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*

August 29, 2024

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. The implementation of the policy led to many experienced contract teachers being replaced by high exam-performing novice teachers. 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, particularly at the expense of teacher experience, 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). We thank Juan Francisco Santos for excellent research assistance. 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 to improve educational outcomes and close the gaps with high-performing regions. Because teacher quality is 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. Several countries have implemented nationwide merit-based hiring systems to select new teachers based on an array of information that often includes test scores on standardized teaching aptitude 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). Hiring systems that heavily weight specific indicators – such as licensing requirements, educational attainment, or performance on standardized exams – may be counterproductive, especially when they lead to decisions that ignore or downplay other dimensions that are more predictive of teacher quality, such as specific human capital acquired through experience on the job (Staiger and Rockoff, 2010). In such cases, teacher-hiring policies aimed at improving incoming teacher quality could have a negative impact on students.

In this paper, we examine the impacts on the attributes of teachers and the

[1]Several Latin American countries (including Chile, Colombia, Ecuador, Mexico, and Peru) introduced teacher entry exams. Other countries using such systems include Germany, Belgium, Austria, and Cyprus (Robalino et al., 2007).

achievements of students from a nationwide reform that sought to enhance teacher quality and improve student learning in the public schools of Colombia. The reform raised teacher salaries and introduced a centralized, merit-based teacher-hiring system that tied teacher-hiring decisions in the civil service to candidates' performance on a national standardized exam evaluating subject-specific knowledge and teaching aptitude. As part of the implementation of the policy, many experienced public school teachers without a civil service long-term contract (i.e., contract teachers) were replaced by novice teachers who performed well on the standardized exam. This civil service 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. Our empirical design assumes parallel trends and stability of the control group. We show evidence of pre-reform parallel trends in students' test scores and college graduation. In addition, we argue that in our setting both teachers and students faced large costs of switching from public to private schools and vice versa. We provide evidence consistent with this claim.

We find that, while the education level, gender, and age of teachers did not change with the reform, 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. After implementing the reform, incoming teachers' test scores rose by 17 percentile points.

The implementation of the reform also led the government to replace many contract teachers who had several years of teaching experience in public schools. Within two years, public schools had replaced nearly 40,000 contract teachers – 13 percent of all teachers – who had been hired before the policy change. In addition, after the first two years of the reform, districts continued to annually replace more than 4,000 contract teachers – representing 37 percent of the teachers who left the profession each year. As a result, while the newly hired teachers had substantially higher pre-college test scores, they were also 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.

We find that the introduction of the teacher-hiring system decreased the overall performance of public school students on high school exit exams and college enrollment and graduation. In the 15 years following the reform, students' average scores fell by 8.2 percent of a standard deviation – roughly equivalent to the decrease in test scores documented in the literature when there is 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 found for exam scores in all other subjects: reading, natural sciences, and social sciences. The new teacher-hiring system also reduced 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. Among those students who had attended public schools after the reform was implemented, there was a 0.9 percentage point decline in the likelihood of graduating from college.

The negative effects of the reform on students' outcomes appear to be driven by

an increase in 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 and evidence that the effects are more pronounced among schools where teacher turnover is expected to be higher (i.e., schools that are more in demand by teachers). Jointly, these results suggest that a larger exposure to inexperienced teachers post-reform can help explain students' learning losses.

Our findings highlight the risks of policies that heavily rely on ex-ante measures of teacher quality for screening and hiring. While the reform led to hiring new teachers with better measures of cognitive skills, these measures did not correlate with improved student learning. The focus on performance in teacher aptitude exams overshadowed other important criteria, such as teaching experience, which is crucial for student achievement (Staiger and Rockoff, 2010; Araujo et al., 2016). Our findings serve as a cautionary tale about the unintended negative effects of such reforms, especially in contexts with limited institutional capacity. Teacher-hiring policies should carefully evaluate the ex-ante information used for hiring decisions, emphasizing the value of experience. Additionally, reforms to hiring practices can trigger other forces, such as increased teacher turnover, which reduces the overall stock of teaching experience.

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.[2] 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. These results are potentially biased due to strong identification assumptions (i.e., that 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 affected by the reform.

Our paper also relates to empirical work regarding the effects of teacher characteristics on student learning. Research in this area has consistently found evidence that teacher quality, as measured by value-added measures, explains a significant fraction of the variation in students' academic performance, education

[2]Under these circumstances, it might be possible that potential learning gains due to teachers hired under a ruled-based scheme, as observed in Mexico and documented by Estrada (2019), can be dampened by some unintended consequences.

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). Beyond quality, experience seems to also matter. Teachers are typically less effective during the first few years of their careers (Hanushek and Rivkin, 2006) and then improve by combining general and specific human capital (Ost, 2014); moreover, students exposed to newly hired teachers can suffer negative learning impacts (Staiger and Rockoff, 2010; Araujo et al., 2016). Some other work on this area suggests that credentials (Rockoff et al., 2011; Clotfelter, Ladd and Vigdor, 2010; Kane, Rockoff and Staiger, 2008) and content knowledge (Metzler and Woessmann, 2012; Bold et al., 2019, 2017) are predictors of student success. However, observable characteristics have also been shown to have a limited scope on enhancing student achievement in other contexts (Rivkin, Hanushek and Kain, 2005; Hanushek and Rivkin, 2006; Jackson, Rockoff and Staiger, 2014; Araujo et al., 2016; Hanushek, 2011). In a close paper to this, Cruz-Aguayo, Ibarrarán and Schady (2017) provide evidence from Ecuador suggesting that students assigned to teachers who performed better on an exam evaluating knowledge and pedagogical abilities do not experience learning gains. We contribute to this literature by showing how teacher characteristics –such as teachers' test score measures and experience– interact when designing selection procedures based fully on merit.

Finally, this paper also relates to the literature on scalability in education. A common result shows that policies that seem to work on a small scale do not replicate when implemented at a larger scale or can have a variety of unintended consequences (Al-Ubaydli, List and Suskind, 2017; Al-Ubaydli et al., 2021; Khanna, 2023; Muralidharan and Sundararaman, 2015; Araujo, Rubio-Codina and Schady, 2021). Hiring schemes based on test scores have been shown to improve the quality of the hires when focusing on small-scale interventions (Hoffman, Kahn and Li, 2017). Our results, however, suggest that these policies could instead have negative effects

when scaled up and implementation constraints may be binding.

2 The Colombian Education System

2.1 Schooling in Public and Private Institutions

School enrollment rates have grown dramatically in Colombia over the past several decades. By 2010, elementary education reached near-universal enrollment, and secondary education enrollment rose from 35 percent to 77 percent in the previous two decades (Bassi, Busso and Muñoz, 2015). Although enrollment levels have increased, students' learning has seen little to no improvement, as shown by the flat evolution of student test scores. On standardized international exams, Colombian students' 75th percentile score lies well below the 25th percentile for students in member countries of the Organisation of Economic Co-operation and Development (OECD). This low performance has persisted over time and pushed Colombia to rank low among countries participating in the Programme for International Student Assessment (PISA) exams.[3] This "learning crisis" is common in much of the developing world but is more pronounced among Latin American and African countries (World Bank, 2018).

Schooling in Colombia 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, and 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. Almost 30 percent of high school students and 69 percent of post-secondary students attend private institutions. Parents and students face considerable differences in tuition and quality when choosing between a private or public school. While public schools are

[3]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.

free, private schools require a tuition payment that can vary substantially.[4]

2.2 Data: Education Administrative Records

The Colombian Ministry of Education collects administrative records from students, teachers, and schools. These records provide the three main data sources used in this paper, all of which can be linked using individual identification numbers.

High school exit exam scores.– Nearly all public and private high school students take a standardized exit exam at the end of high school. The test scores from this exam are our main outcome of interest. The administrative records include data from over 9 million students who took the high school exit exam in the second semester of each year between 2000 and 2019.[5] The exam, known as *Saber 11*, is designed and administered by the Colombian Institute for the Assessment of Education (ICFES) to assess the knowledge of senior students in various subjects, including mathematics, reading comprehension, English proficiency, social sciences, and natural sciences (i.e., physics, chemistry, and biology).[6] The exam matters to students and schools. For students, it represents an enrollment requirement at any higher education institution.[7] Exam results are also the basis of an annual ranking of all high schools, which is published by the national government and affects schools' reputations. Therefore, schools usually prepare their students using material made available by the exam authority. Students can also study independently using such preparation material or pay to enroll at private institutions that provide instruction to take the exam. We standardize subject-specific and overall scores within cohorts.[8] Besides test scores, the data also include students' demographic characteristics such as gender, age, and the

[4] In 2014, the annual tuition among private secondary schools ranged between a few hundred and 16,000 USD (Las 2 Orillas, 2014). As a reference, the minimum yearly salary for 2014 was 3,700 USD and the median was about 5,800 USD (using an exchange rate equal to 1 dollar for 2,000 COP).

[5] 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.

[6] Subject exams in history, philosophy, and geography have also been administered in some years.

[7] The college application process requires selecting a college program. Admissions are typically based on cutoffs in the overall score of the high school exit exam and, for some fields, minimum scores on subject exams.

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

household socioeconomic stratum that serves as a proxy for family income.[9]

College records.– To monitor dropout and graduation rates from higher education institutions over time, the Ministry of Education collects census-like administrative records of students through a system known as *Spadies*, available for the period between 1998 and 2016. These administrative records correspond to more than 5 million students. The information includes the year and semester when students enrolled in a college program, an indicator variable if they graduated, and the graduation date. It also records the student's percentile scores in the high school exit exam and socioeconomic information at admission.

Public teachers' administrative records.– The human resources system of the Colombian Ministry of Education collects administrative records of all public school teachers. Principals from all public schools submit teachers' information twice a year.[10] The data include unique identifiers of about 400,000 teachers between 2007 and 2015 and the exact date when each teacher was hired. This allows us to create a longitudinal data set retrospectively with information on teachers' experience and career development.

2.3 Students in Public and Private Schools

The administrative records of high school students allow us to compare the observable characteristics of students who attended public and private schools. Table 1 presents summary statistics of the students in our sample of analysis.

High school students, both from public and private schools, are, on average, 18 years old when they graduate. Slightly more than half of them 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

[9]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.

[10]This is according to Resolution 166 of the Ministry of Education. These census-like administrative data are recorded in cross-sectional files known as *Anexo 3A*.

in the country's urban zones and main cities.

2.4 Civil Service, Contract, and Private School Teachers

Teachers in Colombia are required to hold a college degree or a pedagogy diploma granted by teaching (vocational) high schools (i.e., *Escuelas Normales Superiores*).[11] A large share of teachers at public and private schools are college graduates from education majors.[12] Students enrolled in education majors are more likely to be women, more likely to come from low-income families, and more likely to have received low scores on the high school exit exam.

Private and public schools have different degrees of autonomy when hiring teachers. While private schools are entirely autonomous in hiring their staff, public school teachers are assigned based on regulations issued by the national government. Consequently, public schools have a much narrower hiring autonomy. Teaching in public schools is done by *civil service* teachers—who have a permanent contract—and by *contract* teachers—who have a fixed-term or temporary contract.

Civil service teaching positions are attractive for monetary and non-monetary reasons. Representative survey data suggest that public school teachers are more satisfied in their jobs, work fewer hours a week, and have higher salaries than private school teachers.[13] 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 only 2 percent lower than the average salary for new graduates with degrees in business and accounting.[14] These teaching positions offer employment stability, annual bonuses, health insurance,

[11] *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.

[12] 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.

[13] See Appendix Table 1.

[14] Recent college graduates employed at public schools had a monthly wage of 629 US dollars in 2010. Appendix Table 2 displays the average monthly earnings of graduates from different fields of employment in the formal sector. Panel B presents averages for all graduates, whereas panel C excludes public teachers.

and a pension system specifically designed for public school teachers.

Civil service teachers may take extended leave for various reasons, such as transfers to administrative roles, incapacity, maternity leave, unpaid leave, suspensions, or other situations. In addition, vacancies may open up because civil service teachers retire or the public system expands its coverage in certain regions, requiring more teachers to serve a larger population. If civil service teachers are not readily available to occupy these vacancies, school districts rely on temporarily hired contract teachers. Similarly to civil service teachers, contract teachers must hold a college degree or a pedagogy diploma and are paid according to the civil service pay grade. In practice, many contract teachers who were intended to work temporarily in a teaching post remained in their positions for many years. Contract teachers are, on average, younger, more likely to have a college (rather than a post-graduate) degree, and work in rural schools than civil service teachers. Both types of teachers are similarly distributed across schooling levels and subject areas.[15]

3 The Process of Hiring Public School Teachers

3.1 Teacher Hiring Before the Reform

The rules and procedures concerning the hiring and promotion of civil service teachers were originally defined in Decree 2277 of 1979. Under this regulation, each school district's hiring of new civil service teachers was decentralized to local governments. The process began with an annual assessment by local school districts to determine the number of vacancies in schools within their jurisdiction. There were no standardized criteria for how local authorities must screen candidates.[16] Both education and experience standards were required to be met to be appointed as a civil service teacher. Entry exams were seldom used but varied by region and had no clear

[15] See Appendix Table 3.

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

evaluation standards.[17] The lack of clarity in rules about civil service teacher hiring led to speculation that the allocation of vacancies was susceptible to manipulation by political interests (Bustamante, 1996; Duarte, 2001, 2003).

While teacher selection to fill vacancies was decentralized, the total number of civil service positions available to each school district and the salaries of those positions were regulated by the central government. Salary 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.[18] There are no salary differences by educational level or by subject area. Promotions in the civil service were tied to additional years of experience, completion of pedagogy courses, and graduate education attainment.[19] Public school teachers' jobs were highly stable as part of the civil service, and only severe misconduct was likely to prevent a teacher from working continuously until he or she reached retirement age. Teacher and student performance played no role in promotions or tenure.

The hiring decision of contract teachers was also decentralized to local governments (Radinger et al., 2018). The process was fairly unregulated, except for the need to comply with educational requirements and the central government's definition of salaries. This decentralized process led to a lack of clarity regarding how vacancies were filled.

3.2 The Merit-based Hiring Reform

The hiring and employment conditions of civil service school teachers were reformed by Decree 1278 of 2002, which introduced a merit-based centralized system that came into effect in 2005. The aim was to improve the quality of public education. The reform linked civil service teachers' hiring, tenure, and promotion to an evaluation

[17] The exams were often canceled due to implementation issues (Tiempo, 1996).

[18] For instance, teachers who lacked professional experience but held an education-related college degree were assigned to a career level seven when hired.

[19] 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 several pedagogy courses.

process. Under the new regulation, the government determines the number of teaching vacancies available nationwide and then announces a *public call* to fill such positions.

Applicants to the civil service teaching career must take a teaching aptitude written exam measuring (i) knowledge of the specific subject that each candidate wants to teach (e.g., math, Spanish, etc.), and (ii) a pedagogy exam. Applicants must score at least 60 (out of 100) in both components of the exam to continue in the process.[20] Only applicants performing above the minimum passing score move on to individual interviews, the next stage of the process.[21] Finally, candidates obtain a weighted average based on exam scores, interview results, and resume evaluation. The written exam, however, heavily outweighs the other components. It corresponds to 70 percent of the final score, whereas 20 percent is given to the resume (including experience) and 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. It is possible that certain vacancies remained unfilled after the public hearing.[22]

Civil service teachers hired in the post-reform period are assessed through a trial period, annual evaluations, and written exams, which are used to determine promotions. School principals must submit a report evaluating these teachers' performance after at least four months in the position.[23] In practice, civil service

[20] A psychometric test is also included along with the written knowledge and pedagogy 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.

[21] 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.

[22] The process can finish either because no more vacancies are available or because all eligible teachers have already been matched to a vacancy.

[23] Principals collect information on a teacher's performance in academic aspects (such as knowledge of the teaching subject, class planning, pedagogy strategies, and evaluation methods), school

teachers rarely fail their trial period evaluation (Garcia et al., 2014; Forero and Saavedra, 2019). After the trial period, teachers in the civil service undergo annual evaluations, which are conducted by their principals. Teachers' employment status depends on not failing two consecutive evaluations. De facto, this mechanism is ineffective at firing low-performing teachers since annual evaluations are assumed to be a means to provide feedback rather than a system to monitor performance.[24] Finally, the reform tied promotions to performance on a written exam evaluating teachers' knowledge.[25]

Salaries continued to be set by the central government after the reform; which also increased the salaries of civil service teachers. 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.[26] Such a wage increase was intended to attract a higher-quality pool of teacher candidates to fill the vacancies at public schools.

The civil service hiring reform did not directly change the way contract teachers were hired. Local governments could still rely on contract teachers to fill up vacancies. Contract teachers could become part of the civil service by going through the new hiring system (which valued performance in the teaching aptitude test over accumulated experience). Teachers hired before 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).

---

administrative duties, and the teacher's involvement with students' families and the environment. Teachers must obtain a minimum score of 60 out of 100.

[24]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.

[25]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 being hired or receiving their last promotion) and obtain a score above 80 out of 100 to be promoted.

[26]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.

Conducting national public calls requires significant time and financial resources due to logistics and exam preparation. Logistically, a sufficient number of vacancies is necessary to justify a national public call. Additionally, collecting vacancy information from each education department is time-consuming. Six public calls to fill vacancies nationwide have occurred since the reform was initially passed (see Figure 1a). The first call was conducted in 2004 when the government announced that 44,596 teachers were needed. Appendix Figure 2 shows that it took more than a year for these teachers to be hired into the civil service. 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.[27] More recently, in 2021, the government announced a new public call to fill approximately 29,000 teaching vacancies.[28]

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).

3.3 Reform Implementation

The implementation of the merit-based reform inadvertently shocked teacher turnover. Novice civil service teachers replaced not only those who retired or left the profession (i.e., *expected* turnover) but also contract teachers who had already accumulated several years of teaching experience (i.e., *unexpected* turnover).[29] Three implementation issues seem to explain this fact.

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

[28] 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 few 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 areas that have suffered the consequences of the armed conflict in the country.

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

Hiring freeze before the reform–. Only a few civil service teachers were hired during the decade before the reform. This led to an increase in the number of contract teachers working in public schools.[30] Because the total number of teachers in public schools remained relatively constant in the period under analysis, vacated positions were largely filled with contract teachers. By 2004, just before the implementation of the reform, there were 55,000 contract teachers, representing 17 percent of all teachers (Jerez, 2004). At that time, many of these contract teachers had more than five years of experience working in public schools.

Insufficient number of public calls.– Three public calls occurred between 2004 and 2008. Approximately 80,000 new civil service teachers were hired from these three calls.[31] By 2007, nearly 40,000 (or 70 percent out of the 55,000) of the contract teachers active in 2004 had been replaced by novice civil service teachers (Jerez, 2004).[32,33] This resulted in a large turnover of the stock of teachers post-reform and an increase of 20 percentage points in the fraction of teachers with less than five years of experience.[34] Consequently, after the first three public calls, in nearly 13 percent of classrooms, novice civil service teachers with less experience replaced more experienced contract teachers.

This pattern in teacher turnover continued after the first public calls. Contract teachers were replaced after every public call, and public schools continued 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 contract

[30] 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 contract teachers were hired.

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

[32] Appendix Figure 4 shows the drop in the number of contract teachers hired between 2004 (pre-reform) and 2007 (post-reform).

[33] 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 who left can be computed as the number of contract teachers who were recorded as working in 2004 but had left by 2007 (40,000) plus the number of civil service teachers who retired (32,000). Appendix Figure 5a shows that approximately 8,000 civil service teachers retire yearly. This would mean that about 32,000 civil service teachers would be expected to have retired over this four-year period. The total number of teachers who left the system was around 72,000, combining civil service and contract teachers. This number is close to the number of vacancies posted by 2007 in the first three public calls in Figure 1a.

[34] See Figure 4b.

teachers. These teachers were considerably younger than civil service teachers, but had already accumulated multiple years of teaching experience.[35]

The new merit-based system could have been designed to implement calls for teachers at a higher frequency to fill vacancies with civil service teachers continuously. Instead, the system initially filled vacancies with contract instructors, who were replaced with civil service teachers in later public calls. The number of contract teachers increased between public calls and dropped when the next wave of new novice teachers was hired.[36] 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.

Inability of the new system to fill all positions–. Even though the public calls have been oversubscribed, some positions could not be filled. Each call has attracted more than double the number of applicants than vacancies, suggesting that civil service teaching positions are attractive and competitive (see Figure 1b). Applicants ranked their preferred schools or positions among those available. This means that some positions might have many interested applicants while others few (or none at all).[37] Many of these positions were filled with contract teachers. Contract 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).[38]

[35]See Appendix Figure 5a for the share of teachers who leave the profession and are hired using a fixed term contract, Appendix Figure 5b for the age distribution, and Appendix Figure 5c for their accumulated years of experience.

[36]See Appendix Figure 4.

[37]Appendix Table 4 shows that applicants who performed one standard deviation better than average were employed in schools where students, on average, performed 0.04 standard deviations better on the high school exit exam. These applicants were also 1 percentage point more likely to be employed in schools closer to the municipality where they took the teaching aptitude exam. The proportion of the total variance explained by these models is small, suggesting a limited degree of teacher sorting.

[38]See also Appendix Table 3.

4 Effects on Teachers' Skills and Experience

The merit-based hiring system increased civil service teachers' average skills. We merge the college records with the public teachers' administrative records and use the percentile score of teachers in the high school exit exam as a proxy of cognitive skills for teachers hired before and after 2005 (i.e., the first year the Colombian government implemented the hiring reform). 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 teachers hired into the civil service. 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 was observed in 2005. Teachers hired after 2004 had test scores of 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 civil service teachers between 2002 and 2015, separating those who were hired before and after the reform. Because we only observe teachers who worked between 2007 and 2015, we calculate the mean and inter-quartile range using the dates when teachers were hired. 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 teaching experience of public school teachers. Figure 4a illustrates how the level of teaching experience at public schools changed. 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 just under 10 percent in 2002 to 30 percent by 2008.[39] As the first wave of new teachers gained experience,

[39]The measures of experience are computed retrospectively. If turnover is higher during the first years of the teaching career, this could potentially lead us to underestimate the share of novice teachers in Figure 4b and distort the experience profile shown in Figure 4a. We do not think this is a big issue in our context for two reasons. First, if the likelihood of dropping from the sample increases with years of tenure, we should observe an even lower share of novice teachers before 2002. However, we find the opposite pattern: the share of novice teachers is high in the late 90s. Second, the estimated share of novice teachers in the 2007-2014 period computed retrospectively using data only from 2015 (see

the fraction of inexperienced teachers in Colombia fell to 20 percent where it remained fairly stable over the later part of our sample period.

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 teachers working in public schools. 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.[40] In addition, while the reform increased teacher turnover, there was little change in the number of teachers in Colombia during this time period.

The merit-based reform affected fundamental inputs for students' learning, such as general teachers' skills (captured by test score measures) and teachers' task-specific human capital (accumulated through on-the-job 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

5.1 Empirical Strategy

Our empirical strategy identifies the effects on student outcomes of the new merit-based teacher-hiring system. We focus on two main set of outcomes: high school exit exam performance and college outcomes. To obtain the latter, we merge

Appendix Figure 6) is very similar to the share of novice teachers computed in the same period using actual data. This implies that teacher turnover might not be enough to underestimate our results.

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

college records with high school exit exam records 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.[41] 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.[42] Formally, we estimate:

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

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 school heterogeneity 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 to 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. This allows for error terms of different students who belonged to the same cohort in the same schools (and were therefore subjected to similar inputs and unobserved shocks) to be arbitrarily correlated.

[41] 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 dates 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.

[42] This strategy uses observations cohort by cohort, and does not follow the same individuals over time.

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 with teaching 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]

5.2 Validity of the Research Design

Our research design requires that 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. The estimation 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 cannot be directly observed, the effects estimated in our event study strategy allow us to test for parallel trends prior to the reform. We provide evidence consistent with the validity of this assumption in the figures presented in Section 5.3.

Consistency also requires that private schools (i.e., the control group) were unaffected by the reform. In particular, the reform did not induce movements of

[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.

teachers or students between public and private schools. While ultimately, it is not possible to directly test this assumption, we consider it reasonable in our specific context as we discuss next.

Stability of teachers.– Figure 5 shows the evolution of the number of teachers and the proportion of female teachers normalized to the year previous to the reform.[44] We do not observe any significant changes in trends after the reform. There was not a massive influx of teachers into public/private schools. Moreover, job transitions between public and private schools were uncommon after the reform.[45] The probability of a public school teacher moving to a private school in the following 3 years was less than 1.5 percent, and for a private school teacher moving to a public school was less than 2 percent.[46] Meanwhile, about 90 percent of public school teachers and more than 60-80 percent of private school teachers remained in their positions, with the remaining exiting the teaching profession altogether. If established (older) private school teachers had switched to the public system as a result of the reform, we should observe an increase in the average age at which teachers enter the public sector. However, this is not the case.[47] Finally, Saavedra et al. (2022) found that only 8 percent of individuals who failed the test to become a public school teacher remained in an education-related occupation the following year. Most switched to informal jobs or found formal employment in the service sector. This suggests that contract teachers who leave their positions are more likely to exit the teaching profession than to switch to private schools.

Stability of students.– Figure 6 presents the evolution of average students' characteristics across public and private schools over time.[48] We do not see any drastic

[44]Calculations in panels 5a and 5b were performed using the census of schools gathered under the form C-600 of the Ministry of Education. The data are available from 2002 onwards, except for 2003.

[45]See Appendix Table 5.

[46]Unfortunately, we lack information to compute these transition probabilities for the period before or during the initial reform.

[47]See Appendix Table 6.

[48]Statistics shown in panel 6a are based on the census of schools gathered under the form C-600 of the Ministry of Education. The data are available from 2002 onwards, except for 2003. Statistics shown in panels 6b, 6c, and 6d were computed from the Colombian household survey. To ensure comparability in the survey, we focus on students between the age of 5 and 20 who live in the 13 main metropolitan areas of Colombia.

changes in the composition of students across both types of schools. The number of students per school evolved similarly before and after the reform, as well as the share of female students, students' household income, and their age.[49] In addition, as we show in Section 5.3, students' test scores in private schools remained fairly constant in the period under analysis. This evidence is consistent with the assumption of stability of the control group in the sample of students.

Moving from a public to a private school is probably uncommon because it is costly. Public schools are free. Tuition fees in private education in Colombia are high and depend on an evaluation performed by the Ministry of Education. On average, monthly tuition fees for private education in Colombia in 2018 were equivalent to $ 584 USD –the equivalent of 2.3 minimum wages at that time.[50]

5.3 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 7 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 normalized gap in test scores between public and private students is close to zero and stable during the pre-treatment period (from 2000 to 2004). Consequently, test

[49]The average number of students in private schools seems to have a steeper gradient relative to public schools, but we cannot reject that these are statistically different from each other.

[50]Data is retrieved from the ministry of education at https://www.mineducacion.gov.co/portal/micrositios-preescolar-basica-y-media/Educacion-Privada/Tarifas-y-Costos-Educativos/219212:Matriculas-y-Pensiones. The average tuition fee was 1,809,690 COP, which represented $ 584 USD. The minimum salary for 2018 was 781,242 COP.

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.[51]

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. The magnitude of the effect is equivalent to the negative impact of being taught by a first-year teacher (Staiger and Rockoff, 2010), to a one standard deviation decrease in teacher quality (Chetty, Friedman and Rockoff, 2014; Petek and Pope, 2023), and about half of the negative effect of switching teaching pedagogy from a traditional lecture style to a new student-centered approach (Berlinski and Busso,

[51] Appendix Figure 8 shows test scores in private and public schools for the period from 2000 to 2013 – a period in which exams where the most comparable. Scores are standardized to the distribution of tests scores in 2000. We observed a spike of more than one standard deviation in year 2010 in both types of schools – which we think reflects a change in the exam in that year and thus we do not report. After 2013 comparisons across years are more difficult to do because of changes to the exam made by ICFES. There is an immediate increase in test scores both in public and private schools (which we think highlights the limited comparability of tests across years). Overall, test scores at private schools remained relatively constant –or slightly increased– while scores in public schools seem to have declined.

2017). Our 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).[52]

Effect on college outcomes – Figure 8 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. While there appears to be some evidence of a pre-trend prior to the policy reform, 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 begins to converge back towards 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 8. 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 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 (OECD and World Bank, 2012). 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.

[52] In Appendix Figure 9 we present dynamic estimates of the effect by subject-specific exam 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.[53]

5.4 Teaching Experience and Student Outcomes

The reform negatively affected students' learning, even though it led school districts to hire teachers who ostensibly had higher cognitive skills (as measured by their own high school exit exam scores). This result is likely driven by students' increased exposure 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 experience (see Figures 4b and 7). 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.[54] 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,

[53] In Appendix Table 7 we present complementary results for these estimations in a constant sample of individuals across outcomes.

[54] 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.

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 8a. 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.[55] 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 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

[55]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.

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 2007. 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 10.

This differential dosage effect is further confirm when analyzing the effect of differential exposure to novice teachers by subject (i.e., math, Spanish, English, social and natural sciences).[56] We estimate the same model at the student-by-subject level and interact the indicator variable of public school with indicator variables for the quintiles of the share of novice teachers *per subject*. Panel A of Appendix Table 8 presents the results. It shows that a higher exposure to novice teachers in a given subject negatively impacts the high school exit exam test scores in that same subject.[57]

School vacancies that were more demanded by applicants are expected to face larger teacher replacement rates relative to less demanded schools where contract teachers were left in place. Following the argument in Appendix Table 4, we proxy high demand with whether the public school is urban and with an indicator variable for whether the school is in the top quintile of the high school exit exam scores distribution prior to the reform. We present the results in panels B and C of Appendix Table 8. For the fully saturated model, we find statistically larger negative impacts of the reform in urban schools and in schools with higher pre-reform test scores.

[56]The share of novice teachers by subject remains very stable in time. The average proportion of novice teachers who teach English, math, natural sciences, social sciences and Spanish are 15.77%, 24.19%, 24.49%, 19.07% and 16.48% respectively.

[57]We additionally present dynamic effects for the first and fifth quintiles in Appendix Figure 11.

Altogether, these 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 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. The debate around teacher characteristics predicting future teacher quality has shown opposing views with some arguing that characteristics like credentials predict teacher success (Rockoff et al., 2011; Clotfelter, Ladd and Vigdor, 2010; Kane, Rockoff and Staiger, 2008), whereas others are more pessimistic highlighting the limited scope on student achievement (Rivkin, Hanushek and Kain, 2005; Hanushek and Rivkin, 2006; Jackson, Rockoff and Staiger, 2014; Araujo et al., 2016; Hanushek, 2011; Taylor and Tyler, 2012).[58]

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, policymakers 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 civil service positions 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 contract teachers and then refilled with civil service teachers from the

[58] 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.

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 contract teacher to fill the position for one to five years and then a civil service 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.[59]

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 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,

[59]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 7).

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

Al-Ubaydli, Omar, John A. List, and Dana L. Suskind. 2017. "What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results." *American Economic Review*, 107(5): 282–86.

Al-Ubaydli, Omar, Min Sok Lee, John A List, and Dana Suskind. 2021. "The science of using science: A new framework for understanding the threats to scaling evidence-based policies." In *The Scale-Up Effect in Early Childhood and Public Policy*. 104–125. Routledge.

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, Marta Rubio-Codina, and Norbert Schady. 2021. "70 to 700 to 70,000: Lessons from the Jamaica Experiment." In *The Scale-Up Effect in Early Childhood and Public Policy*. 211–232. Routledge.

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.

Berlinski, Samuel, and Matias Busso. 2017. "Challenges in educational reform: An experiment on active learning in mathematics." *Economics Letters*, 156: 172–175.

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.

Bold, Tessa, Deon Filmer, Ezequiel Molina, and Jakob Svensson. 2019. "The lost human capital: Teacher knowledge and student achievement in Africa." *World Bank Policy Research Working Paper*, 8849.

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Christophe Rockmore, Brian Stacy, Jakob Svensson, and Waly Wane. 2017. "What do teachers know and do? Does it matter? Evidence from primary schools in Africa." *Does it Matter*.

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.

Clotfelter, Charles T, Helen F Ladd, and Jacob L Vigdor. 2010. "Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects." *Journal of Human Resources*, 45(3): 655–681.

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.

Hoffman, Mitchell, Lisa B Kahn, and Danielle Li. 2017. "Discretion in Hiring." *The Quarterly Journal of Economics*, 133(2): 765–800.

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.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. "What does certification tell us about teacher effectiveness? Evidence from New York City." *Economics of Education Review*, 27(6): 615–631.

Khanna, Gaurav. 2023. "Large-scale education reform in general equilibrium: Regression discontinuity evidence from india." *Journal of Political Economy*, 131(2): 549–591.

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

Metzler, Johannes, and Ludger Woessmann. 2012. "The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation." *Journal of development economics*, 99(2): 486–496.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2015. "The aggregate effect of school choice: Evidence from a two-stage experiment in India." *The Quarterly Journal of Economics*, 130(3): 1011–1066.

OECD and World Bank. 2012. *Reviews of National Policies for Education: Tertiary Education in Colombia.* OECD.

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.

Ost, Ben. 2014. "How Do Teachers Improve? The Relative Importance of Specific and General Human Capital." *American Economic Journal: Applied Economics*, 6(2): 127–51.

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

Radinger, Thomas, Alfonso Echazarra, Gabriela Guerrero, and Juan Pablo Valenzuela. 2018. *OECD Reviews of School Resources: Colombia 2018.* OECD Publishing.

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.

Saavedra, Juan Esteban, Dario Maldonado, Lucrecia Santibañez, and Luis Omar Herrera-Prada. 2022. "Premium or Penalty? Labor Market Returns to Novice Public Sector Teachers." *Journal of Human Resources*.

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

Taylor, Eric S., and John H. Tyler. 2012. "The Effect of Evaluation on Teacher Performance." *American Economic Review*, 102(7): 3628–51.

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

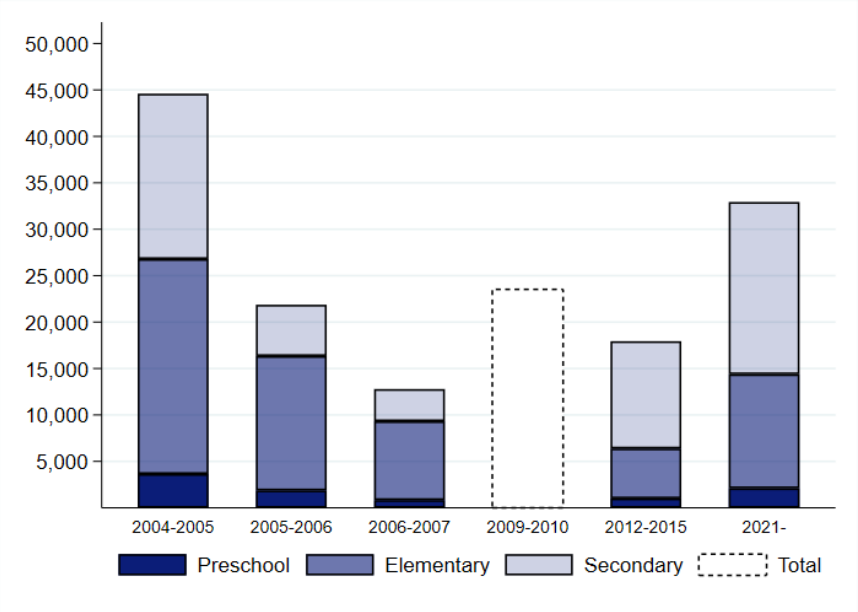
Velasquez, Maria, Oscar Arcos, Adriana Rodriguez, and Fabian Gonzalez. 2010. *Docentes y Directivos Docentes. Acerca de su Condicion y Perfiles sobre Resultados de Convocatorias de Concursos de Meritos.* CID, Universidad Nacional de Colombia.

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

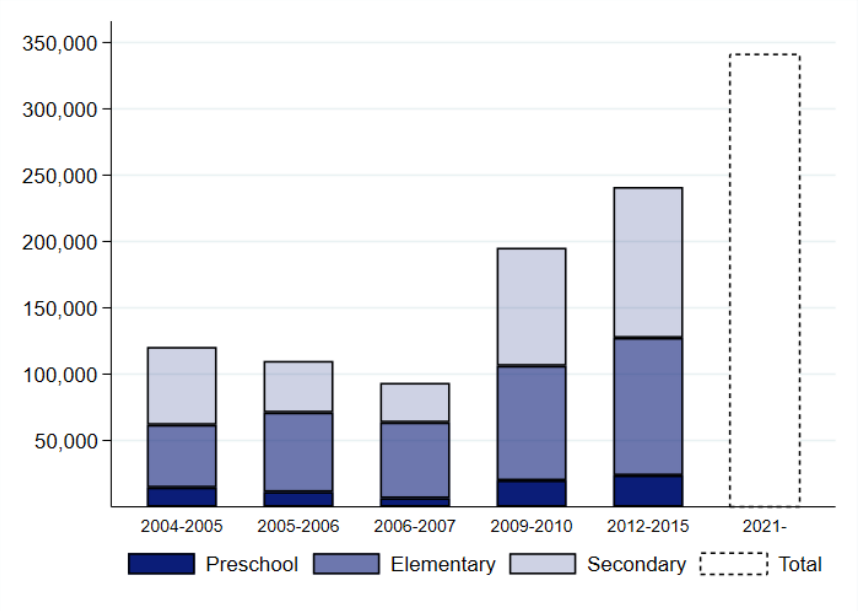
Figures and Tables

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

(a) Teaching Vacancies

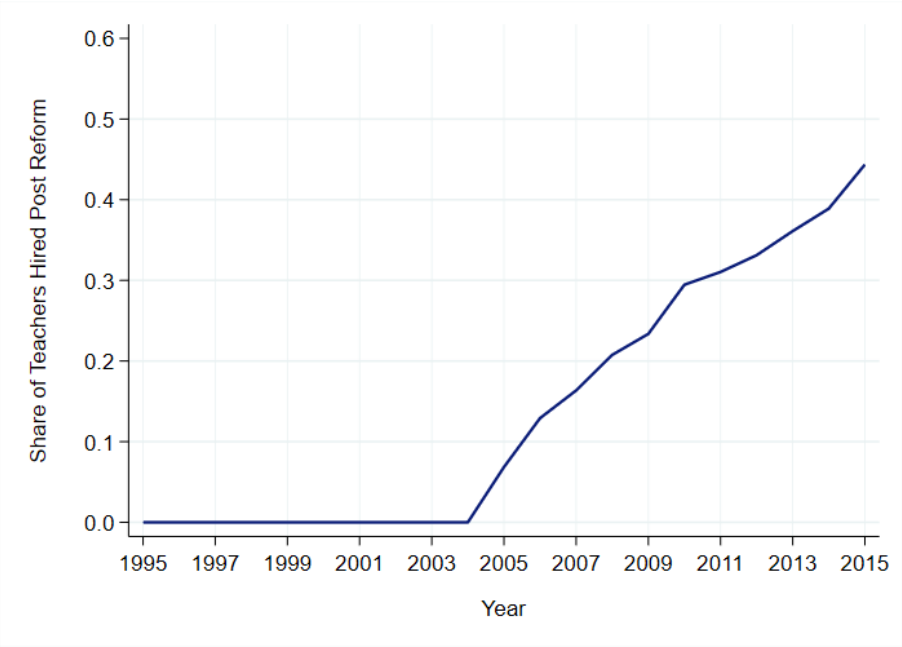


(b) Applicants to Teaching Positions



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.

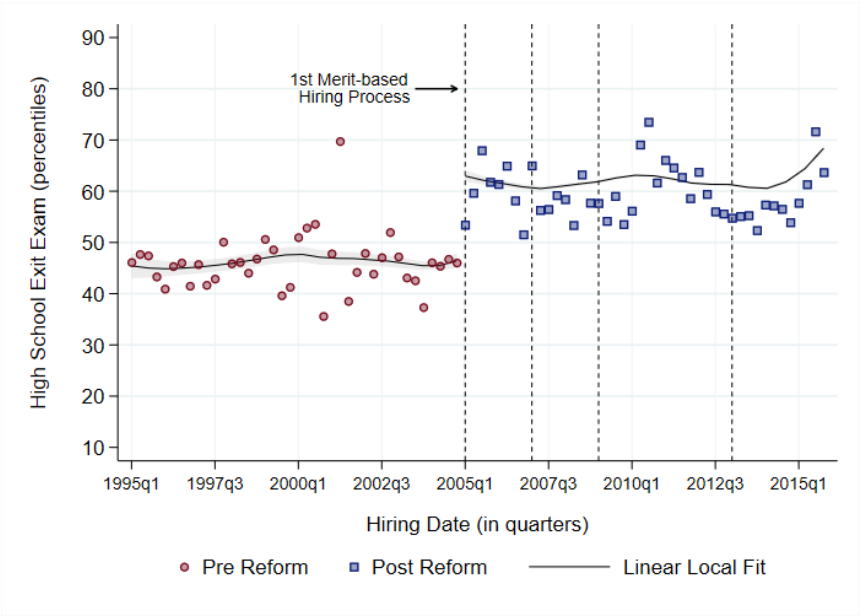
Figure 2: Share of Teachers Hired After the Reform's Implementation



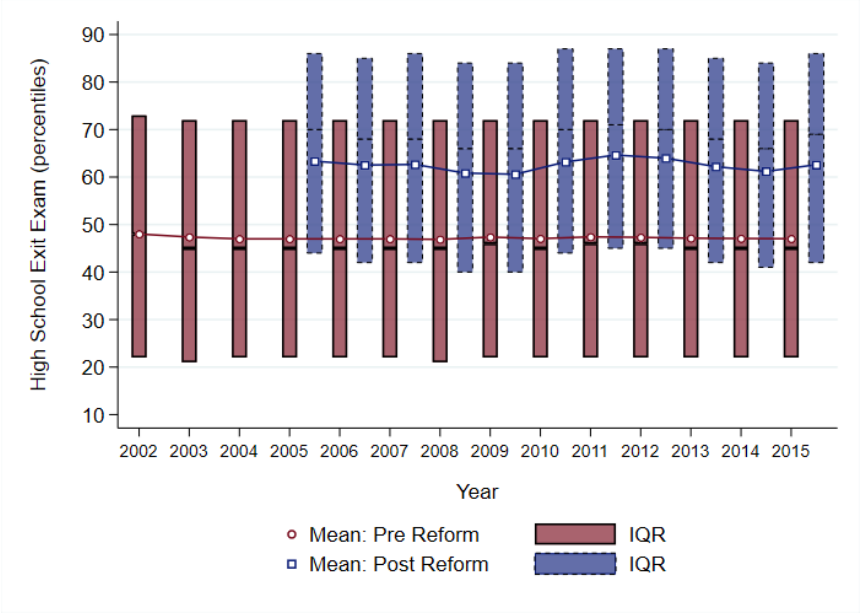
Notes. The solid line represents the annual share of teachers hired after the reform was implemented in 2005.

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

(a) Test Scores by Teacher's Hiring Date



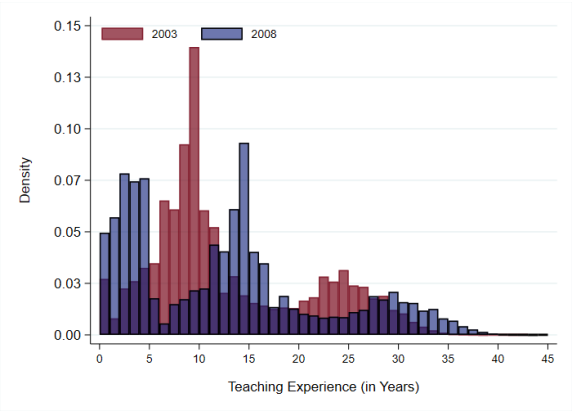
(b) Distribution Pre- and Post- Merit System



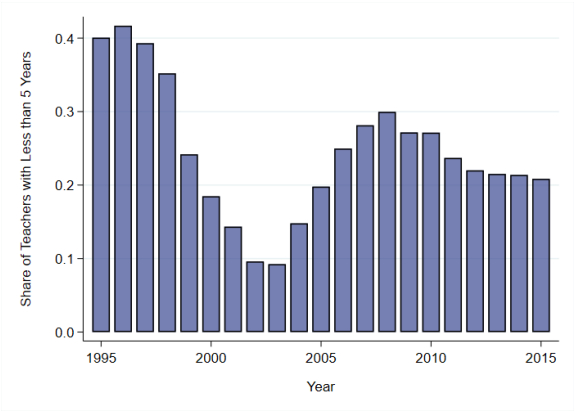
Notes. Panel 3a plots the average percentile in the high school exit exam of teachers hired in each quarter between 1995 and 2015. Solid lines represent local linear regressions fitted using individual-level data of teachers hired before and after 2005. Confidence intervals at the 95% level are displayed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



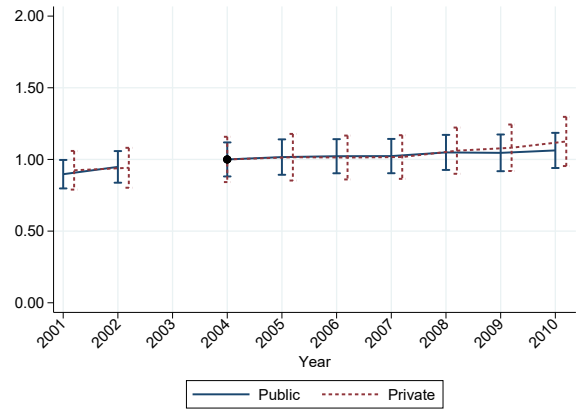
(b) Share of Novice Teachers



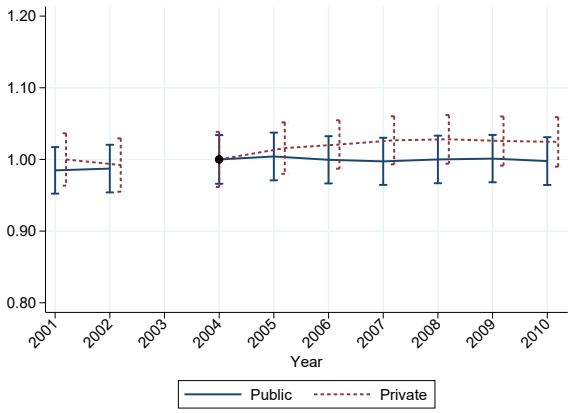
Notes. Novice teachers are defined as those with less than five years of experience. For years prior to 2007, teaching experience for a given year is computed retrospectively using the difference between that year and the year in which the teacher started, but observed between 2007 and 2015. Teaching experience after 2007 is directly observed. Panel 4a shows the density of experience among teachers working in 2003 and 2008. Panel 4b plots the share of novice teachers working at public schools in any given year between 1995 and 2015.

Figure 5: Teacher Characteristics Across Public and Private Schools

(a) Number of Teachers per School



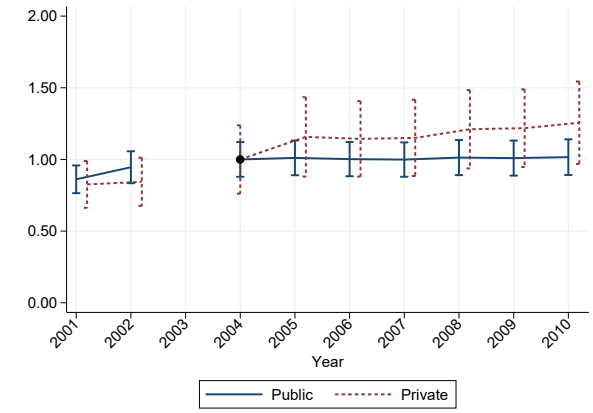
(b) Share of Female Teachers



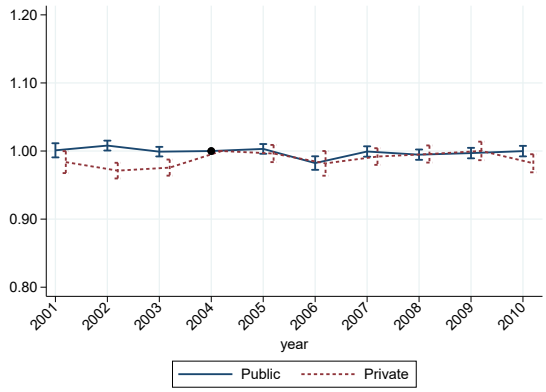
Notes. These figures plot the evolution of teachers' characteristics, separately for those in the public and private sector. The values are normalized to 2004, the last year pre-reform. Panels 5a and 5b are computed using the census of Colombian schools (form C-600), which does not have information for the year 2003. Confidence intervals correspond to the 95 percent confidence level.

Figure 6: Student Characteristics Across Public and Private Schools

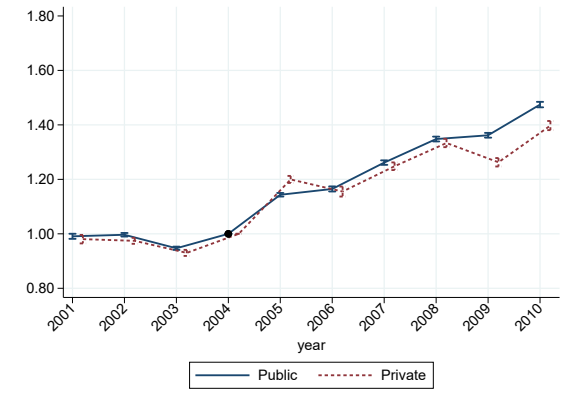
(a) Number of Students per School



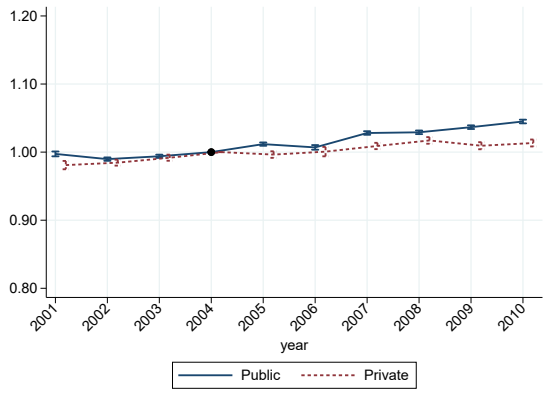
(b) Share Female Students



(c) Students' Household Income

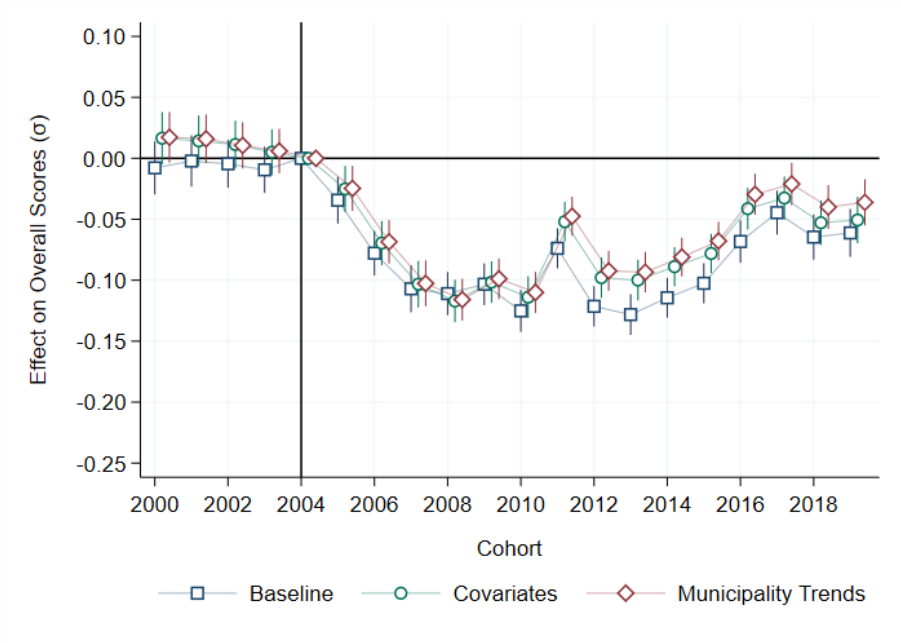


(d) Students' Age



Notes. These figures plot the evolution of students' characteristics, separately for those in the public and private sector. The values are normalized to 2004, the last year pre-reform. Panel 6a is computed using the census of Colombian schools (form C-600), which does not have information for the year 2003. Panels 6b, 6c, and 6d are computed using the Colombian household survey (Gran Encuesta Integrada de Hogares). To ensure comparability across years, we limit the sample to students between the ages of 5 and 20 who live in the main 13 main metropolitan areas. Household income is winsorized at the 1st and 99th percentile. Confidence intervals correspond to the 95 percent confidence level.

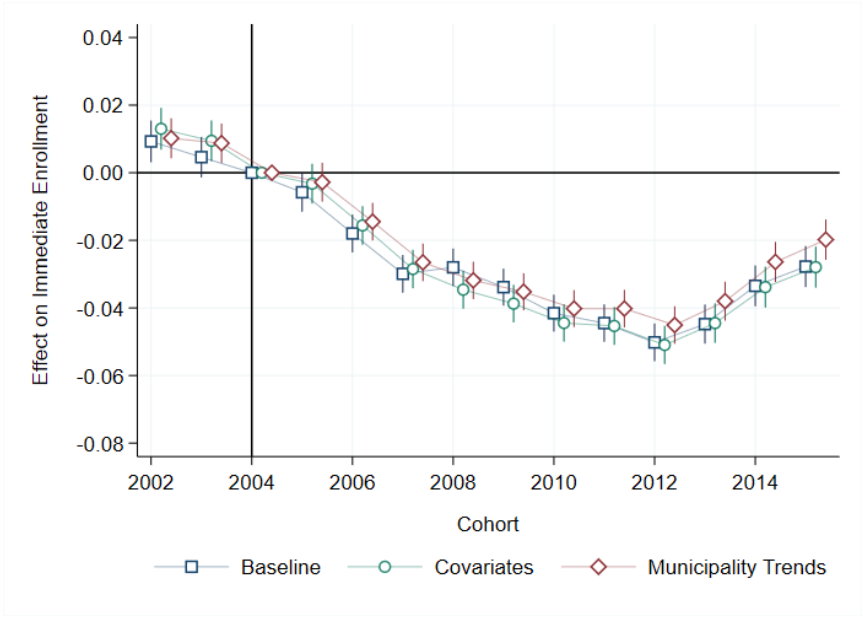
Figure 7: Dynamic Effects of a Merit-based Teacher-hiring Policy on Students' Test Scores



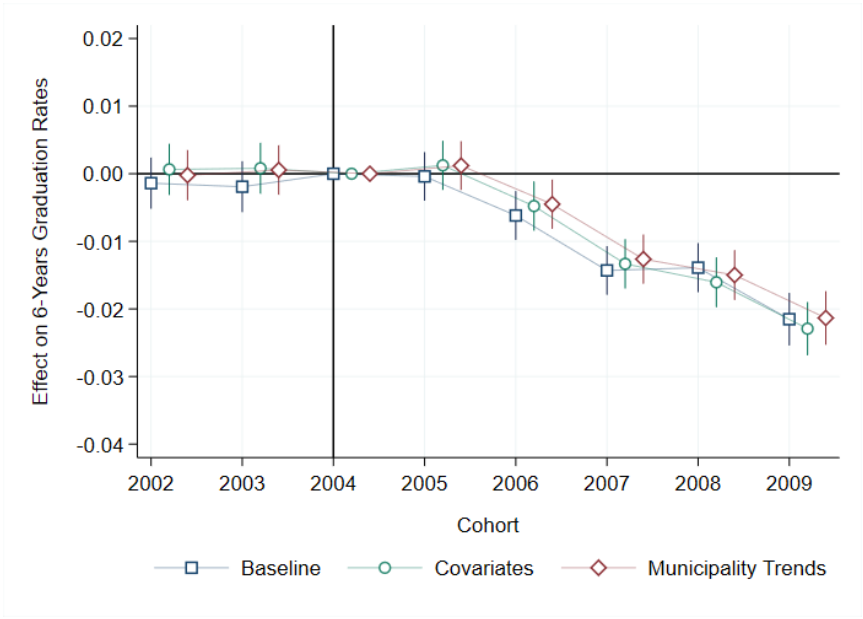
Notes. Ordinary Least Squares estimates of the dynamic effects, δ_τ, of equation 1. The estimation uses observations cohort by cohort, and does not follow the same individuals over time. The outcome variable is the overall score on the high school exit exam. Overall scores are computed as the average 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 8: 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 estimation uses observations cohort by cohort, and does not follow the same individuals over time. The outcome variable in Panel 8a indicates whether a student enrolls in a college program within six months after graduating high school. In Panel 8b, 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 (Pooled)

	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 Nights	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***	–0.032***	–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]
Mean of Dependent Variable for Public Students	0.115	0.115	0.115	0.265	0.265	0.265	0.075	0.075	0.075
R-squared	0.147	0.155	0.157	0.207	0.228	0.229	0.083	0.087	0.088
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
Municipality Trends		Yes	Yes		Yes	Yes		Yes	Yes
School 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	0.332	0.346	0.349	0.256	0.289	0.292	0.212	0.230	0.232
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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

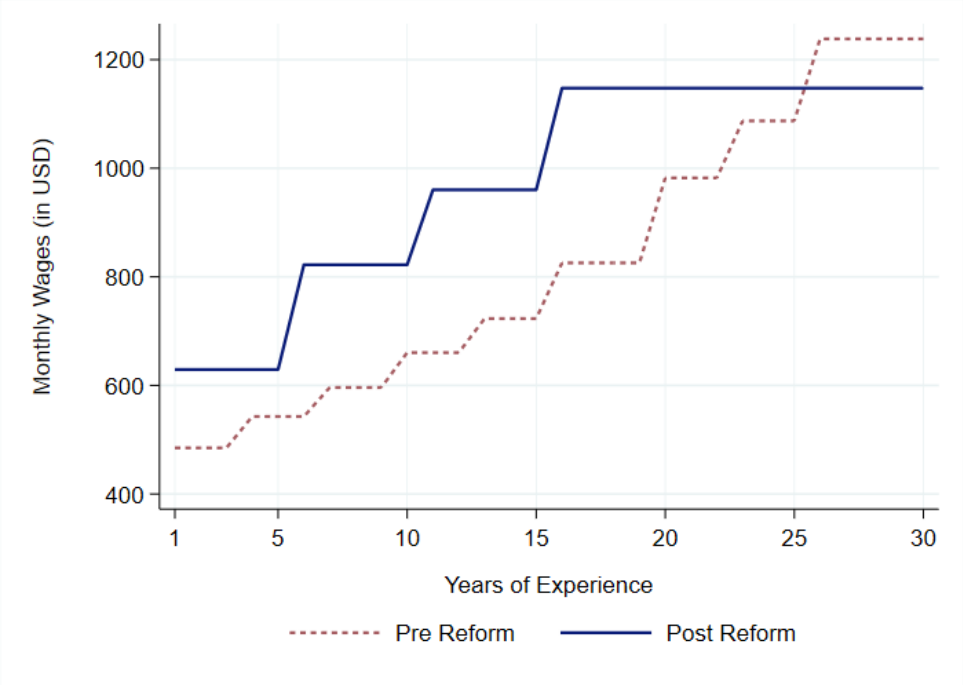
Table 5: Heterogeneous Effects of a Merit-based Hiring Policy on Students' College Outcomes

	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]
Mean of Dependent Variable for Public Students	0.114	0.114	0.114	0.266	0.266	0.266	0.075	0.075	0.075
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
Municipality Trends		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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

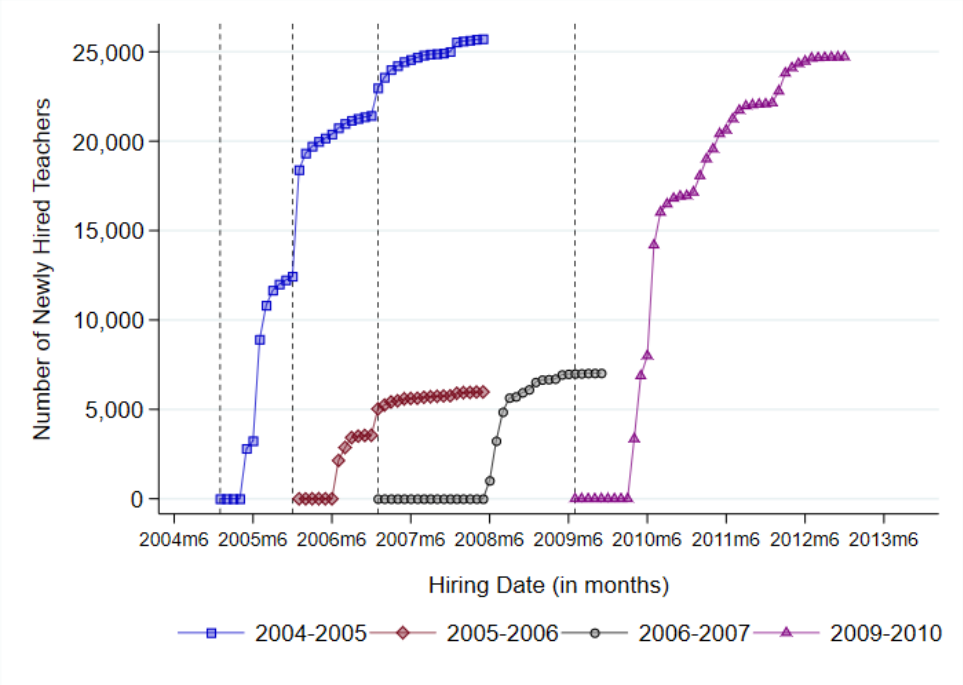
A Additional Figures and Tables

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



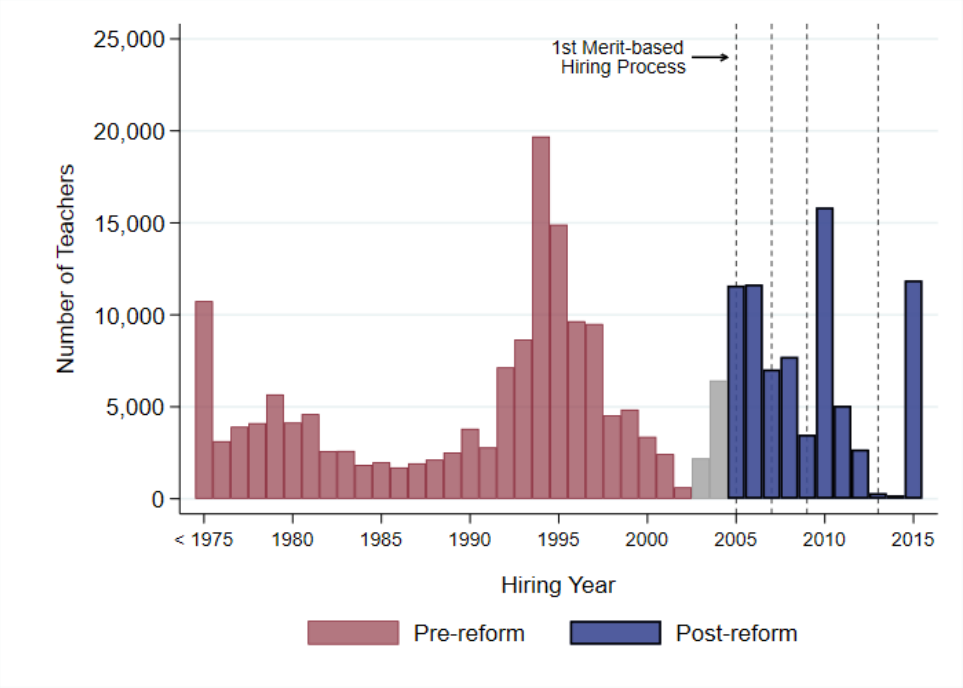
Notes. Wage-experience profiles are computed based on public school teachers' salaries in 2010, found in Decrees 1367 and 1369. The profile of teachers hired post-reform assumes promotions every five years. The daily average of the exchange rate in 2010, 1 $USD = 1898.7 $COP, is used to present salaries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

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



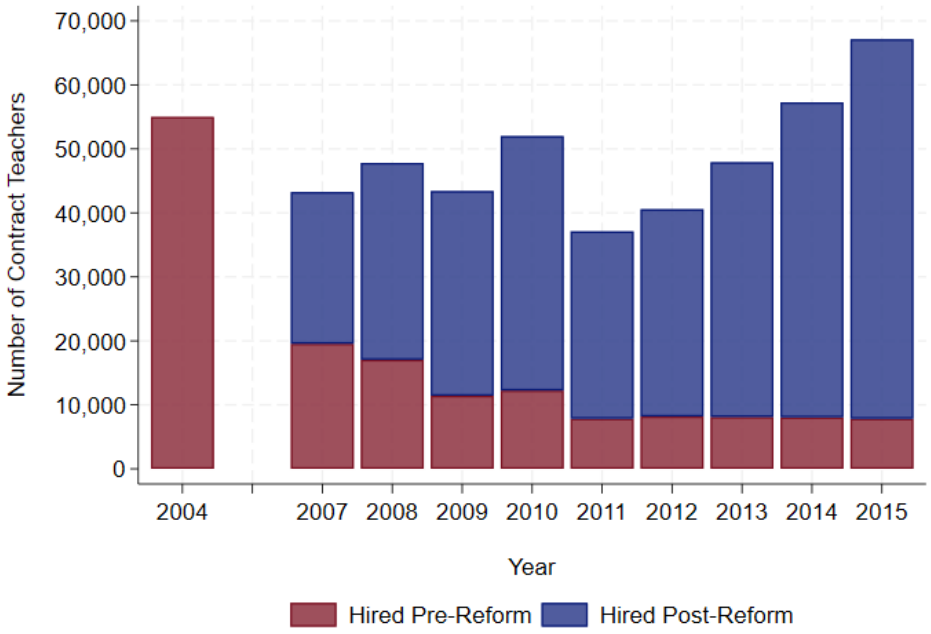
Notes. This figure plots the (cumulative) number of teachers hired across time in each merit-based hiring process between 2004 and 2014. Vertical dashed lines represent the month when individuals hired took the entry exam used by the Colombian government to screen applicants.

Appendix Figure 3: Stock of Civil Service Teachers by Hiring Date



Notes. This figure plots the density of Civil Service 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 were contract teachers initially but transitioned to civil service teachers by 2015.

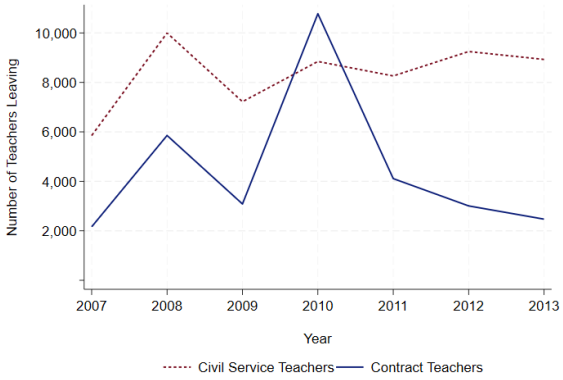
Appendix Figure 4: Stock of Contract Teachers Hired Before and After the Reform



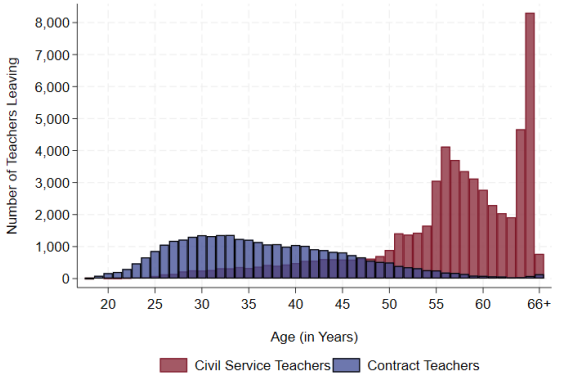
Notes. This figure displays the number of contract teachers working at public schools in any given year between 2007 and 2016. The data from 2007 to 2015 correspond to the census of teachers. The data point for 2004 comes from Jerez (2004).

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

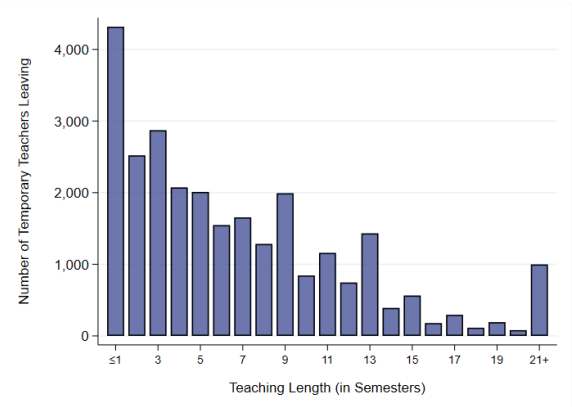
(a) Number of Teachers Leaving Across Time



(b) Distribution of Teachers by Age

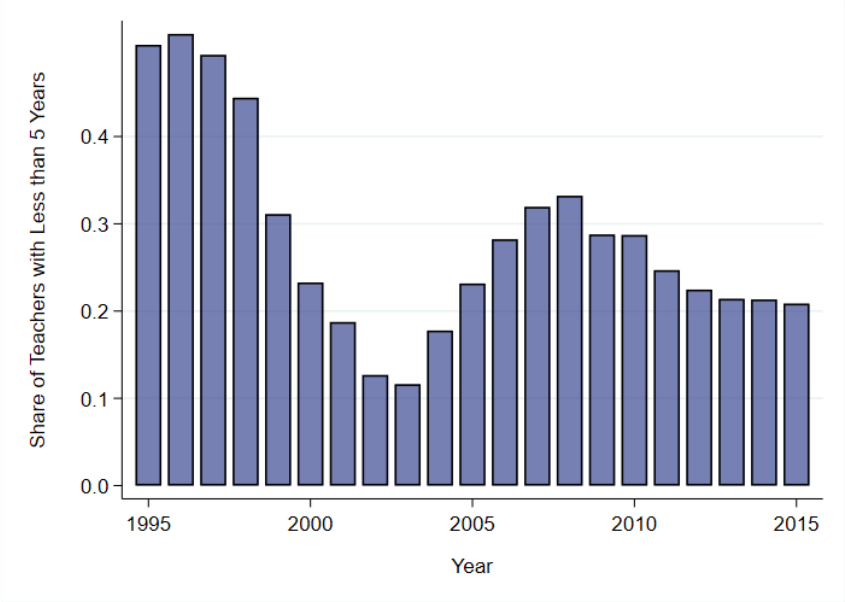


(c) Contract Teachers' Work Experience at Moment of Leaving



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

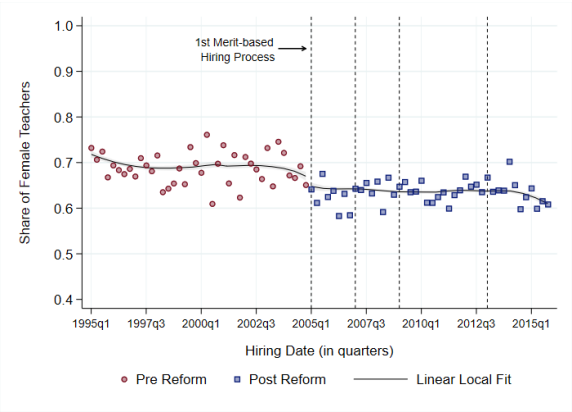
Appendix Figure 6: Share of Novice Teachers (Using only the census in 2015)



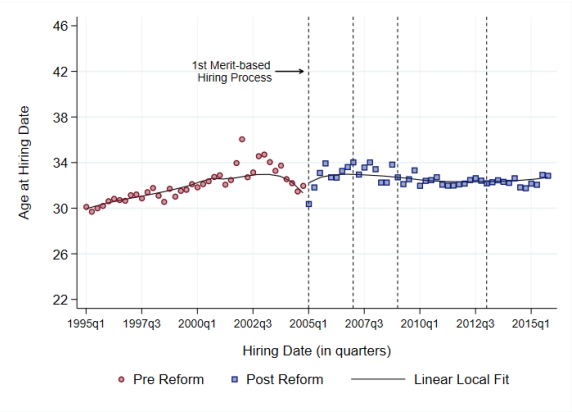
Notes. Novice teachers are defined as those with less than five years of experience. Teaching experience for a given year is computed retrospectively using the difference between that year and the year in which the teacher started, but observed in 2015, exclusively.

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

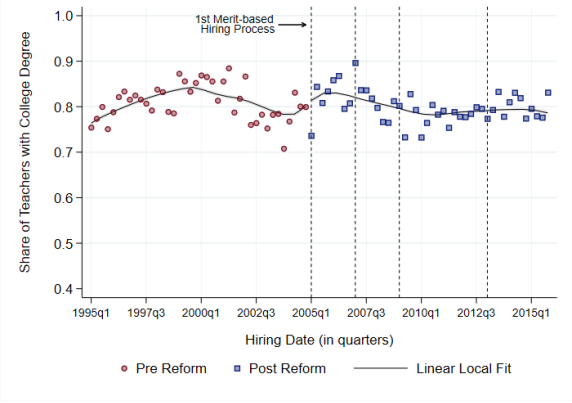
(a) Gender



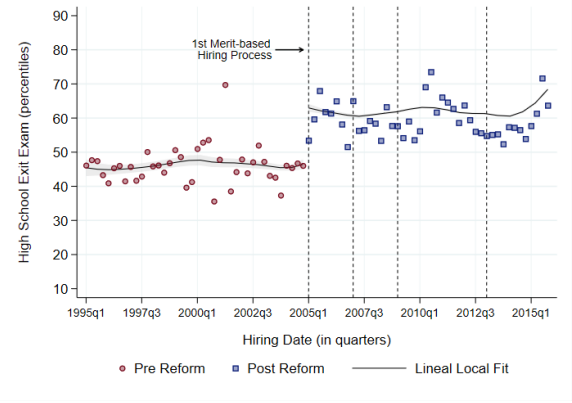
(b) Age at Hiring Date



(c) College Degree

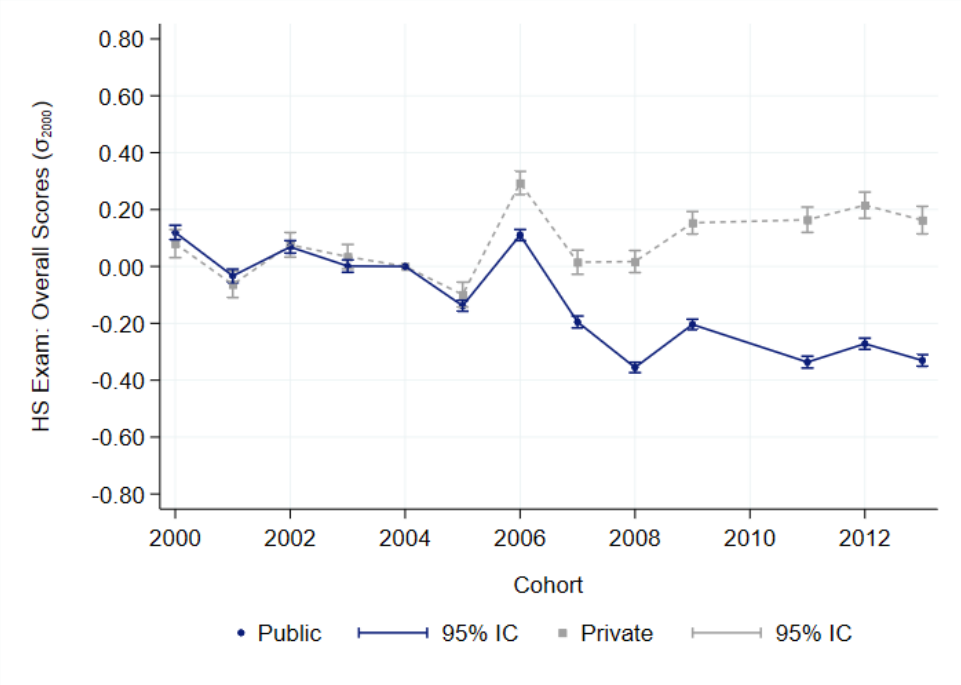


(d) Test Scores



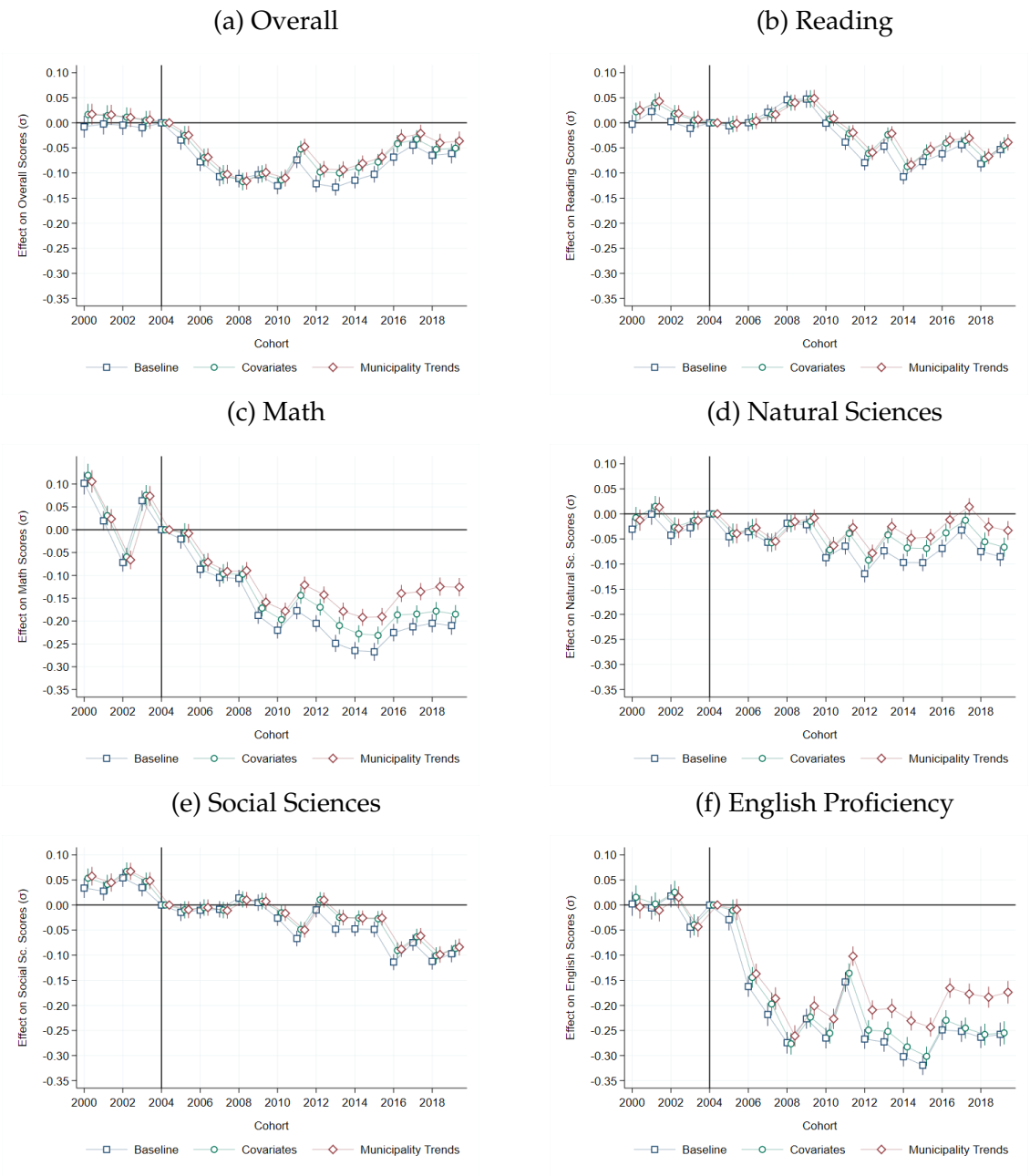
Notes. Plotted dots in Panel 7a represent the share of female teachers hired in the same quarter between 1995 and 2015. Panel 7b plots the average age (at hiring) of teachers hired in the same quarter. Panel 7c plots the share of teachers holding a college degree. Panel 7d plots the average percentile 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: 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. The point estimate for 2010 is omitted because we observed a spike of more than one standard deviation in both types of schools which might reflect a change in the exam in that year and thus we do not report it. 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.

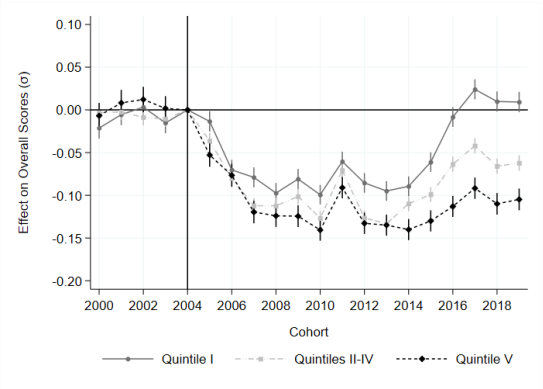
Appendix Figure 9: Dynamic Effects of a Merit-Based Teacher Hiring Policy on Students' Test Scores



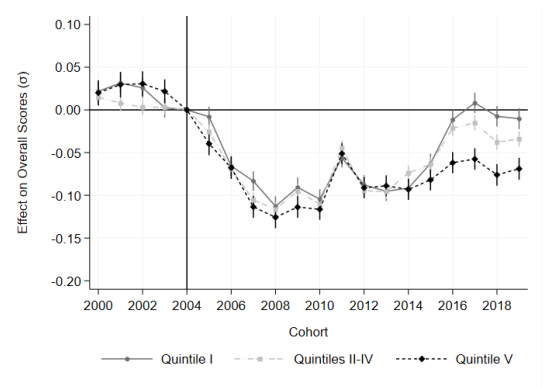
Notes. Ordinary Least Squares estimates of the dynamic effects, δ_τ, of equation 1. Outcome variables correspond to overall performance on the high school exit exam and test scores in all evaluated subjects. Overall scores are computed as the average performance in five subject exams: reading comprehension, 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 10: Dynamic Heterogeneous Effects of a Merit-Based Teacher Hiring Policy on Students' Outcomes

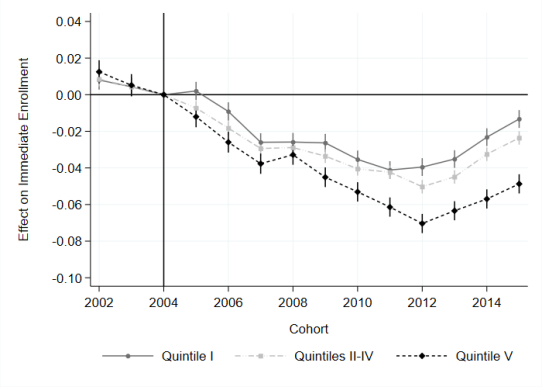
(a) Baseline Model: Overall Scores



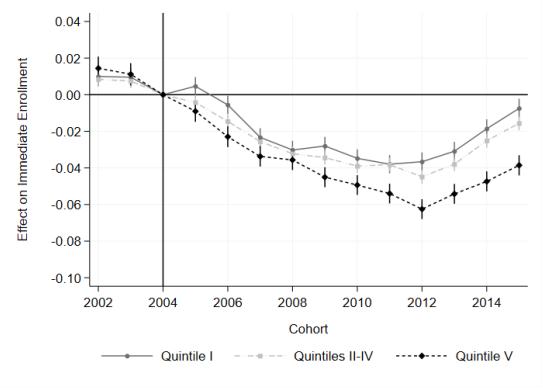
(b) Full Model: Overall Scores



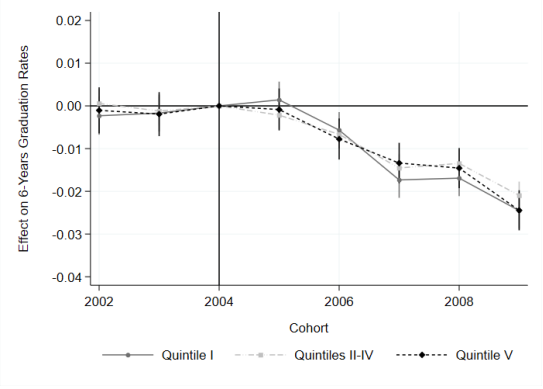
(c) Baseline Model: College Enrollment



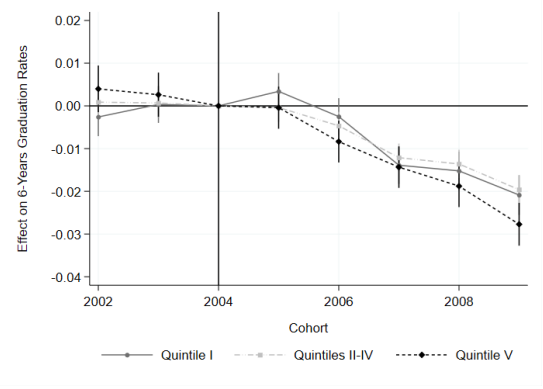
(d) Full Model: College Enrollment



(e) Baseline Model: College Graduation

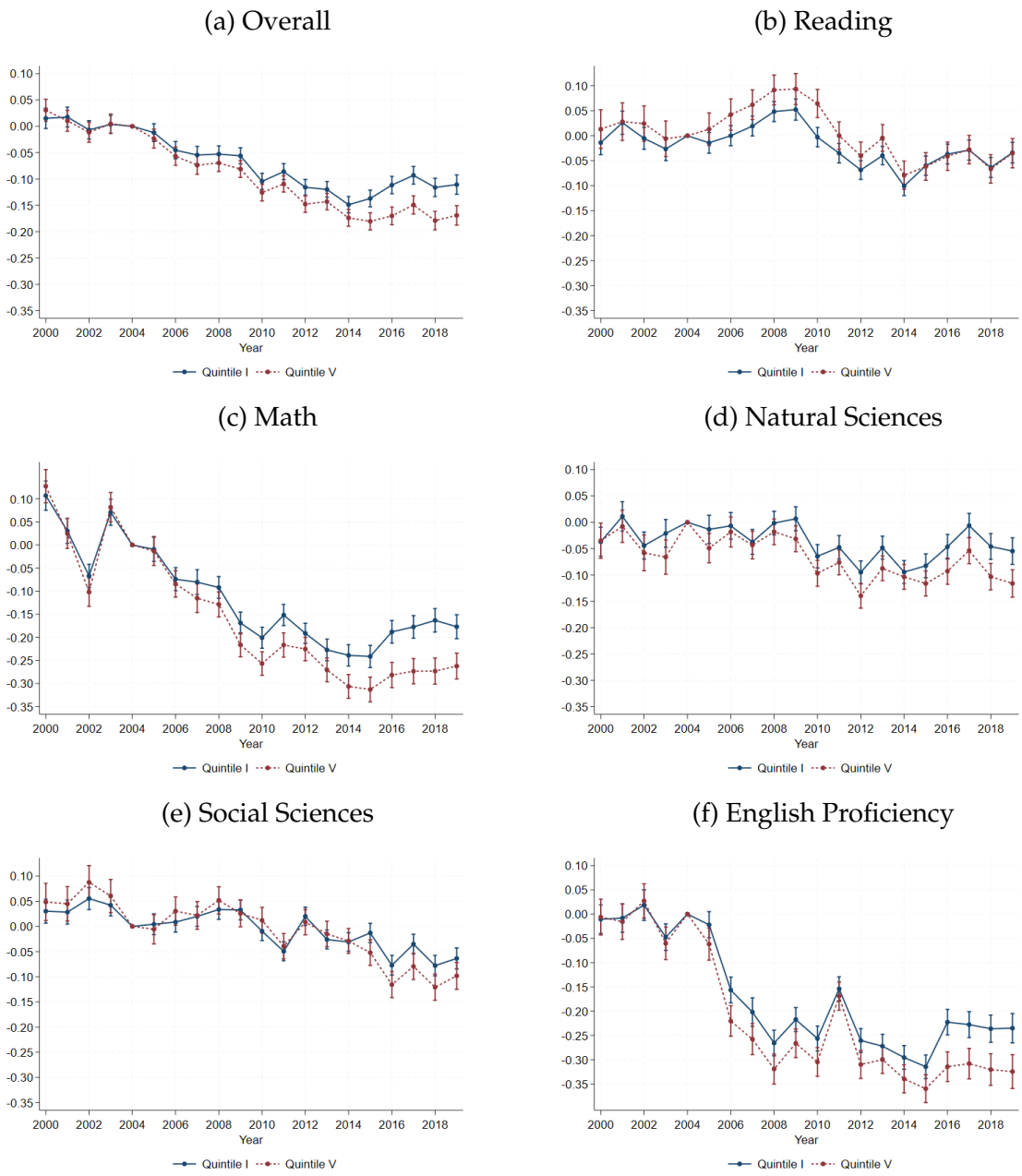


(f) Full Model: College Graduation



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

Appendix Figure 11: Heterogeneous Effects by Share of Novice Teachers Exploiting Across Subject Variation



Notes. Ordinary Least Squares estimates of the dynamic effects, δ_τ^I and δ_τ^V, of equation $Y_{ijst} = \mu_t + \mu_{sj} + \sum_\tau(\delta_\tau^I \times \text{Novice}_{sj}^I + \delta_\tau^{II-IV} \times \text{Novice}_{sj}^{II-IV} + \delta_\tau^V \times \text{Novice}_{sj}^V) \times \mathbb{1}[\tau = t] \times \text{Public}_s + \varepsilon_{ijst}$. Three mutually exclusive groups are defined based on quintiles of the fraction of novice teachers at a student's school in a particular subject 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 in a subject. We define novice teachers as teachers hired within the last five years. Novice_{sj}^I is an indicator for whether a student is classified in Quintile I, while Novice_{sj}^V indicates if the student is in Quintile V. 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 specification includes school-subject and year fixed effects. 95% confidence intervals are displayed around plotted coefficients and are computed using standard errors clustered at the school $\times$ year level.

Appendix Table 1: Descriptive Statistics of Public School Teachers

	Public Teachers	Private 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 2: Statistics of College Students and Graduates by Field of Study

	Field of Study :							
	Education	Agricultural Sciences	Business & Accounting	Social Sciences	Health	Engeneering	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) :								
$t = 1$	571.20	533.42	642.11	662.95	918.46	753.93	709.11	768.49
	(290.10)	(341.15)	(437.53)	(411.57)	(634.68)	(505.70)	(482.58)	(486.32)
$t = 2$	627.52	597.29	726.20	749.16	951.30	876.26	822.45	878.39
	(324.14)	(423.19)	(506.82)	(468.82)	(658.83)	(593.21)	(568.09)	(562.87)
$t = 3$	688.72	665.26	818.19	839.38	1000.84	1005.02	959.61	1005.04
	(354.50)	(473.19)	(583.35)	(533.37)	(671.40)	(682.36)	(657.89)	(660.74)
$t = 4$	747.01	737.45	922.46	942.36	1070.26	1143.93	1095.48	1133.17
	(382.51)	(513.31)	(665.26)	(609.88)	(698.07)	(768.55)	(750.33)	(749.12)
Panel C : Graduates (excluding public teachers)								
Age at graduation	26.05	25.39	25.42	24.77	24.52	24.74	24.35	24.68
	(3.47)	(3.11)	(3.57)	(3.28)	(2.84)	(2.95)	(2.87)	(2.60)
Earnings Aft. Grad. (t) :								
$t = 1$	549.34	532.29	642.08	665.42	918.66	755.62	709.40	779.32
	(292.20)	(342.18)	(437.49)	(415.75)	(634.92)	(507.34)	(482.65)	(492.66)
$t = 2$	593.20	595.34	726.07	753.44	951.47	878.58	822.65	893.30
	(318.57)	(425.48)	(506.62)	(475.30)	(658.97)	(595.28)	(567.05)	(574.13)
$t = 3$	641.85	663.54	818.33	845.20	1001.06	1008.21	960.86	1024.16
	(361.01)	(476.23)	(584.05)	(540.17)	(671.63)	(684.25)	(659.46)	(671.80)
$t = 4$	698.04	737.42	922.35	951.04	1070.42	1148.62	1097.42	1158.30
	(408.59)	(517.87)	(665.44)	(621.09)	(697.91)	(771.73)	(753.14)	(764.71)

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 and Panel C correspond to the pool of students graduating from college between 2007 and 2014, however, Panel C excludes public teachers. 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 3: Public Teachers' Descriptive Statistics

	Civil Service Teachers		Contract Teachers		Mean Difference
	Mean	S.D.	Mean	S.D.	p-value
	(1)	(2)	(3)	(4)	(1) - (3)
Teacher's Characteristics :					
Age	46.97	9.50	37.44	8.80	0.00
Female	0.66	0.47	0.65	0.48	0.00
Degree :					
Teaching Vocational High School	0.21	0.40	0.27	0.44	0.00
Professional	0.55	0.50	0.69	0.46	0.00
Post Graduate	0.25	0.43	0.04	0.20	0.00
Schools's characteristics :					
Rural	0.27	0.44	0.50	0.50	0.00
Urban	0.69	0.46	0.48	0.50	0.00
Teaching level :					
Preschool	0.05	0.23	0.04	0.20	0.00
Primary	0.45	0.50	0.43	0.49	0.00
High School	0.41	0.49	0.48	0.50	0.00
Subject :					
Natural Sciences	0.16	0.37	0.15	0.36	0.00
Social Sciences	0.24	0.43	0.17	0.38	0.00
Spanish	0.21	0.41	0.21	0.41	0.64
English	0.15	0.36	0.19	0.39	0.00
Math	0.24	0.43	0.27	0.45	0.00
Observations	2,428,502		444,716		2,873,218

Notes. Summary statistics pooling public teachers from 2007 to 2015, from the public teachers administrative database (Anexo 3A). Statistics are given for civil service and contract teachers in columns (1)-(2) and (3)-(4) respectively. Column (5) shows the p-value for the mean difference of each variable across these two groups. Degree categorises teachers according to the highest level of education achieved. Schools' characteristics classify whether the teacher works in a rural or urban school. Finally, Teaching level and Subject show the proportion of teachers teaching at preschool, primary school or high school as well as the subject they teach: natural sciences, social sciences, spanish language or literature, english language, and math.

Appendix Table 4: Teacher Sorting into End-of-High School Exam

	Dependent Variable :				
Panel A :	Average High School Exit Exam of Placement School				
	(1)	(2)	(3)	(4)	(5)
Score in Civil Service	0.038***	0.040***	0.026***	0.038***	0.027***
Entrance Exam	[0.002]	[0.001]	[0.001]	[0.001]	[0.001]
R-squared	0.012	0.020	0.212	0.015	0.221
Observations	118,319	117,878	118,319	118,319	117,878
Demographic controls	No	Yes	No	No	Yes
Department FE	No	No	Yes	No	Yes
Year FE	No	No	No	Yes	Yes
Panel B :	$\mathbb{1}$[Same Municipality Before and After Civil Service Exam]				
	(1)	(2)	(3)	(4)	(5)
Score in Civil Service	0.010***	0.011***	0.013***	0.012***	0.015***
Entrance Exam	[0.002]	[0.001]	[0.001]	[0.001]	[0.001]
R-squared	0.000	0.019	0.057	0.005	0.078
Observations	124,837	124,316	124,837	124,837	124,316
Demographic controls	No	Yes	No	No	Yes
Department FE	No	No	Yes	No	Yes
Year FE	No	No	No	Yes	Yes

Notes. Estimated coefficients where obtained by taking the sample of teachers who passed the public call exam between 2005 and 2013 and matching them with the mean standardized saber 11 score of the school where they began working in that year. Columns (1) through (4) are sub-specifications of column (5): $y_{isdt} = \mu_d + \mu_t + \delta ScoreCivilService_{isdt} + \beta X_i + \varepsilon_{isdt}$. $ScoreCivilService_{isdt}$ is the standardized score of the civil service entrance exam of teacher i, at the year t, which began working at school s in department d.In Panel A y_{isdt} is the average standardized high school exit exam score of the school where the teacher was placed. In Panel B y_{isdt} is an indicator variable of whether a teacher is in the same municipality before and after taking the civil service exam . μ_d and μ_t are department and year fixed effects respectively. The coefficient of interest is δ. The controls X_i include the teachers' age and sex. Clustered errors in brackets.[a]

[a]The proportion of teachers who teach in the same municipality where they took the test is 49.5%. By quintiles of their score in the civil service entrace exam it is (QI, QII, QIII, QIV, QV) = (47.4%, 46.7%, 47.5%, 50.9%, 54.4%)

Appendix Table 5: Transition Matrix by public and private teacher

	Percentage Transitioning Within the Following Three Years:		
Year	Public	Private	Other Sector
A) Public			
2007	0.906	0.006	0.088
2008	0.916	0.007	0.077
2009	0.896	0.009	0.095
2010	0.903	0.011	0.086
2011	0.926	0.007	0.068
2012	0.928	0.006	0.065
B) Private			
2009	0.018	0.592	0.390
2010	0.004	0.798	0.199
2011	0.006	0.748	0.246
2012	0.005	0.643	0.352

Notes. The transition matrix was created using the public teacher database and the Colombian social security records (known as the "planilla integrada de liquidación de aportes" (PILA)). Teachers starting as public can be identified since 2007 and teachers starting as private since 2009 due to the initial year of the aforementioned databases respectively. Public teachers are said to continue if observed as such within the next three years, otherwise their CIIU code is checked, if it is teaching-related then they are classified as a private teacher, otherwise they are classified in Other Sector. Private teachers have a CIIU code related to education and are not identified as a public teachers. They are considered to continue as private teachers if their CIIU code is still related to education and they are not in the public teacher data within the next three years. They are said to change to the public sector otherwise. If their CIIU code changes to non education related or it is missing then they are classifed as Other Sector

Appendix Table 6: Age of new public teachers through time

	Mean	Median	S.D.	Observations
	(1)	(2)	(3)	(4)
Entered public sector in :				
1995	30.38	30	6.34	17148
1996	31.06	30	6.60	12605
1997	31.78	31	7.11	10717
1998	32.32	32	7.12	5863
1999	32.19	31	7.09	5809
2000	33.37	33	7.46	4477
2001	34.58	34	7.55	3775
2002	38.52	38	8.80	1374
2003	36.88	36	7.78	8019
2004	33.84	33	8.07	20309
2005	33.44	33	7.89	16846
2006	34.29	33	8.12	17422
2007	35.49	34	8.17	11424
2008	34.51	33	8.53	14975
2009	34.71	33	8.72	9636
2010	33.95	33	7.82	29493
2011	34.43	33	8.27	13305
2012	34.71	33	8.45	11029
2013	34.80	33	8.80	11629
2014	33.81	32	8.53	10804
2015	33.56	32	7.91	25107

Notes. Summary statistics through time (from 1995 to 2015) of the age of public teachers the year they began working as such. From the public teachers administrative database (Anexo 3A).

Appendix Table 7: Effect of Merit-based Teacher Hiring Policy on Students' College Outcomes Using a Constant Sample

	Dependent Variable :								
	College Enrollment						College Graduation		
	Immediate			2-year					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Public $\times$ $\mathbb{1}[t \geq 2005]$	–0.025***	–0.027***	–0.021***	–0.019***	–0.027***	–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]
Mean of Dependent Variable for Public Students	0.095	0.095	0.095	0.251	0.251	0.251	0.075	0.075	0.075
R-squared	0.149	0.155	0.156	0.200	0.216	0.217	0.083	0.087	0.088
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
Municipality Trends		Yes	Yes		Yes	Yes		Yes	Yes
School Trends			Yes			Yes			Yes

Notes. Ordinary Least Squares estimates of the effect of a merit-based teacher hiring policy on college outcomes, based on the following equation: $Y_{ist} = \mu_t + \mu_s + \delta\,(\text{Public}_s \times \mathbb{1}[t \geq 2005]) + X_i'\gamma + \varepsilon_{ist}$. Public_s indicates whether a student is enrolled in a public high school. Outcome variables are displayed at the top of each column and correspond to college enrollment and graduation indicators with different time windows. Immediate enrollment indicates whether a student enrolls in a college program within six months of graduating high school. 2-year enrollment indicates whether a student enrolls in college within the next two years. College graduation indicates whether a student graduates from college in the following six years after completing high school. All regressions include school and year fixed effects. Specifications with covariates control for the student's age, gender, socioeconomic stratum, and schooling time (i.e., whole day, morning, afternoon, night, or weekends). All results are based on information of students from cohorts 2002 to 2009. Standard errors are displayed in square brackets and clustered at the school $\times$ year level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Appendix Table 8: Heterogeneous Effects of Merit-Based Teacher Hiring

	Dependent Variable : Test Scores (σ)		
Panel A :			
	Share of Novice Teachers		
	(1)	(2)	(3)
Public $\times$ $\mathbb{1}[t \geq 2005] \times$ Novice^{I}	–0.093***	–0.095***	–0.072***
	[0.004]	[0.004]	[0.004]
Public $\times$ $\mathbb{1}[t \geq 2005] \times$ Novice^{II-IV}	–0.113***	–0.108***	–0.092***
	[0.004]	[0.004]	[0.004]
Public $\times$ $\mathbb{1}[t \geq 2005] \times$ Novice^{V}	–0.124***	–0.110***	–0.082***
	[0.004]	[0.004]	[0.005]
P-value $\alpha^I = \alpha^V$	0.000	0.000	0.008
Observations	36,912,611	36,912,611	36,912,611
Panel B :			
	Rural and Urban Schools		
	(1)	(2)	(3)
Public $\times$ $\mathbb{1}[t \geq 2005] \times$ Rural	–0.144***	–0.125***	–0.067***
	[0.004]	[0.005]	[0.005]
Public $\times$ $\mathbb{1}[t \geq 2005] \times$ Urban	–0.104***	–0.103***	–0.085***
	[0.003]	[0.003]	[0.003]
P-value $\beta^U = \beta^R$	0.000	0.000	0.000
Observations	44,753,295	44,753,295	44,753,295
Panel C :			
	Pre-Reform High School Exit Exam Scores		
	(1)	(2)	(3)
Public $\times$ $\mathbb{1}[t \geq 2005] \times$ Pre-Reform Score^{I}	–0.130***	–0.147***	–0.033***
	[0.004]	[0.005]	[0.005]
Public $\times$ $\mathbb{1}[t \geq 2005] \times$ Pre-Reform Score^{II-IV}	–0.106***	–0.101***	–0.088***
	[0.003]	[0.004]	[0.004]
Public $\times$ $\mathbb{1}[t \geq 2005] \times$ Pre-Reform Score^{V}	–0.122***	–0.113***	–0.125***
	[0.007]	[0.006]	[0.006]
P-value $\delta^I = \delta^V$	0.196	0.000	0.000
Observations	38,988,335	38,988,335	38,988,335
Covariates		Yes	Yes
Municipality Trends			Yes

Notes. Ordinary-least-squares estimates of the following equations: for Panel A $Y_{isjt} = \mu_t + \mu_{sj} + (\alpha^I \times Novice^I_{sj} + \alpha^{II-IV} \times Novice^{II-IV}_{sj} + \alpha^V \times Novice^V_{sj}) \times \text{Public}_s \times \mathbb{1}[t \geq 2005] + X_i'\gamma + \varepsilon_{isjt}$. For Panel B $Y_{isjt} = \mu_t + \mu_{sj} + (\beta^R \times Rural_s + \beta^U \times Urban_s) \times \text{Public}_s \times \mathbb{1}[t \geq 2005] + X_i'\gamma + \varepsilon_{isjt}$, and for Panel C $Y_{isjt} = \mu_t + \mu_{sj} + (\delta^I \times \text{Pre-Reform Score}^I_{sj} + \delta^{II-III} \times \text{Pre-Reform Score}^{II-III}_{sj} + \delta^{IV} \times \text{Pre-Reform Score}^{IV}_{sj}) \times \text{Public}_s \times \mathbb{1}[t \geq 2005] + X_i'\gamma + \varepsilon_{isjt}$. Public_s indicates whether a student is enrolled in a public high school. Mutually exclusive groups are defined based on: 1) quintiles of the fraction of novice teachers at a student's subject and school by 2008 for (panel A); indicator variables of whether the student is enrolled in a public or private school (panel B); Quartiles of the schools' average high school exit exams from 2000 to 2004 (panel C). The base category are private schools. 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-subject 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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.