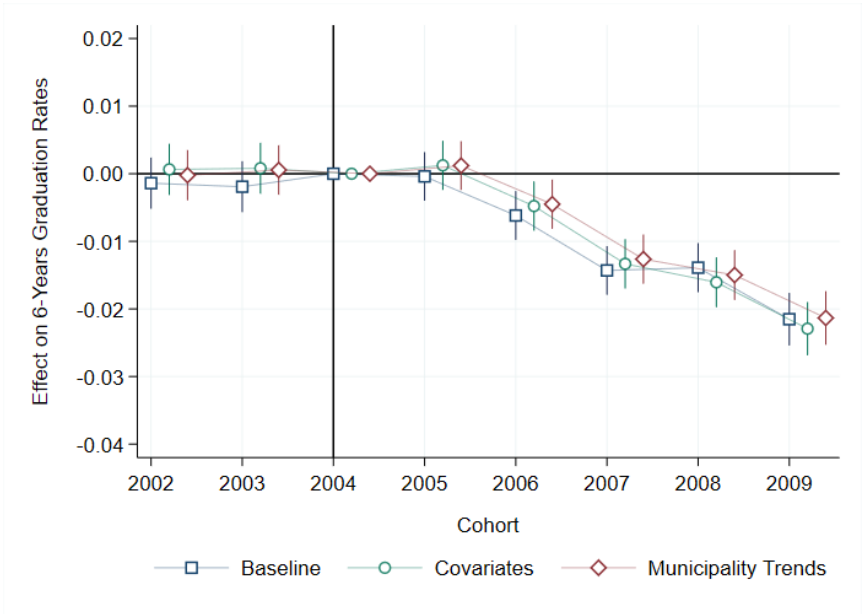
Notes. Ordinary-least-squares estimates of the dynamic effects, δ_τ, of equation 1. The outcome variable is the overall score on the high school exit exam. Overall scores are computed as the average performance in five subject exams: reading comprehension, mathematics, natural sciences, social sciences, 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 or not it was completed. 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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

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. Two-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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 4: Heterogeneous Effects of a Merit-based Hiring Policy on Student 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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

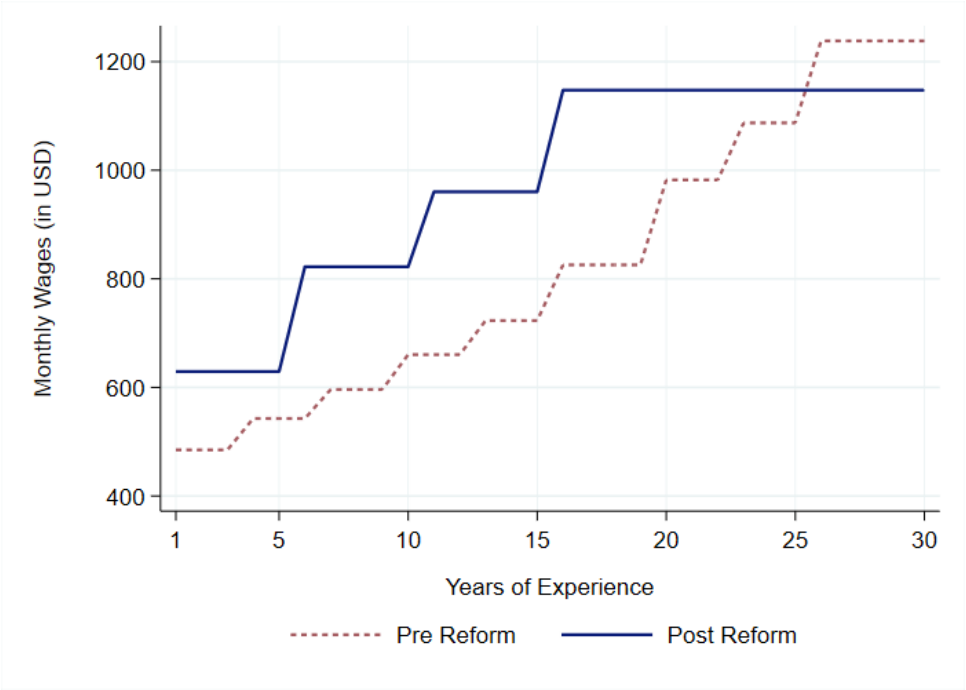
Table 5: Heterogeneous Effects of a Merit-based Hiring Policy on Students' College Outcomes

	Dependent Variable :								
	College Enrollment						College Graduation		
	Immediate			2-year					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Public $\times \mathbb{1}[t \geq 2005] \times$ 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. Two-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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

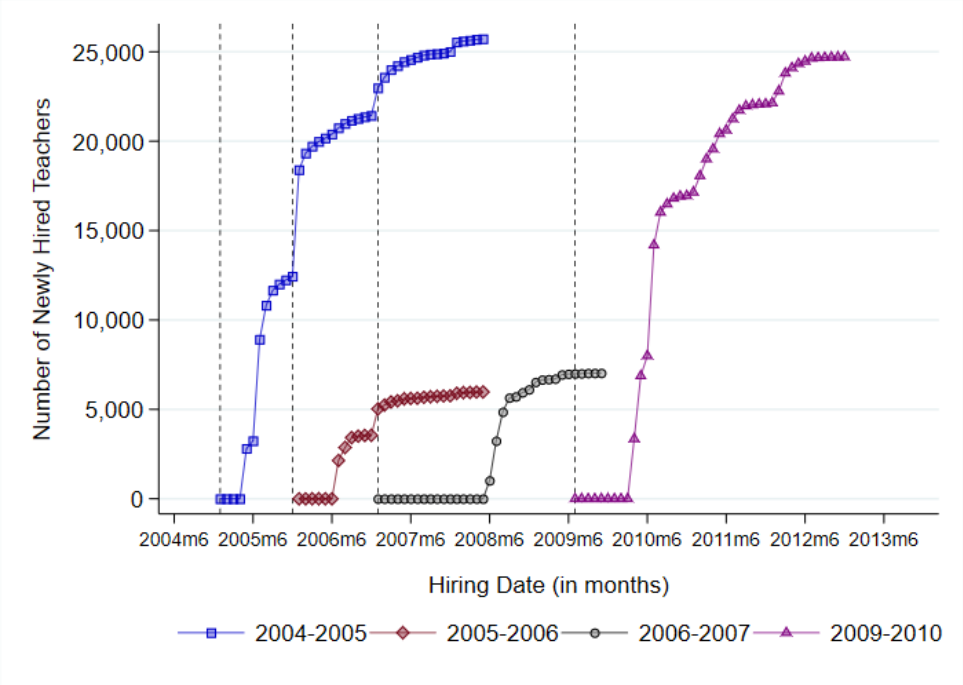
A Additional Figures and Tables

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



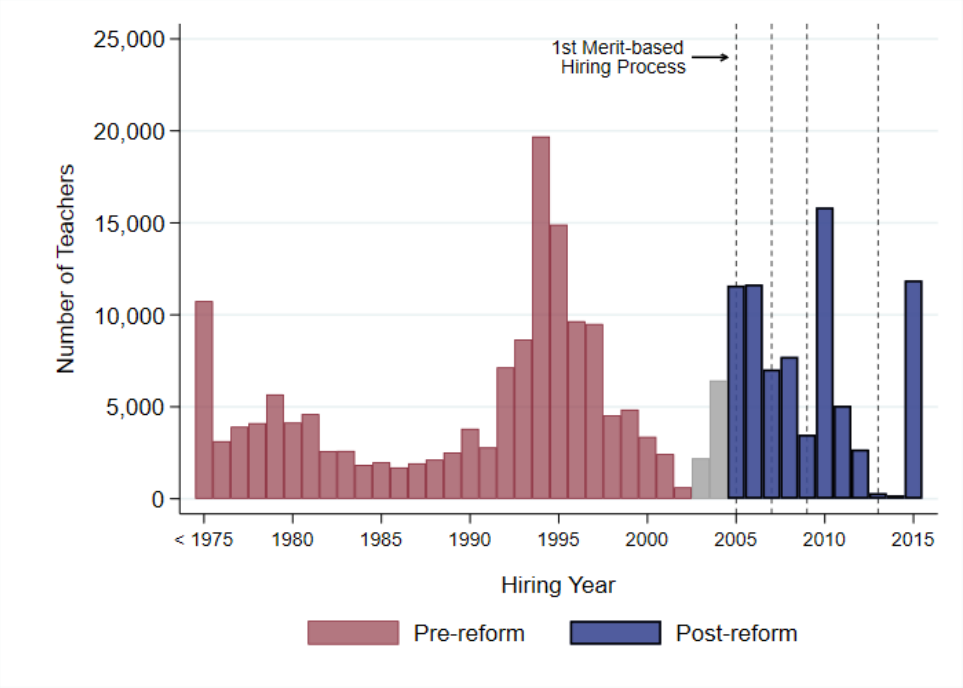
Notes. Wage-experience profiles are computed based on public school teachers' salaries in 2010, found in Decrees 1367 and 1369. The profile of teachers hired post-reform assumes promotions every five years. The daily average of the exchange rate in 2010, 1 $USD = 1898.7 $COP, is used to present salaries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

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



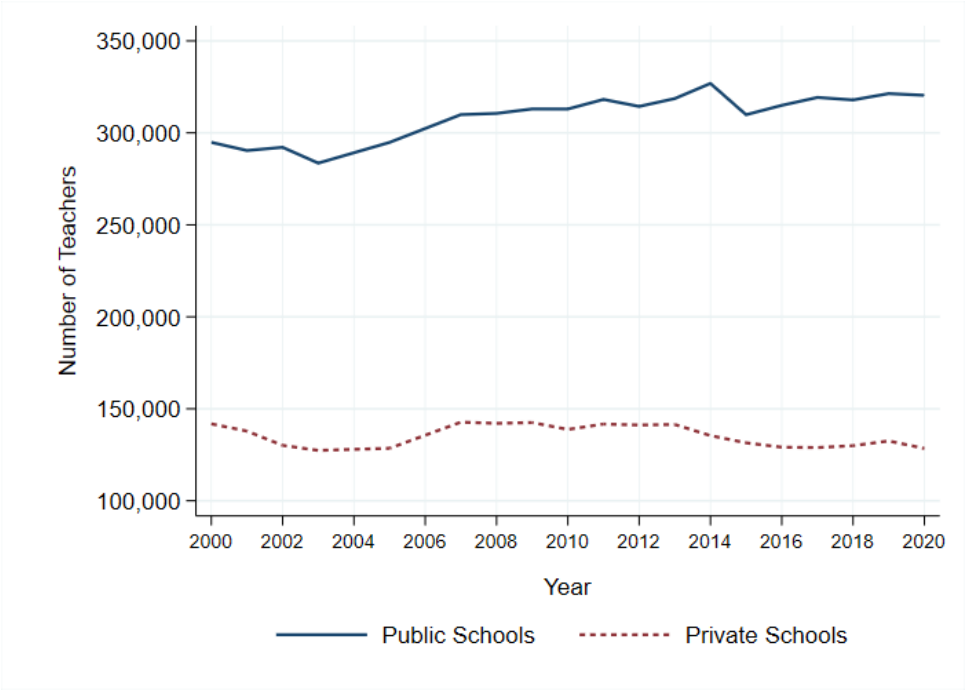
Notes. This figure plots the (cumulative) number of teachers hired across time in each merit-based hiring process between 2004 and 2014. Vertical dashed lines represent the month when individuals hired took the entry exam used by the Colombian government to screen applicants.

Appendix Figure 3: Stock of Permanent Teachers by Hiring Date



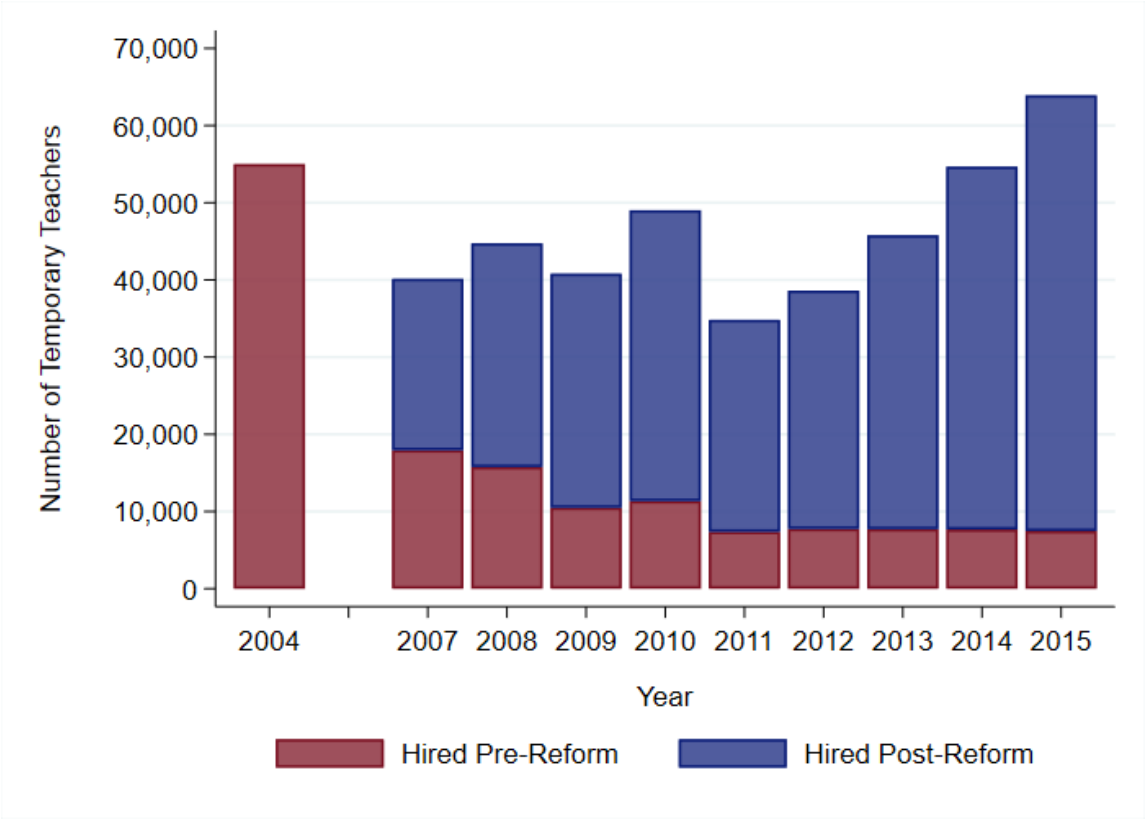
Notes. This figure plots the density of permanent teachers working in 2015 by hiring date. Dashed vertical lines represent the years when a new merit-based hiring process starts. The gray bars correspond to teachers who started working after the new hiring policy was introduced in 2002 but before the first merit-based process was carried out in 2005. Most likely, these teachers initially held a temporary contract but transitioned to a permanent one by 2015.

Appendix Figure 4: Number of Teachers by Type of School



Notes. This figure 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).

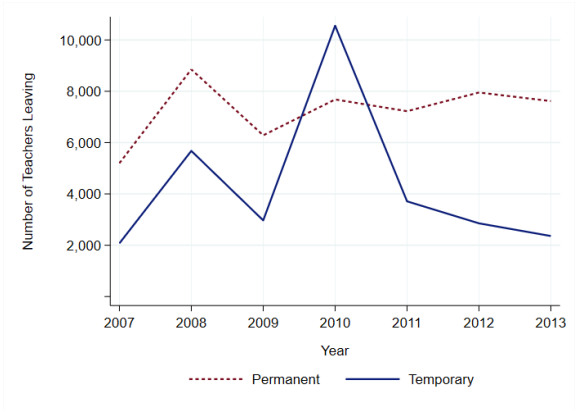
Appendix Figure 5: Stock of Temporary Teachers Hired Before and After the Reform



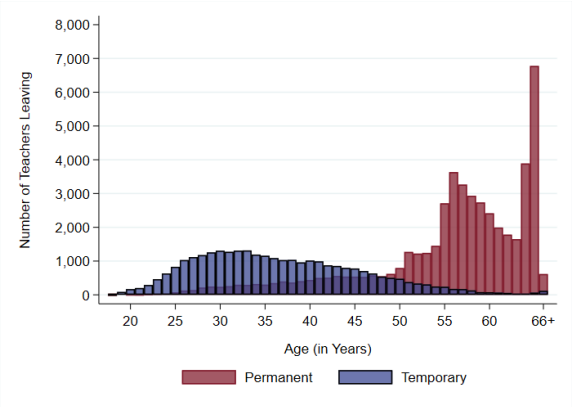
Notes. This figure displays the number of temporary teachers working at public schools in any given year between 2007 and 2016. The data from 2007 to 2015 correspond to the census of teachers. The data point for 2004 comes from Jerez (2004).

Appendix Figure 6: Teachers Leaving Public School Positions Over Time by Type of Contract

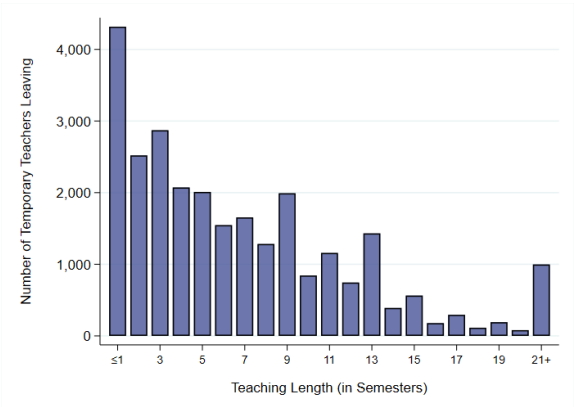
(a) Number of Teachers Leaving Across Time



(b) Distribution of Teachers by Age



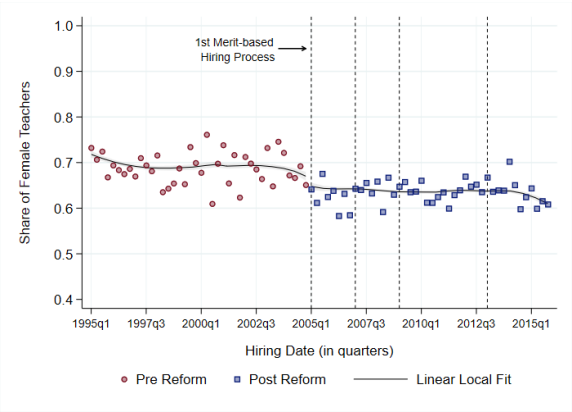
(c) Temporary Teachers' Work Experience at Moment of Leaving



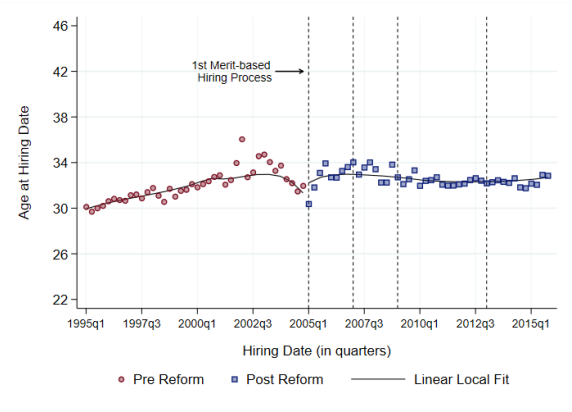
Notes. This figure plots information on individuals who stop working as public school teachers between 2007 and 2013. We assume a teacher leaves or stops working when this is not observed during two consecutive years in the teacher census data. Panel 6a presents the number of teachers leaving across time by type of contract (i.e., permanent or termporary). Panel 6b plots the number of teachers leaving by age and type of contract. Panel 6c plots the distribution of the time that temporary teachers work at public schools before leaving (in semesters).

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

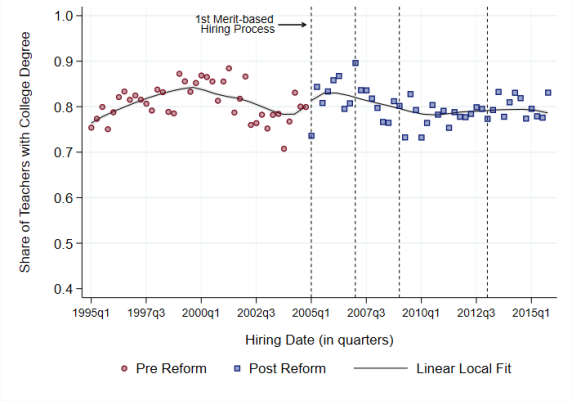
(a) Gender



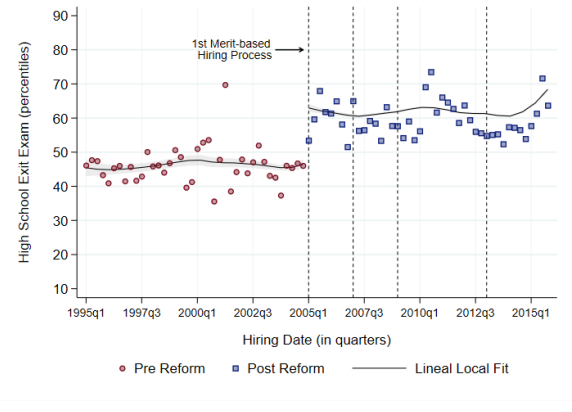
(b) Age at Hiring Date



(c) College Degree

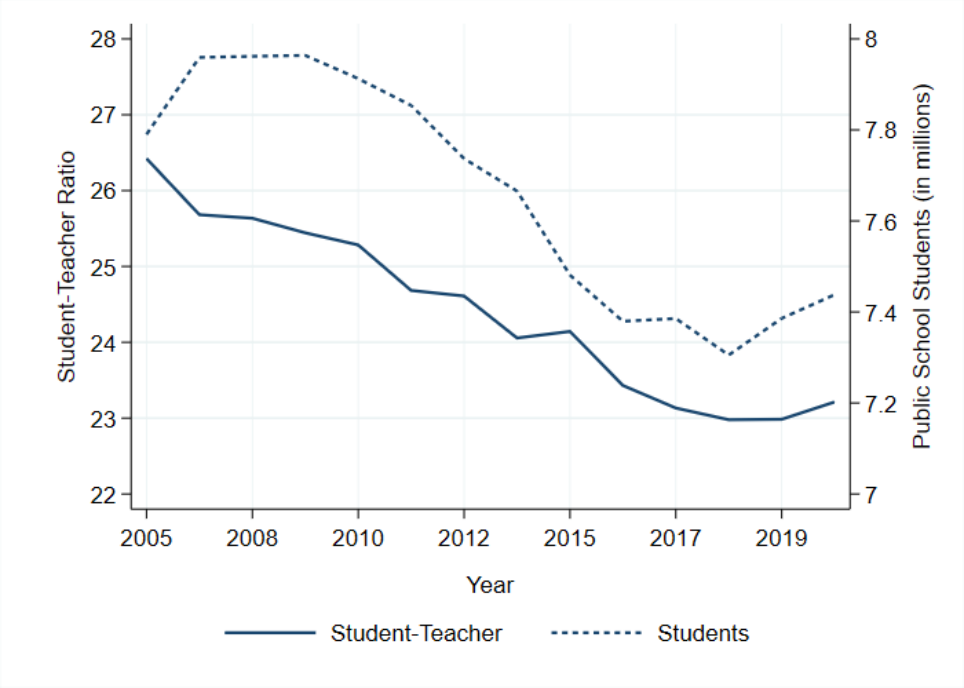


(d) Test Scores



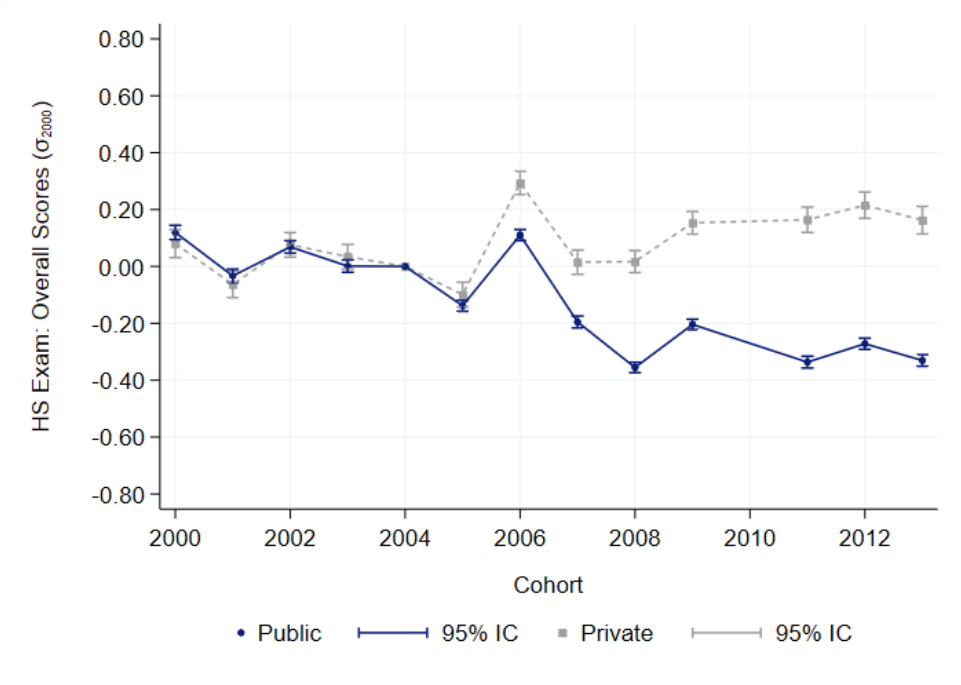
Notes. Plotted dots in Panel 7a represent the share of female teachers hired in the same quarter between 1995 and 2015. Panel 7b plots the average age (at hiring) of teachers hired in the same quarter. Panel 7c plots the share of teachers holding a college degree. Panel 7d plots the average percentile 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: Student-Teacher Ratio at Public Schools



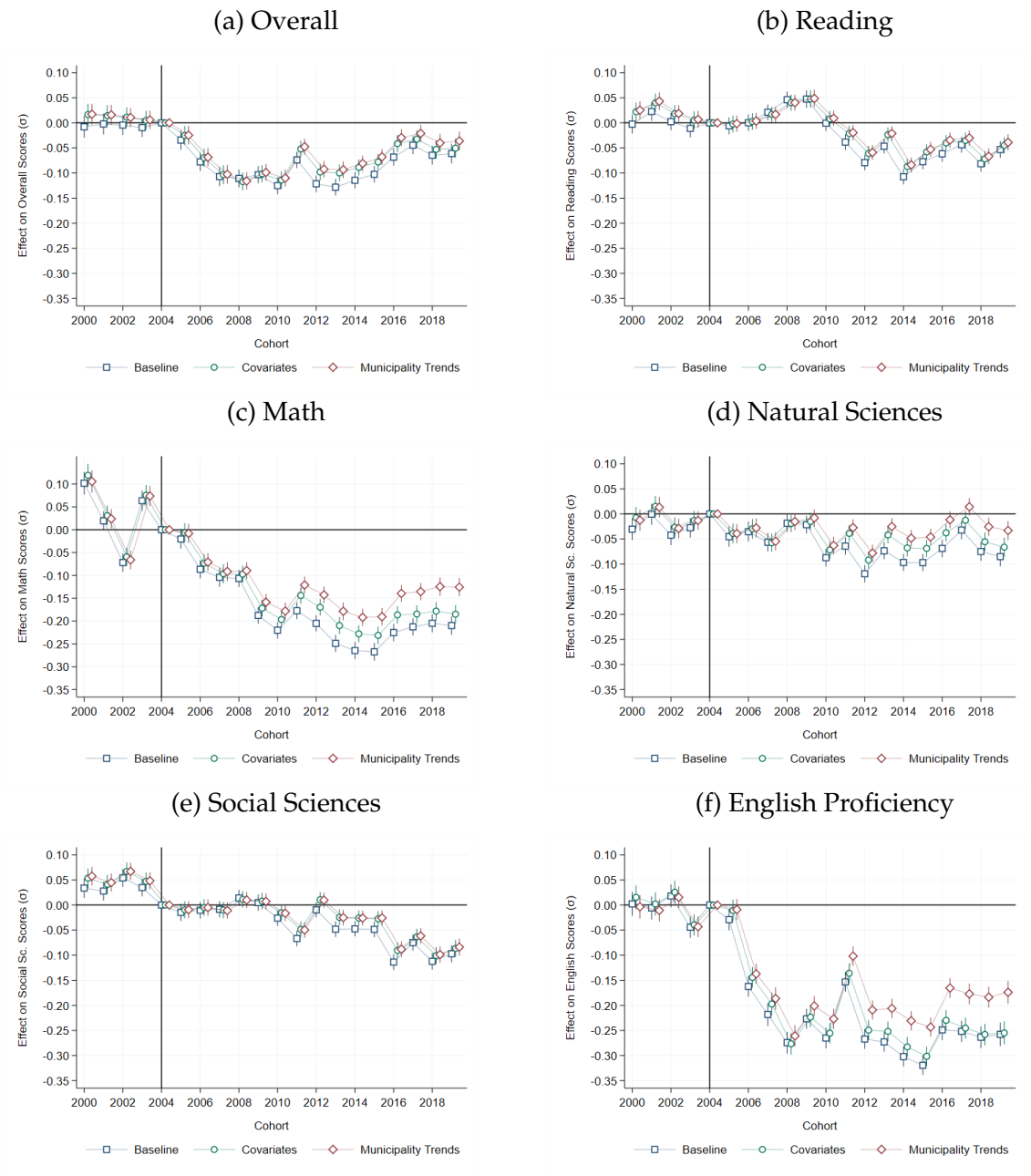
Notes. This figure displays the student-teacher ratio and the number of students enrolled at public schools between 2005 and 2019. Enrollment information is based on reports publicly available by DANE.

Appendix Figure 9: Stable Unit Treatment Value Assumption



Notes. Ordinary Least Squares estimates of independent regressions for students in public and private schools. The outcome variable is the residualized score of the overall performance in the high school exit exam on students' characteristics (i.e, age, gender, socioeconomic stratum, and schooling time) and municipality linear trends. Students' overall performance of each cohort was standardized with respect to the year 2000. Plotted coefficients correspond to cohort fixed effects. Regressions include school fixed effects. 95% confidence intervals are displayed for each coefficient and were computed from clustered standard errors at the school $\times$ year level.

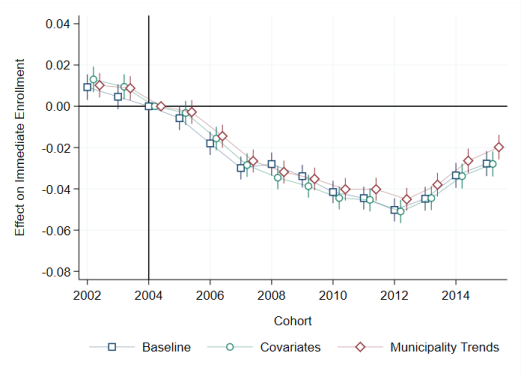
Appendix Figure 10: Dynamic Effects of a Merit-Based Teacher Hiring Policy on Students' Test Scores



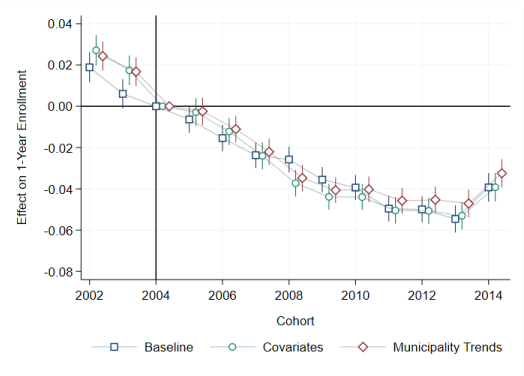
Notes. Ordinary Least Squares estimates of the dynamic effects, δ_τ, of equation 1. Outcome variables correspond to overall performance on the high school exit exam and test scores in all evaluated subjects. Overall scores are computed as the average performance in five subject exams: reading comprehension, 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 11: Dynamic Effects of a Merit-Based Teacher Hiring Policy on Students' College Enrollment Rates

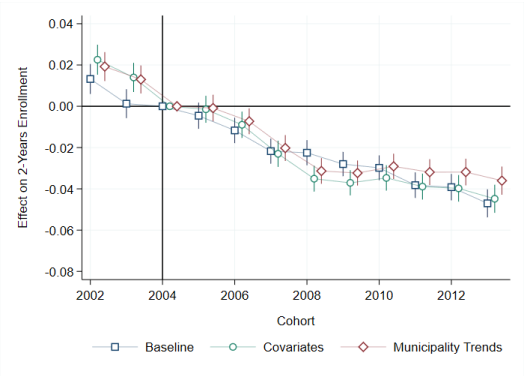
(a) Immediate Enrollment



(b) Enrollment within One Year



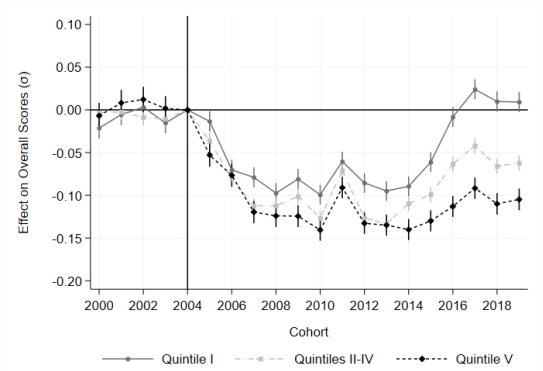
(c) Enrollment within Two Years



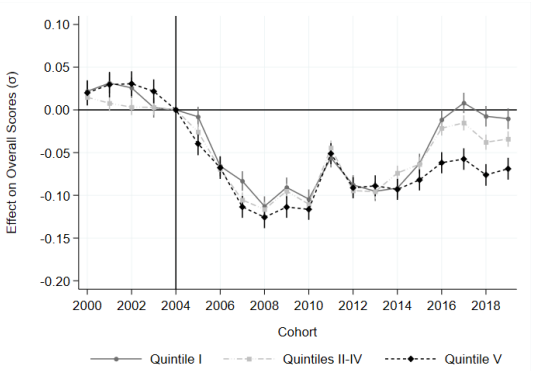
Notes. Ordinary Least Squares estimates of the dynamic effects, δ_τ, of equation 1. Outcome variables are indicators for whether a student enrolls in college within six months (immediate), one year, or two years, after taking the high school exit exam. The baseline specification includes school and year 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 12: Dynamic Heterogeneous Effects of a Merit-Based Teacher Hiring Policy on Students' Outcomes

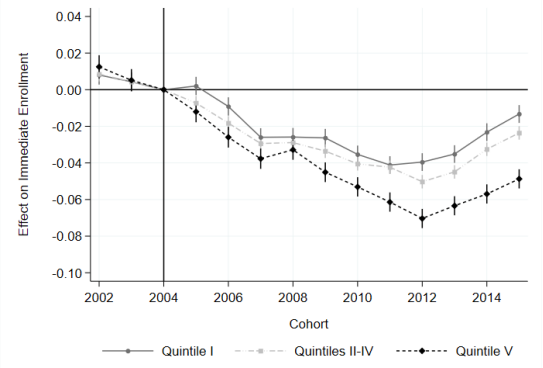
(a) Baseline Model: Overall Scores



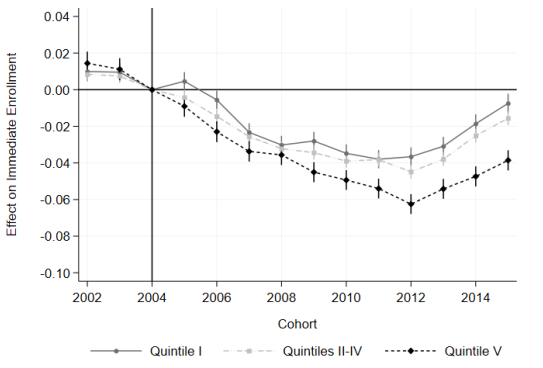
(b) Full Model: Overall Scores



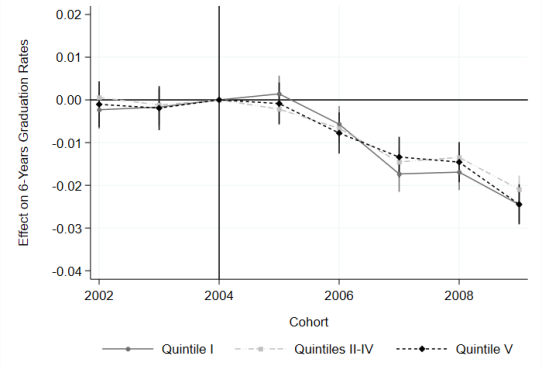
(c) Baseline Model: College Enrollment



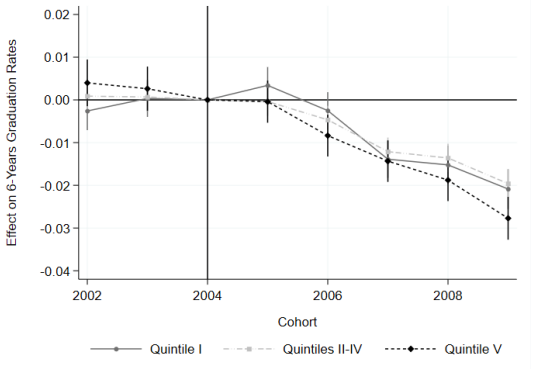
(d) Full Model: College Enrollment



(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 × 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 × year level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.