Economic Analysis and Policy 73 (2022) 62–93

Contents lists available at ScienceDirect

Economic Analysis and Policy

journal homepage: www.elsevier.com/locate/eap

Analyses of topical policy issues

Turning around teacher quality in Latin America: Renewed confidence and lessons from Colombia☆

Zelda Brutti [a,*,1], Fabio Sánchez Torres [b]

[a] Pompeu Fabra University, BSE and IEB, Department of Economics, Edifici Jaume I, floor1, Carrer de Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain
[b] Universidad de Los Andes and CEDE, Bogotá, Colombia

ARTICLE INFO

Article history:
Received 28 July 2021
Received in revised form 21 October 2021
Accepted 23 October 2021
Available online 3 November 2021

JEL classification:
I25
I28
J08
J45

Keywords:
Teacher quality
Teacher selection
Public education reform
Standardized testing
Latin America
Colombia

ABSTRACT

We present new evidence on the effectiveness of large-scale reforms of public teaching careers towards improving student outcomes in Latin America. Since 2002, Colombia's public teacher candidates are screened through an entry competition featuring a standardized examination, and their subsequent career progression is subject to performance evaluations; the fact that educators hired before the reform remained exempt from the new rules provides a unique research setting for policy evaluation. We are able to exploit unusually detailed data on the universe of secondary schools and teachers in Colombia; our empirical strategy exploits variation in student performance and teacher types across different subjects, within any given school and year. We pin down a sizeable performance premium of quality-screened teachers with respect to their traditional colleague: the former improve student achievement by about 7% of a standard deviation within school and year. We are able to exclude various types of selection biases, cohort effects and other confounders. We show that the Colombian entry examination is an effective candidate screening device and that teachers' achievement on the exam is predictive of their performance on the job. We provide suggestive evidence on how the reform affected other policy-relevant aspects such as teacher candidate quality and career attrition.

1. Introduction

School teachers are the single most important factor generating educational progress, as well as the main responsible for the Latin American disadvantage with respect to other regions of the world (Bruns and Luque, 2014). Historically, the public teacher profession in this region has embodied many of the issues that can affect public sector employment, such

☆ The authors gratefully acknowledge helpful suggestions by Jerome Adda, Raquel Bernal, Adriana Camacho, Leopoldo Fergusson, Andrea Ichino, Rachid Laajaj, Dario Maldonado, Eric Maurin, Steve Pischke, Luis Rojas and a number of anonymous referees. Many thanks are also due to seminar participants at the London School of Economics, the European University Institute, the University of Los Andes, and at conferences hosted by the Econometric Society, the European Economic Association and LACEA. We thank Nicolás Acevedo Rebolledo for dedicated research assistantship. Zelda Brutti is grateful for financial support received during the course of this project by the Centro Studi Luca d'Agliano (Turin, Italy).

* Corresponding author.
E-mail addresses: zelda.brutti@upf.edu (Z. Brutti), fasanche@uniandes.edu.co (F. Sánchez Torres).

[1] Most of this work was carried out while affiliated with the European University Institute (Florence, Italy), the University of Barcelona and the IEB (Barcelona, Spain).

https://doi.org/10.1016/j.eap.2021.10.008

as limited aspirations and poor work ethic. A growing body of empirical research has shown that noncompetitive hiring mechanisms, high job security, low-powered incentive structures and weak monitoring are systematically associated to a range of faults and inefficiencies in the public sector, especially in developing countries (Finan et al., 2017).

Despite this evidence, large reforms of the public teaching careers are still uncommon on the continent (Elacqua et al., 2018). In fact, beyond the pioneering initiative started by Colombia in 2002, one can count only a handful of radical and nationwide reform projects over the last two decades, including Chile (2004–2016), Mexico (2008), Ecuador (2011) and Peru (2012).

This widespread inertia is explained by the fact that profound revolutions to the public teacher profession are extremely challenging endeavors from a social and political point of view, typically met with vehement opposition from powerful teacher unions. Legarralde and Palamidessi (2006) find the social mobilization and political obstruction capacity of Latin American teacher unions to be the most relevant factor in the lack of modernization in the sector. For instance, Estrada (2019) discusses how the Mexican reform of the teacher hiring system became possible only after a sequence of media scandals and policy reports, through which the discretionary and corrupt practices controlled by the national teacher union raised the disapproval of the general public. Similarly, in Ecuador the government faced fierce resistance from the national teacher union when it proposed a transition to more transparent, standardized and merit-based methods of hiring and evaluating teachers; the reform went through after unusually strong pressure from media and politicians, and large financial investments to compensate teachers with better retirement benefits (Chiriboga Montalvo and Pinto Haro, 2019). Repealing the new rules introduced by the 2012 *Ley de Reforma Magisterial* was a prominent motivation behind the several-months-long 2017 teacher strike in Peru, and in Chile most aspects of the current teacher regulation are the result of intense negotiations between congress and the main teacher union (Mizala and Schneider, 2019). Further, the social and political complexities surrounding the profession often also imply scarce availability of systematic data on individual public teacher performance. In the light of the above considerations, is not surprising that there is little empirical evidence to date on the success of "modernizing" public teacher careers, in terms of raising student achievement on the Latin American continent. In this study we report the impact of a radical reform that transformed the career of public school teachers in Colombia, a country that over the last two decades has made education a main priority to improve the economic and social prosperity and pledged more resources to this sector than any other policy area (OECD, 2016).

The reform we study aimed at improving the quality of public school teachers by introducing merit-based incentives to it, most importantly strict entry screening based on the performance on a standardized examination. Our main goal is to answer the question of whether teachers hired through standardized entrance tests and subject to the new merit-based regulation, featuring both good-performance incentives and bad-performance threats, exhibit a quality premium with respect to their traditional colleagues. Throughout this study, we measure teacher quality through its impact on student achievement.

In our analysis, we are able to exploit a large wealth of administrative data on teachers and students over the period 2008–2016. Due to the agreement reached with unions (*FECODE, Federación Colombiana de Educadores*), incumbent teachers remained completely immune to the new set of rules, so that these applied only to new entrants hired after 2002. On one hand, this fact significantly limited the speed at which the institutional change has become visible in the education system, but on the other it provides a very favorable research setting for us to explore the effects of the reform. During our observation window – and still today – traditional and new-regulation teachers coexist in the Colombian school system, in varying shares across different schools and even within schools, across different subject areas taught. We exploit this variation to create statistical associations between new-regulation teacher shares and student performance at the secondary school level. Our data allows us to operate at a detailed within-school level of observation and thus obtain reliable estimations of the impact of new-regulation teachers on pupil achievement. We find that being taught by quality-incentivized teachers, rather than by their traditional colleagues, does indeed improve students' performance.

The magnitude of the effect is sizable and compares to the difference in being taught by experienced rather than novice teachers. It also corresponds to about half the difference made by a high-quality versus a low-quality teacher, according to mainstream estimates from the US context (Hanushek and Rivkin, 2012). In contrast to previous findings on Mexico and Ecuador, we show that the Colombian entry exam is an effective quality screening device: performance of teacher candidates on this exam helps predicting their performance on the job, in magnitudes that are very well aligned with recent findings on US teacher certification exams (Goldhaber et al., 2017a,b; Cowan and Goldhaber, 2016; Goldhaber and Anthony, 2007). Further, we provide descriptive evidence on an improvement in educational qualifications among teachers belonging to the new career regulation with respect to their similarly aged traditional colleagues.

Previous evidence on the effects of the Colombian reform was very limited. The comprehensive report by García Jaramillo et al. (2014) provides an excellent review of the main features, merits and challenges of the teacher profession in Colombia; however, it does not attempt any impact analysis of the new regulation on student achievement. An impact study in the same spirit as ours was carried out by Ome (2012, 2013) and in contrast to us, the author does not find any effect of the new-regulation teachers on test scores at secondary school level. We attribute this discordance to the fact that Ome (2012, 2013) had at his disposal a less convenient dataset: first, he focuses on an observation window which is both shorter and closer to the reform with respect to ours, so that the effects of the reform may not have had a chance to materialize. Second, the lower level of detail in the data forced this earlier study to compare student achievement across time rather than within the same academic year as we do, so that the effect of the reform was presumably confounded by a number of changes affecting schools over the years.

Our study represents a relevant contribution to the existing literature on teacher quality in developing and middle-income countries. It is among the first studies on the Latin American context drawing rigorous evidence from nationwide, administrative data on public schools and teachers, rather than focusing on subsets of special schools or geographical areas. Moreover, the Colombian reform of the teaching profession represents a particularly clear-cut example of introducing merit-based selection procedures and incentives to public servants' careers, which speaks to the recent literature on the personnel economics of the state (Finan et al., 2017). Importantly, the pre-reform context in Colombia was not different from the ones still in place in several neighboring countries at the time of writing[2], making of the results of this study valuable policy guidance for future educational reforms on the continent.

The remainder of this paper structured as follows. In Section 2, we provide a review of the literature related to our study. In Section 3, we proceed to describe how Colombia changed its system from 2002 onward. In Section 4, we describe the data we used for our analysis. In Section 5, we illustrate the empirical methodology we employed to assess the impact of the Colombian reform on student achievement. In Section 6, we offer our findings on the impact of the Colombian reform and in Section 7 discuss two possible channels explaining them. Section 8 offers concluding remarks and policy insights that can be drawn from this study.

2. Related literature

There is a new generation of public teaching careers being gradually instituted on the Latin American continent. These new career regulations are characterized by a number of common elements, such as being goal-oriented; establishing merit as the main criterion for entry, retention and advancement in the profession; featuring high-stake continuous evaluation protocols; featuring monetary or non-monetary performance incentive schemes (Cuenca and Carrillo 2017). In Appendix A we review the available evidence on three of the most comprehensive educational reforms directly targeted at improving teaching quality that have been carried out in Latin America over the last two decades: Chile, Mexico and Ecuador, highlighting the value added of the Colombian experience with respect to each of these. The currently available evidence on the effectiveness of such radical and large-scale reform initiatives is decidedly scarce: in many instances, the reforms are too recent in time to draw any conclusions yet; in other cases, there is a lack of an appropriate data collection structure that would enable rigorous impact analysis; in most situations, the slow and gradual implementation of the new rules on the entire teacher population makes it challenging to distinguish their effects from other institutional changes that occur contemporaneously in the country.

More in general, by quantifying the impact of Colombian 'New Regulation' teachers on student performance we primarily contribute to the literature on teacher quality, and in particular to the branch of research aiming to understand which type of policies are effective in creating, attracting, hiring and retaining better teachers to the public system. There is a growing consensus that the quality of teacher hiring processes are decisive in improving students' outcomes and recently, protocols have been shifting from being mainly credential and certification-based to promoting direct knowledge testing. Most of the available evidence on teacher testing is based on US studies and confirms the higher effectiveness of this approach in selecting better candidates.[3] The Colombian reform we analyze in this paper represents an unusual case of a middle income country radically changing its public teacher hiring process and tilting its incentive package towards meritocracy.[4]

In the perhaps closest study to ours, Estrada (2019) analyzes the Mexican context and exploits the temporary co-existence of discretionary and test-based hiring systems to carry out a difference-in-difference analysis. Estrada's study represents one of the very few on public official recruiting in Latin America using administrative, large scale data. He finds that test-based selection yields massive advantages in terms of student achievement, *despite* the test's inability to predict teacher performance, thus concluding that Mexican teachers hired through the discretionary system must belong to the bottom of the quality distribution. However, Estrada (2019) focuses exclusively on a subset of small schools, characterized by only one teacher per classroom, and catering to small and often disadvantaged communities ("*Telesecundaria*" schools), so that it remains unclear whether his results would apply to wider contexts as well. Indeed we show that, unlike in the Mexican case, the score on the Colombian selection exam is a good proxy of later teacher performance, reviving the debate on test-based selection in the Latin American region. Discordantly with respect to both the Mexican and the Colombian experiences, Araujo (2019) finds that in the Ecuadorian case, test-screened teachers do not perform any better than their peers in general, and yield benefits only when matched with students from disadvantaged backgrounds. A previous study of the Colombian teacher reform was carried out by Ome (2013, 2012), who uses three years of data on student performance and exploits variation across years for identification. Contrary to us, he finds no effect of the new-regulation teachers on test scores at high school level: we attribute this contrast to the fact that our analysis is able

[2] Such as Bolivia, Brazil, Venezuela, Nicaragua, El Salvador and Panama. For situations similar to pre-reform Colombia across South Asia and the rest of the world, refer to the discussions in Murillo et al. (2007), Hanushek (2009) and Hanushek and Woessmann (2012).

[3] Among others, see Boyd et al. (2007), Rockoff et al. (2011), Goldhaber et al. (2017a), Goldhaber et al. (2017b), Cowan and Goldhaber (2016), Goldhaber and Anthony (2007).

[4] Secondarily, our results also speak to the question of how much of the variation in student outcomes is explained by teacher effects (see Chetty et al. (2014), Rivkin et al. (2005), Rockoff (2005) and, for an excellent review, (Hanushek and Rivkin, 2012)) and add to the evidence on which individual traits are predictive of better teachers (see among others Rockoff et al. (2011), Kane et al. (2008), Gordon et al. (2006)).

to deal with the selection of teachers across schools over time, which we find biasing estimates towards zero. Given the small amount of rigorous, large scale evidence and the lack of consensus on the effectiveness of test-screening of public teachers in Latin America, we believe our study represents a relevant addition to this literature.

When looking at how the different components of the new career regime influence teachers' incentives and behaviors, we also contribute to the relatively recent literature on the personnel economics of the state, which engages with aspects of the state–employee relationship such as selection, incentive structures and monitoring (Finan et al., 2017). Focus of this literature is how the characteristics of public sector careers such as payment, job security and job status influence the type of applicants and service quality. So far, most of the evidence on these matters is provided by field experiments. Examples are the ones conducted by Ashraf et al. (2018) in Zambia, finding that career-track motivations resulted in more talented hires of healthcare workers and better health outcomes; by Dal Bó et al. (2013) in Mexico, finding that higher salaries were able to attract more skilled and motivated civil service job applicants; by Leaver et al. (2021) in Rwanda, Muralidharan and Sundararaman (2011) in India and Lavy (2009) in Israel, finding large, positive effects of "pay for performance" schemes applied to teachers. Conversely, Deserranno (2019) finds that the expectation of higher earnings discouraged prosocial applicants for village promoter roles, resulting in lower effort and retention in Uganda. We present suggestive evidence indicating that the education level of public teacher candidates improved following the merit-oriented Colombian reform, a result which is in line with the majority of the experimental evidence mentioned above.

Finally, our paper contributes to the literature addressing less-expected consequences of screening mechanisms such as higher turnover in civil servants. In principle, any initial testing requirement for candidates to public service may present some unintended consequences, which explains the difficulty of designing distortion-free screening mechanisms (Bruns and Luque, 2014; García Jaramillo et al., 2014). For example, Larsen (2013) finds that the implementation of more stringent qualifications for teachers led some high-quality junior teachers to abandon the profession, but it also improved the qualifications of the remaining teachers and led to increased student scores. A possible mechanism for this phenomenon is that increasing the licensing requirements increase the opportunity costs associated with choosing this profession, as argued by Goldhaber and Liu (2003) and Eide et al. (2004). Moreover, the implementation of testing requirements in teacher hiring may lead to changes in their career trajectories. For example, Wiswall (2007) finds that this increasing opportunity cost drives out skilled teachers, which leads to shorter average tenures. Hendricks (2016) confirms this phenomenon, where testing requirements initially drive out lower ability teachers, but increase attrition among higher ability teachers after the first year. However, Boyd et al. (2005) suggest that higher ability teachers do not leave teaching for good, but transfer to higher-performing schools. Also this strand of research dries out quickly when leaving the US and moving to the south of the continent. In this light, in line with several of the above studies, we add to the literature by documenting that the new merit-based incentive package introduced in Colombia has on average increased teacher turnover with respect to the traditional system, is in general able to retain better candidates relative to low-performing ones, but is not successful with top performers, who show the highest rates of attrition.

3. The 2002 reform of the teacher career

Colombia has been a Latin American pioneer in radically innovating its public teacher career regulation, and still lies among a mere handful of truly radical initiatives marking the region over recent decades. The 'Old Regulation' (OR) of the Colombian public teacher career dated back to 1979, and was characterized by loose access requirements, opaque candidate evaluation processes and overprotection of tenured teachers. The OR required public teachers to hold the appropriate education level set by the law, but the appointment of candidates to vacancies was considered a mere administrative procedure in the hands of local authorities. As a consequence, hiring processes were under the responsibility of regional governors or city mayors, lacking any rigor or standardized public competition among candidates (Congreso de Colombia, 1994; Consejo de Estado de Colombia, 2004; Cifuentes Cubillos, 2014). This system lent itself to dynamics of corruption and clientelism, beyond not fostering the investment in individual teaching skills and subject knowledge (Duarte, 2001). Once a teacher had entered the career, upgrades in the seniority scale were based solely on years of service and on education titles, without teaching performance playing any role. Public teachers were also guaranteed their job until retirement — except in the extreme cases of ascertained severe misconduct, involving criminal behavior or sexual offenses.

Overall, the legal framework was characterized by little transparency in procedures, excessive protection of employed teachers and lack of incentives towards the improvement of teaching skills. Clientelism and politicization of teacher appointments represented substantial and well-known issues; far too often public schools were used as 'placement pools' for relatives and connections of influential personalities (see excellent description and discussions in Duarte (2001, 2003)). In fact, the new 1994 General Law for Education (Congreso de Colombia, 1994) had started to indicate public contests as the necessary method for the appointment of new teachers, but this feature was never enacted in practice, and the legal reference for public teacher careers remained the traditional Estatuto Docente 2277/1979. The following subsections describe how this scenario changed after the 2002 reform of the public teacher career, which introduced the Nuevo Estatuto Docente 1278/2002.

3.1. Entry competition

Teachers belonging to the 'New Regulation' (NR), entering the profession from January 1st 2002 onward, must go through a public selection contest in order to be assigned a public vacancy. The contest is based on a score system and serves the purpose of establishing a rank among applicants, which determines the order in which they are allowed to choose their preferred vacancies. The stages of the selection contest and their weights are summarized in Table A.4. In the first stage of the selection procedure, candidates sit an exam which is identical across the whole nation and is administered and evaluated centrally by governmental agencies[5]; it is structured into three modules testing teaching aptitude, subject knowledge and psychometric values[6]. Candidates who do not score a minimum of 60 points out of 100 are excluded from the rest of the competition. The exam stage represents 65% of candidates' global contest score. Further points are earned for academic credentials such as degrees, attendance of training courses, academic publications, past teaching experience and evaluations (if available), and for holding specific awards, according to official standards set by the Colombian 'Civil Service Commission'[7] (CNSC), the governmental body in charge of this stage of the context. The credentials stage weighs 20% of the total contest score. The final face-to-face interview of candidates is also responsibility of CNSC, who may nominate local delegate bodies to decentralize the process. Typically, universities and other certified higher education institutions are chosen to form ad-hoc interview committees. The committee questions each candidate verbally, and the evaluation accounts for the remaining 15% of the global score on the competition.

The new standardized entry competition has been the most salient innovation to the public teaching career, and is universally considered the most important quality-selection instrument currently in place (García Jaramillo et al., 2014). In fact, the first entry competition took place in 2004, so that the 2002–2004 period may be referred to a "transition period" between regulations.

3.2. Changes after entry

Beyond the strict entry procedures, the New Regulation (NR) also introduced continuous evaluation practices, aiming at maintaining a satisfactory performance by teachers over time, as well as providing them with incentives to improve further. However, the reality of post-entry evaluation has proven itself to be far less rigorous than in the intentions of the law, as we further discuss here below and along the paper.

Once successfully assigned to one of the available vacancies, the aspiring teacher starts a probation period that lasts for a minimum of four months and up to the end of the academic year. The candidate's performance is evaluated by the school principal through a standard assessment procedure, outlined by the Ministry of Education (*Evaluación de periodo de prueba*), consisting in a process of collecting evidence on performance, continuous feedback and final assessment through a point system. Conditional on a positive outcome, the new teacher takes permanent possession of the chair, but shall be dismissed in the case of a negative assessment. Once in a permanent position, teachers face further assessment schemes. The first consists in yearly performance assessments filed to the local education authority, and again sees school principals in the role of evaluators (*Evaluación anual de desempeño laboral*). The reports have a constructive nature, and are meant to help teachers improve on their weaker aspects; however, a positive assessment is necessary in order to progress in the pay scale, and according to the law, two consecutive years of poor evaluations lead to dismissal of the teacher. Both the probation period and the yearly evaluation practices just described are highly controversial, for two main reasons. First, the assessments are perceived as being threatening and 'punitive', and were met with hostility by teacher unions ever since 2002 (Tiempo, 2003). Second, they load large responsibilities on the shoulders of school principals, who find themselves in the uncomfortable position of having to cast judgments that may carry heavy consequences for teachers; principals of public schools have very little incentives to take up the trouble of causing a teacher to be fired, facing potential institutional unrest and social or even legal backlash. As a consequence, a meticulous implementation of these two evaluation practices has been rare in the country, and the threat of dismissal following negative feedback remained idle (Arizmendi Domí nguez and Saa García, 2018; ECE, 2017; Echeverri Álvarez et al., 2017; Subsecretaria de Calidad Educativa, 2015; García Jaramillo et al., 2014; Sánchez Cotrino, 2014; personal conversations (2015) with the 2002–2010 Minister of Education).

The second continuous evaluation scheme introduced by the NR are the standard examinations that teachers must pass in order to qualify for a rank promotion — and thus obtain a salary boost. The examination, known as 'skill evaluation' *(Evaluación de Competencias*[8]), assesses teachers' pedagogical, disciplinary and behavioral skills; over the majority of our

5 ICFES (Instituto Colombiano para la Evaluación de la Educación) and CNSC (Comisión Nacional del Servicio Civil). The exam registration fee is below 9 USD (2012–2013 contest). The exam questions are elaborated by the National University, the largest public university in Colombia.

6 The 'aptitudes' module aims at assessing the candidate's ability to appropriately deal with language and numbers, and his knowledge of basic pedagogical concepts. The second module evaluates proficiency and skills of the candidate in his subject specialty. The psychometric test evaluates the candidate's hypothetical responses when facing pedagogical or institutional issues arising in the everyday teaching life.

7 Comisión Nacional del Servicio Civil (CNSC). It is an autonomous and independent body, located at the highest level in the structure of the Colombian State. It has legal identity and administrative, financial and managerial independence, and it is not part of any sector of the government authority. *(Description translated from the institutional webpage of CNSC, http://www.cnsc.gov.co/index.php/institucional/direccionamiento-estrategico/quienes-somos-cnsc; fetched on 19 Jan 2015).*

8 Regulated until 2015 by the Ministry of Education Decree 2715/2009. From 2015 onward, regulated by the Ministry of Education Decrees 1075 and 1657/2015, and Resolution 15711/2015.

period of observation the assessment took place through a written exam, however this was replaced by the evaluation of class recordings from 2015 onward. For NR teachers, passing the skill evaluation assessments represent an additional prerequisite for rank promotion, on top of having had positive yearly performance assessments and of the Old Regulation requirements[9]. This new promotion obstacle has been considered very challenging by NR teachers: the minimum passing score is very high and data shows that only about 20% of candidates to the exam were able to meet the threshold score between 2010 and 2015 (Elacqua et al., 2018; Gallego Henao and Vásquez Jaramillo, 2015; MEN, 2014; Cifuentes Cubillos, 2014). It is also worth noticing that skill evaluation exams started to be implemented only in 2010 — frustrating any expectation of salary increase up to that year. Finally, for many years, salary increases were limited by the scarce funding available to finance them (MEN, 2012a; Maussa Díaz, 2012). Altogether, over our period of observation the new continuous evaluation scheme tied to rank promotions has been far from smoothly implemented, and has created distance between theoretical pay scale trajectories of NR teachers and the ones achievable in reality.

Summarizing the new post-entry continuous evaluation practices in Colombia, it is hard to establish strong priors about the direction into which they may have contributed shaping teacher quality. On one hand, they have not been as strict as on paper about the negative consequences for low-performing teachers, particularly the threat of dismissal. On the other hand, meritocracy along the career path has proven blurrier than what the reform envisioned, due to institutional delays, financing constraints and the strong opposition by teacher unions. We will come back to these considerations in our final section.

3.3. Career progression and pay scale structure

The career- and corresponding pay scale structures for NR teachers are different with respect to OR ones, and the two are illustrated in Table A.5. The most important changes between the OR and NR can be summarized with making a postsecondary specialization the minimum education requirement for public teachers, thus increasing the minimum years of education from 11 (secondary education) to 13 (postsecondary specialization), and introducing a potentially much faster ascent in the pay scale — which we illustrate in Fig. A.1. The NR salary structure was planned so to make the public teacher profession more attractive, and especially so for highly educated and highly talented individuals — who would be able to enter at higher salary levels and progress quickly even further (Ome, 2013). In general, the question of whether financial incentives are an important determinant of public teacher quality has been addressed by several studies in various countries, with mixed conclusions.[10] In the reality of the Colombian case, as described in the previous subsection, ascending in the pay scale under the New Regulation has been generally perceived as a very arduous process with a low success rate. It is thus difficult to predict whether the appealing new compensation outlook depicted on paper is obtaining the desired effect of attracting and retaining more talented teachers in the public system; Saavedra et al. (2017) find that New Regulation teachers *narrowly* passing the entry exam threshold obtain a wage premium with respect to their similar non-teacher counterparts. The suggestive evidence we present in Section 7 does indicate that better educated individuals have been approaching the public teaching career under the New Regulation, and the more attractive salary prospects may have been playing a role in this trend.

3.4. Vacancy-filling process

Teaching vacancies arise when existing teachers retire or leave for personal reasons, when they ask and obtain to be transferred to another institution, or when the school expands its capacity. Schools report their vacancies to their local education authority (LEA), which publishes a list of vacancies to be filled either permanently or only temporarily (sick leaves, maternity leaves, leaves of absence).[11] The individual schools and school principals do not have any decisional power in the hiring or vacancy-filling processes (García Jaramillo et al., 2014). Vacancies are considered more attractive when belonging to schools whose geographic location or urban neighborhood is better, and whose students come from more advantaged socioeconomic backgrounds (García Jaramillo et al., 2014).

If a permanent vacancy remains empty, it will be occupied by a temporary teacher until a transfer eventually occurs, or until a novice teacher is assigned to the vacancy following a national hiring contest. Temporary vacancies are filled by temporary teachers until the permanent occupant returns. The rules concerning temporary teachers are rather lax, both in terms of minimum credentials required to enter the position and in terms of the candidate screening processes. The intention is to guarantee the presence of a teacher in the classroom even at short notice, in situations of emergency and in disadvantaged territories which lack fully accredited candidates; however, there are unwanted consequences stemming from this legal loophole, and we describe them in the next section.

9 Holding the appropriate education level, having attended the relevant teacher training courses, and having spent a given number of years of service in the previous rank (MEN, 2013b).

10 Among others, see Behrman et al. (2016) on Chile, Bau and Das (2017) on Pakistan, Hanushek et al. (2004) and Figlio (2001) on the US.

11 Permanent vacancies are published by LEAs twice each year, and teachers who already hold a permanent position in a different school may ask to be transferred into one of these. Teachers with longer tenures are prioritized in the choice of vacancies; the second priority criterion, for NR teachers only, is the national entry test score. All permanent vacancies which are open at the moment in which the national hiring contest is announced need to be reported by schools to their LEA, and published by LEAs for the purpose of the competition.

3.5. Student sorting

This paper does not address the aspect of student sorting across different schools and of how such sorting patterns may have been affected by the 2001 reform of public teacher careers.[12] We are not aware of any literature answering the question of whether the gradual inflow of New Regulation teachers has impacted the perception of school quality by students and their families. Even though there is no anecdotal evidence of public awareness on this matter being high or on large student shifts to have occurred, it cannot be excluded a priori that student sorting might be taking place as a consequence of the changing distribution of NR versus OR teachers across different schools. In an exploratory exercise, we look at how the proportion of NR teachers relates to the number of students sitting the Saber11 examination across the different public schools and years in our sample. The first two columns of Table A.6 show school-fixed effects estimations, that is, in these specifications we exploit the variation in NR teachers *within the same school* across the different years in our sample. We find that, conditional on other teacher characteristics, higher proportions of NR teachers tend to correlate with higher numbers of students sitting the final examination, both when we estimate a single linear relationship (Column 1) and when we allow for more flexibility by dividing the shares of NR teachers in deciles every year and estimating a separate relationship for each decile (Column 2). Both strategies point to the conclusion that, for a given public school, moving from the lower to the upper part of the NR teacher share distribution over the years increases the average number of test takers by between 3 and 4.5 students, which corresponds to 4%–6% of the mean or 6%–9% of the median candidate cohort. The third and fourth column of Table A.6 show results of a different approach exploiting *across-school variation*, namely of measuring how the growth in the candidate cohort between two fixed periods in time relates to the proportion of NR teachers observed in different schools. In Column 3, we used the growth in test takers between the pre-reform year 2001 and the first post-reform year we observe, 2008, while in Column 4 we used the growth between the first and last year of New Regulation we observe, 2008 and 2016. Again we find a tendency of higher growth rates in test taker cohorts for schools in which the share of NR teachers is larger, although the precision of estimates is not high. Figs. A.2 and A.3 illustrate the results of Columns 2, 3 and 4, each time taking the second decile of the NR distribution as the baseline and excluding deciles 1 and 10 for the sake of clarity[13]. Our tentative analysis on student sorting suggests that the presence of NR teachers may be noticed and perceived positively by the student population and their families, so to cause a gradual inflow of students towards schools where such teachers are concentrated. However, a more sophisticated analysis is needed in order to determine the degree of causality in this process and to understand the origin of the student flows; the absence of teacher data on private schools prevents us from carrying out a more careful cross-sector analysis. In fact, this represents a promising avenue for future research on the Colombian education reform experience.

In any case, it is important to notice that none of the results presented in this paper depends upon the assumptions of stable quality in schools' student bodies or absence of sorting of students across schools. As will be further illustrated in Section 5, our estimation strategy based on exploiting variation in student outcomes within a given school and in a given year allows for arbitrary dynamics in student sorting to be occurring over time.

4. Data

Our data on student outcomes is drawn from the Colombian high school exam known as Saber11, which is sat by students at completion of secondary education.[14] Saber11 test data is collected by ICFES[15] and it is the most complete among the standardized tests being conducted at different school grades; it is widely accepted as the reference examination to evaluate the quality of Colombian secondary education. Individual Saber11 scores are recorded separately by subject and range from 0 to 100; they are standardized by subject at the national level, so that each student's score is informative about her position relative to the national average. We use 9 years of test scores, 2008 to 2016, on each of the 5 subject areas which students are required to take: Mathematics, Natural Sciences, Critical Reading, Social Sciences and English language.

Data on teachers is available thanks to a yearly information reporting procedure enforced by the Ministry of Education: public schools are required to give details on their pupils, staff and infrastructure through the standardized formats set by the Ministry (*Resolución MEN 166 /2003* and following updates). However, teacher census data is available only from 2008 onward, which causes the observation window of our panel to start that year. We employ individual data on teachers engaged at any public high school in the country for the years 2008 to 2016 in a subject belonging to the 5 areas previously mentioned. Each record includes the teacher's demographics, education level, school of placement, teaching subject, type

[12] Thank you to an anonymous reviewer for suggesting clarification on this point.

[13] Deciles 1 and 10 are particularly noisy in all estimations and are affected by small or new schools, in which the share of NR teachers takes the extreme values 0 or 1.

[14] That is, after 11 years of schooling. Schooling years 10 and 11 are not compulsory and are attended by around 41% of the eligible school population (2012 data: Sistema Nacional de Indicadores - Tasa de cobertura neta - Ministerio de Educación Nacional - http://menweb.mineducacion.gov.co).

[15] Instituto Colombiano para la Evaluación de la Educación. The same agency also administers the national tests taken by students at different stages of their career. It is a governmental agency with social scope within the sector of public education; a national, public entity with administrative independence, bound to the Ministry of Education. *(Description summarized and translated from the institutional webpage of ICFES, http://www.icfes.gov.co/informacion-institucional/marco-legal; fetched on 19 Jan 2015).*

of contract, date first hired in the public education system, and salary level. We excluded teachers who are recorded as not being active in an educational structure in any given year, because of leaves of absence or other personal reasons. We are left with a sample of more than 151,800 teachers, among which about 76,200 belong to the Old Regulation and about 75,600 to the New Regulation. Descriptives at the individual level are illustrated in Table A.7. Using their unique national ID document, we were able to match 91.9% of all New Regulation high school teachers to their entry exam scores.

From this exercise we drew an additional novel contribution to the literature on the Colombian education system. We are the first to document the fact that over 30% of all NR teachers holding a public teaching position over the period 2008–2016 have not been successful in any of the entry contests. As described in the previous section, the law allows individuals who did not pass the entry exam to fill teaching vacancies temporarily, and only in the absence of qualified candidates willing to fill the positions, in order to avoid teacher shortages especially in disadvantaged areas of the country. Table A.8 shows that, as expected, the vast majority of NC teachers occupies temporary teaching positions (94.4%), a small fraction is recorded as being on probation (1.6%) and the remaining as occupying a permanent position (4%).[16] Some of our analysis will show that student performance associated with these teachers is significantly lower with respect to their certified colleagues.

Finally, the match between students and teachers works as follows. Given we do not have information on class groups, we are not able to achieve an exact match between teachers and the students actually taught. However, we construct matches at the school, subject and year level: that is, between teachers teaching at secondary school i, year t and subject s, and the Saber11 student test scores in the same school i, year t and subject s. The 207,500 units of observation in our main sample are *[school, year, subject]* tuples: we collapse our individual student and teacher data to obtain averages at that level. Our outcome variable is "Average test score" in the *[school, year, subject]* and our main regressor of interest is "Share of New Regulation teachers" teaching in that *[school, year, subject]*, which is given by the ratio of teachers hired under the NR regime to total teachers in the observation unit. Descriptives at this level of observation are illustrated in Table A.9.

5. Empirical strategy

Our goal is to assess the impact that introduction of the New Regulation of the teacher career has had on student performance, and there are two main challenges to this goal. The first is to obtain a rigorous estimate of the performance premium of NR teachers with respect to their OR colleagues, which is a far from trivial task, given that various kinds of selection affect the distribution of teachers in the country. The second challenge is to convincingly exclude the possibility that the performance premium of NR teachers may be due to experience or cohort effects, since NR teachers are systematically younger and hired later with respect to their OR colleagues: from a policy perspective, we care to determine whether the change in *rules* yielded any change in performance.

In the following paragraphs, we will describe how our main empirical specification is suited to settle the first task of identifying the NR premium accurately. The basic relationship we focus on is between student test scores y and the share of New Regulation teachers teaching those students. If NR teachers are on average more effective than OR ones, we expect the relationship to be positive and significant; on the other hand, if NR teachers are less effective than their OR colleagues we shall estimate a negative relationship, and if the two groups are on average similar in their effect on student performance, we expect no relationship between the two variables. However, one must account for the potential selection of teachers across schools and across time, which would result in a non-random match between NR teachers and students; García Jaramillo et al. (2014) present evidence on more senior and better-scoring teachers being more concentrated in institutions with better locations and more advantaged students. We overcome this hurdle by leveraging the detailed administrative data at our disposal and operating at the *within school, within year* level. In every school and every academic year, the *same* students are taught and tested on different subject areas, and each area has a different share on NR teachers active in it: the across-subjects heterogeneity in performance an in NR teacher share is the key to our identification. More formally, the model we estimate has the following specification:

$$y_{its} = \beta_0 + \beta_1 NR_{its} + \sum_{k=1}^{K} \beta_k X_{kits} + \boldsymbol{\alpha_{it}} + \boldsymbol{\alpha_s} + e_{its} \tag{1}$$

where the unit of observation is a school(i)-year(t)-subject(s); y_{its} is the average student test score; NR_{its} is the share of New Regulation teachers in each unit of observation; X_{kits} are K unit-level controls including the average level of education, age and experience of the teacher group; α_{it} is a vector of school–year fixed effects; α_s is a vector of national-level subject fixed effects; e_{its} is the residual error term. This specification enables us to rule out any school-level characteristics, of both time-invariant and time-varying nature, which induce selection of teachers into schools or cause correlation between the share of NR teachers and student scores.[17]

[16] The non-zero percentages on the permanent and probation categories are explained by further exceptions granted to teachers in the most rural areas and by those teaching specific technical subjects.

[17] Similar models not featuring school fixed effects are exposed to bias, due to non-random selection of NR teachers into schools. Those featuring school- but not school–year fixed effects (as in Ome (2012)) are able to account for time-invariant sources of bias but are still exposed to selection bias arising from time-varying factors or shocks, such as changes in school management or in school resources.

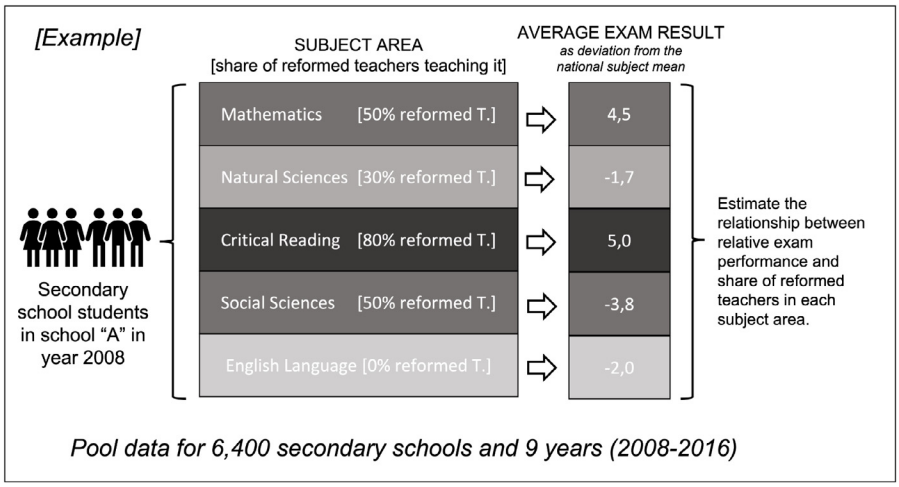

Fig. 1. Illustration of the empirical strategy.

Fig. 1 illustrates the core intuition behind our empirical strategy. Imagine we focus on a specific secondary school "A" in the school year "2008". All students who graduate that year will sit the Saber11 examination, which is composed of five different subject modules. It is important to notice that all students in that cohort must take all the subjects: another way to see it is that each subject is taken by the same group of students. Now, in school "A" there will be some difference in the performance of these test-takers in the subject of Mathematics compared to Natural Sciences, where the national average performance in Math and Sciences are the respective baselines. Further, in that same school "A" and year "2008" the subject of Mathematics is taught by a group of teachers that are in part New Regulation and in part Old Regulation (say, 50%–50%), while Natural Sciences sees a different share of New versus Old Regulation teachers (say, 30%–70%). In our econometric model, we estimate how much of the variation in student performance across subjects is explained by the variation in New Regulation teachers across subjects. Using the remaining information in our teacher dataset, we are able to establish separate relationships between student performance and other characteristics of the teacher team, so that these do not confound the effect of Regulation type. Simply put, we pool the results of within-school-and-year performance comparisons carried out in around 6,400 secondary schools and across 9 different years. Exploiting this very large sample allows our statistical model to accurately identify systematic associations between student performance and the various characteristics of the teaching teams, such as regulation type, age and experience. Importantly, we also extend these characteristics to include the time period in which teachers first entered the profession (which is different from the length of their experience), in order to partial out the effect of pertaining to specific "hiring cohorts" or "teacher generations". Further, the strategy of comparing performance across subjects, within single schools and school years also enables us to exclude the influence of the characteristics of the school itself, such as its location, its student body, its principal; of each subject area at the national aggregate level, such as the difficulty of the national curriculum, the national average quality of teachers in that subject, and so on; any effects that are specific to the year "2008", such as particularly difficult or simple exams set by ICFES, the effect of economic downturns, and so on.

The main assumption behind our statistical analysis is that individual schools are not able to manipulate the shares of NR versus OR teachers within the same year, across different subjects. In previous sections we have described how the institutional characteristics allow for a certain degree of performance-related selection of teachers into schools[18] but leave no room for such selection to happen within the same school and year. Fundamentally, at this level of data detail, we are facing a quasi-experimental setting in which the variation in the proportions of NR teachers across different subjects is unrelated to previous student performance or to subject-specific performance trends. In practice, the different shares in New Regulation teachers present in each school, year and subject are the result of the retirement patterns that each school witnessed over time.

Beyond the descriptive evidence drawn from the institutional setting, we also back the assumption of no selection within schools and years empirically, in Appendix C, after illustrating the results of the Colombian reform. In the same section, we also address any residual concern about differences between OR and NR teachers that are not being accounted for by Model (1), especially along the time dimension. Our main model includes school–year fixed effects, subject-specific time trends, and controls for the age and experience of teachers in the shape of 10-year and 5-year bins respectively. These features go a long way towards dealing with the main differences between OR and NR colleagues; however, given the relevance of this point from a policy perspective, we present an extended battery of robustness checks and placebo tests targeted specifically at ruling out experience or cohort effects as the drivers of our results.

[18] Namely the fact that higher-scoring NR teachers choose their vacancies first, as well as the voluntary transfers mechanisms prioritizing more tenured teachers and teachers with higher entry test scores.

Table 1
The effect of New Regulation teachers on student performance.

	Simple	School selection	School and time selection		
	(1)	(2)	(3)	(4)	(5)
New Regulation	−1.25***	−0.10**	0.05***	0.07***	0.07***
	(0.09)	(0.05)	(0.01)	(0.02)	(0.02)
Age and experience				Yes	Yes
Postgrad education				Yes	Yes
School FE		Yes	Yes	Yes	Yes
Year FE		Yes	Yes	Yes	Yes
School–year FE			Yes	Yes	Yes
Subject FE			Yes	Yes	Yes
Subject-specific trends					Yes
N.obs	6,471	47,481	207,545	207,545	207,545
N.groups	.	6,471	47,481	47,481	47,481
R-squared	0.03	0.86	0.89	0.89	0.89

Note: In column (1), each observation is one school 'i': selection of teachers across schools and time is not accounted for. In column (2), each observation is one school 'i' in year 'y': selection of teachers across schools is accounted for, but selection within schools over time still affects estimates. In columns (3) to (5), each observation is one subject 's' in school 'i' in year 'y': these specifications correct for both selection across schools and within schools over time. Age and experience controls are 10-year (age) and 5-year (experience) bins. SE clustered by school in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6. The impact of the Colombian reform on student achievement

Table 1 shows our main results, expressed in within-school-and-year standard deviations. We identify a positive and significant effect of New Regulation teachers on student test scores: when students switch from being taught by OR teachers to NR teachers in a specific subject, on average their performance improves by 7% of a standard deviation compared to their performance in other subjects that same year in that same school.

By gradually introducing fixed effects and controls, Table 1 allows us to appreciate the effects of non-random teacher selection. Results in Column (1) do not account for any type of selection: not surprisingly, there is a strong and negative correlation between NR teachers and student scores. On average, NR teachers tend to obtain positions in lower-quality schools with respect to their long-tenured OR colleagues.[19] Results in Column (2) correct for selection of teachers across schools, which we find accounting for the largest part of the overall negative correlation between NR teachers and student test scores, but does not correct for selection within schools over time. Our estimates indicate that, with respect to their OR colleagues, NR teachers tend to select into and *remain* in schools whose performance is not only on average lower at any point in time, but which also feature worse development paths.[20] Finally, Columns (3) to (5) reflect our full model, which is able to correct for both selection across schools and within schools over time, by exploiting year-specific variation across subjects. We estimate the relationship between student performance and NR teachers to be positive and statistically significant, at a magnitude of about 5% of a standard deviation before accounting for their age, experience and level of education (Column 3). Omitting the individual characteristics just mentioned is likely to lead to an underestimation of NR teachers' performance, given that these are on average younger and less experienced than their OR colleagues, and had less time to acquire higher education degrees: we expect these factors to weaken the true positive impact that NR teachers have on student performance. In fact, once these personal characteristics are taken into account through control variables, we obtain our preferred estimate of a 7% of a standard deviation impact (Column 4), which is robust to the inclusion of subject-specific time trends for each school (Column 5).

The magnitude of the impact of NR teachers with respect to OR ones is quite large. It is worth reminding that our analysis is measuring the *average* performance difference between teachers pertaining to the two different career regulations, and that it is generally hard to spot objective personal attributes of teachers that robustly predict their performance. We can relate our estimates to the least controversial observable attribute of teacher effectiveness, experience: the NR effect we identify compares to the difference in being taught by teachers with more than 5 years of experience with respect to first-time teachers (Hanushek and Haycock, 2010). Another useful point of reference are the most rigorous estimates on individual teacher quality distributions on the US context (see Hanushek and Rivkin (2012)

19 Due to prioritization of tenured teachers in the vacancy selection process — see Section 3.4.

20 The analyses by Ome (2013, 2012) do not account for this type of selection, which we find to be important and going against the true positive relationship between NR teachers and student performance. We believe this may explain his null finding, along with the shorter observation period and thus lower testing power he is able to exploit. This type of over-time selection also affects estimations at the school–subject level, in which the key variation is in the share of NR teachers in a specific school and subject, across different years. Other issues affecting estimations exploiting over-time variation in this context are the fact that they compare outcomes of different and heterogeneous cohorts of students – leading to potential bias – and that the variation in NR teacher shares over time is not particularly high — leading to underpowered results.

for a review). The average student achievement premium stemming from NR teachers in Colombia is about half as large as the effect of being taught by a teacher at the 75th percentile rather than at the 25th percentile of the quality distribution in the US.

7. Mechanisms

In the following subsections we discuss some evidence on the mechanisms behind the main results. Two main scenarios come to mind when thinking of the potential drivers of the better-quality teaching under the new career rules: in the first scenario the pool of candidates to the teaching profession did not change as a consequence of the reform, but the new screening procedures, especially at entry, achieve a selection of higher quality teachers with respect to the previous setting. In the second scenario, the new regulation caused the type of candidates to the teacher profession to change in first place. Although we are not able to provide a completely satisfying answer to these questions due to data limitations, we proceed to illustrating some statistical evidence implying that both scenarios are likely to coexist and jointly explain the final result.

7.1. Quality screening at entry

The public entry competition, and especially its written exam component, is certainly the most prominent among the novelties that were introduced by the 2002 reform of the teacher career. The purpose of the exam is allowing the most capable candidates to proceed in the selection process, while keeping the less desirable ones out of the profession. However, it is well known that designing effective teacher candidate assessment methods is hard, and recent research on Mexico (Estrada, 2019) and Ecuador (Araujo, 2019) has shown that national entry exams may not be up to the job of selecting the best candidates. The New Regulation in Colombia might have induced a performance gain with respect to the Old Regulation even in the event of a 'bad' entry examination, simply by changing the *perception* of the difficulty and the competitiveness of the teaching career, and by making access to the job more standardized and less corrupt — which is what appears to have happened in Mexico. The question of whether the Colombian entry exam is actually able to select the better candidates is therefore a relevant to answer, if we wish to shed further light on the mechanisms at work behind the success experience of this country.

First, we retrieved data on the exam success rates on the first five entry contests, and show these in Table A.10. It is evident that the minimum score threshold required to pass (60 points over 100) represents a tough hurdle for many candidates, and success rates are as low as 28% on average.

Nevertheless, a low passing rate does not necessarily imply that the entry exam is actually performing a selection based on candidates' quality, i.e. that it is effective in picking out teachers who 'produce' better student performance later in time. We thus investigate the relationship between NR teacher entry exam scores and student test scores, replicating the same model specifications that produced our main results in Table 1. Table 2 shows the results of regressing student performance in each subject on the average exam score[21] of its associated NR teachers; the sample of analysis includes all school–year–subject observations in which there is a non-zero share of NR teachers.[22]

The results show a positive and significant relationship between teachers' performance in the entry exam and their students' performance in the high school examination. First, as we did for the main results, by comparing the more naïve specifications in Columns (1) and (2) to our full model in Columns (3) to (6) we learn about teacher sorting across schools and within schools across years. There is a significant selection of better candidates into better performing schools, explaining the strong relationship identified in Column 1; further, we confirm school-level sorting over time, with higher skilled teachers holding onto schools with a better over-time development record (Column 2). Once we dig into the within-school-and-year level (Columns 3–6), the relationship between teacher entry exam score and student performance stabilizes at a coefficient of 0.04: that is, being taught by teachers whose average entry exam score is 1 standard deviation higher in the national distribution increases students' subject performance by 4 percent of a standard deviation, within their school and year. The magnitude of the associations we find is very well aligned with recent findings in US studies, which place the effect of a one standard deviation increase in teachers' exam test scores at between 1 and 7 percent of a standard deviation in student performance (Goldhaber et al., 2017a,b; Cowan and Goldhaber, 2016; Goldhaber and Anthony, 2007).

In order to verify whether the relationship between teacher scores and student scores is linear across the whole score spectrum, we also show results obtained by dividing teacher performance into score deciles. Results are shown graphically in Fig. 2 and numerically in Table A.11 in the Appendix. We find that higher entry score deciles consistently lead to better student performance, in a relationship that appears almost linear even when not constrained to be so.

In sum, the Colombian national entry examination is indeed performing a positive selection on candidates' ability; we find it is as effective as its US counterparts such as the edTPA and the NBPTS in choosing better candidates for the

[21] For each teacher we use his/her most recent test score, as the most up-to-date measure of skills. Like student scores, teacher scores are expressed in standard deviations of their national distribution.

[22] In Appendix B we discuss how unobserved quality of Old Regulation teachers may affect this relationship and present the corresponding robustness check.

Table 2
The effect of teachers' entry exam scores on student test scores.

	Simple	School selection	School and time selection			
	(1)	(2)	(3)	(4)	(5)	(6)
Teachers' score	0.84***	0.08***	0.04***	0.04***	0.04***	0.04***
	(0.05)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
New Regulation						0.01
						(0.03)
Age and experience				Yes	Yes	Yes
Postgrad education				Yes	Yes	Yes
School FE	No	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes	Yes	Yes
School–year FE	No	No	Yes	Yes	Yes	Yes
Subject FE	No	No	Yes	Yes	Yes	Yes
Subject-specific trends	No	No	No	No	Yes	Yes
N.obs	6,415	45,734	164,226	164,226	164,226	164,226
N.groups	.	6,412	45,733	45,733	45,733	45,733
R-squared	0.07	0.87	0.90	0.90	0.90	0.90

Note: In column (1), each observation is one school 'i'. In column (2), each observation is one school 'i' in year 'y'. In columns (3) to (6), each observation is one subject 's' in school 'i' in year 'y'. The sample includes only school–year–subject observations in which there is a non-zero share of NR teachers. Age and experience controls are 10-year (age) and 5-year (experience) bins. SEs clustered by school in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

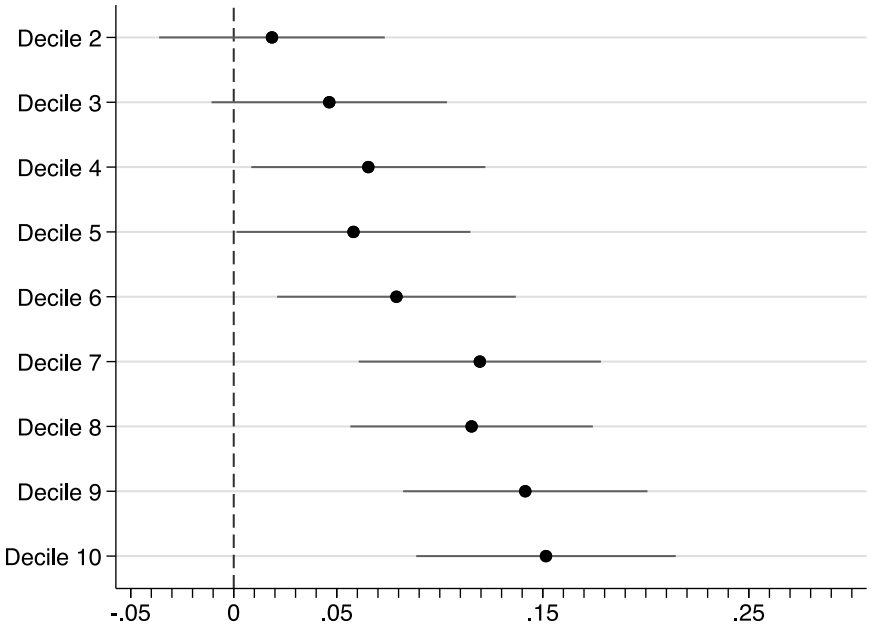

Fig. 2. The effect of teacher entry exam scores on student performance. *Note:* The effect of average teacher entry exam scores on student performance, higher average score deciles relative to the first. Point estimates and 95% confidence intervals with Bonferroni correction for multiple hypothesis testing. The corresponding numerical results are shown in Table A.11 in the Appendix.

teaching profession.[23] Higher scores on the entry exam consistently translate into better student performance at high school, so that the entry selection mechanism introduced by the 2002 Colombian reform is producing the desired result of prioritizing better-deserving candidates when assigning teaching vacancies across public schools in the country.

[23] In Appendix B we address the issue that due to observing only entry test scores of NR teachers, our estimates of the true relationship between teacher skills and student performance might be biased. We show that the direction of the bias is actually unclear and depends on the correlations between the skills of NR and OR teachers. We re-estimate the relationship between teacher scores and student performance on a subsample of observations in which *all* teachers are NR: we find a slightly stronger relationship, moving the slope of interest from 4 percent to 5 percent of a standard deviation.

This validation becomes all the more relevant when we recall to mind our finding that around 30% of New Regulation teachers were engaged in public schools during the 2008–2016 period without having actually passed the entry exam, exploiting the legal loophole of temporary positions (see Section 4). Given our results on the effectiveness of the exam in selecting the better candidates, we would expect a difference in the performance premium of 'certified' NR teachers, who were able to pass the exam, and the one of 'non-certified' NR teachers, who were not. We verify this hypothesis in Table A.12, where we run our main specifications after separating certified from non-certified teachers. Indeed, the results reveal that the performance premium with respect to OR teachers is concentrated on certified NR teachers, and is small and insignificant for non-certified NR teachers.[24] In fact, this finding can be viewed as an additional cross-check on the validity of the Colombian entry exam as a selection instrument implementing meritocracy. The exam performance of non-certified teachers is below par, and indeed, on average these individuals are not able to produce good student performance any better than their loosely-selected Old Regulation colleagues.

7.2. The quality of candidates

Besides selecting the more able candidates through standardized competition, by changing rules and incentives of the teacher career the New Regulation is likely to have changed the quality of the pool of teacher candidates. As described in Section 3, tougher initial selection, higher quality requirements along the career path and test-based but potentially faster increases in salary are meant to discourage unproductive applicants from pursuing the public teacher career. Both anecdotal evidence and previous research Ome (2013) suggest that the new teacher career is indeed attracting higher proportions of specialized graduates and skilled professionals; on the contrary, the typical Old Regulation teacher profile is a graduate majoring in Education sciences.[25]

We would like to verify whether these patterns are visible in the data we have at hand, taking acquired education as a proxy for quality and marketable skills. Unfortunately, it is not possible for us to directly compare *candidates* to the OR teacher career and *candidates* to the NR teacher career, given that data on the former group does not exist, and that background information on the second group is limited to basic demographics. As a second-best exercise, we compare the education level of OR and NR teachers. Given the common practice of acquiring further education degrees over the course of the teaching career, we limit our comparison to teachers below the age of thirty-five, both in order to partially compensate for the average age differences between the two groups, and to get closer to *candidates*' profiles. Our education descriptives are plotted in Fig. 3. Indeed we notice a significantly higher share of university degrees among NR teachers (93.6%) in comparison to their age-comparable OR colleagues (85.1%).

Apart from the shift away from secondary and upper-secondary degrees into university degrees, a noticeable fact is the distinctly higher shares of *specialized* education degrees among NR teachers — that is degrees majoring in subjects other than Education, which typically attract better high school graduates[26]: these are held by 22.5% of NR teachers and only 13.4% of OR teachers. While these descriptive results are in line with the type of change in applicants pool postulated by previous authors, it is good to keep in mind that they also incorporate the outcome of the two different entry selection processes, and the early career attrition patterns that affect each of the two groups.

In sum, we are able to detect better education achievements among junior teachers of the NR, both in terms of average number of years dedicated to study and in terms of a shift from the less prestigious education majors to subject-specific ones. It is reasonable to presume that these changes are at least in part to be ascribed to a change in the skills profile of people attracted to the career of teaching in Colombian public schools, in line with the objectives of the reform and with the earlier anecdotal evidence on this matter.

8. Concluding remarks

The good news is that the process of turning around teaching quality in Latin America has unequivocally started, under the pressure of international institutions, economic advancement and higher awareness among local communities. Despite continued resistance by unions and local politicians, the traditional pacesetters in the public teaching sector, winds of change have come to stay and have started an unprecedented wave of reform initiatives across the region, based on the principles of rewarding merit, continuously evaluating and implementing continuous professional development. Assessing the effectiveness of such revolutions through empirical research is necessary but highly challenging, as most countries' experiences do not lend themselves to rigorous impact evaluation. The interference of contemporaneous

[24] The results of Columns (1) and (2) reveal that the selection of non-certified teachers into schools and across time is even more negative than for Cert New Regulation teachers. Non-certified teachers sort into worse schools and into schools on worse over-time performance patterns. Non-certified teachers are also younger, significantly less experienced and less educated than certified teachers (see Descriptives in Table A.7), which explains the negative point estimate persisting in Column (3), which shows results at the school–year level but does not account for these important characteristics.

[25] On average, students choosing Education majors at university had scored significantly lower on high school exams with respect to their peers majoring in other college specializations. *Source*: MEN - SPADIES information system; García Jaramillo et al. (2014).

[26] We take education level and Saber 11 scores as the best proxies for quality. We skip the analysis on average *college* graduation marks, which is another potential dimension of heterogeneity. Additionally, we are abstracting from the discussion of how much pedagogical value added a teacher candidate owns for having majored in Education sciences rather than in a specialized field. Also see footnote 25.

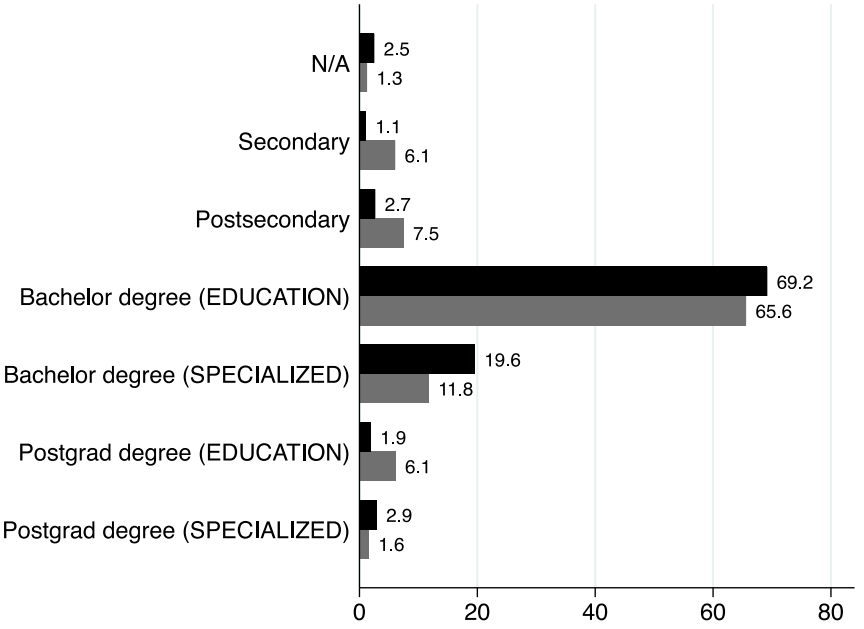

Fig. 3. Teachers' education levels. *Note:* Highest education level obtained by the end of observation in our panel, by NR (black) and OR (gray) teachers. University degrees are separated by majors in Education versus subject-specific majors. The sample considered is individuals below 35 years of age by the end of observation in our panel.

institutional changes, coupled with the often-slow implementation of innovations and the scarcity of high-quality data, poses difficulties to the policy analyst that are in many instances hard to overcome. This is why studying an experience such as the Colombian one, which happened to be particularly clean-cut in its implementation, was followed through high-quality data collection processes and whose effects have had more than a decade to develop, is extremely valuable.

In this paper we studied the effects of introducing merit-based selection and incentives to the public teaching career in Colombia, through a reform which is innovative on the Latin American continent both in terms of contents and reach and in a context characterized by an approximately equal proportion of "traditional" and "reformed" teachers. Leveraging the accuracy of our data and exploiting the differences in teacher type distribution across subject areas, we estimate that being taught by "reformed" rather than "traditional" teachers improves Colombian student achievement considerably, in a magnitude comparable to being taught by experienced versus novice teachers. Our findings thus speak to a clear success of this initiative. We proved the existence of a causal relationship between the new merit-based regulation and the improvement in teaching quality across the country, and we are able to exclude that the performance premium is due to either selection of teachers into schools or the differences in age, experience and hiring cohort between the two groups of educators.

Our results indicate that candidate selection at entry through the new standardized testing procedure has been the most effective measure implemented, and, importantly, we show that the Colombian entry test is indeed a good predictor of future teacher performance. The degree of correlation between performance on the entry test and teaching effectiveness is in line with similar findings for the US (Goldhaber et al., 2017a; Cowan and Goldhaber, 2016; Goldhaber and Anthony, 2007). This is the first positive result of this kind on the Latin American context, in contrast to the negative one reported for Mexico (Estrada, 2019), and represents an encouraging addition to the literature on teacher selection procedures in this region. More in general, our analysis of the Colombian experience may offer insight to education policymakers in neighboring countries such as Argentina, Bolivia, Brazil, Venezuela, Nicaragua, El Salvador and Panama — which to date are characterized by public teacher career regulations that very much resemble the pre-reform Colombian one.

It is also appropriate to dedicate some final thoughts to a less unambiguous aspect of the Colombian reform experience. In our description of the institutional setting in Section 3.2 we had noticed how the implementation of merit-based selection *after* the start of the teaching career has been far less clear-cut than the one enforced at entry. The opposition of teacher unions, the perception of "punitive" intentions in the new rules, and some degree of short-sightedness about their practical implementation were all factors contributing to creating a significant distance between the new career circumstances 'on paper' and the 'de facto'. Motivated by the uncertainty surrounding the after-entry dynamics, in Appendix D we develop and discuss a set of exercises that focus on dropout rates of teachers after they have reached a permanent position. We find that in general, dropout patterns are quite similar between the Old and the New Regulation, suggesting that in-career stability has not been particularly affected by the new rules. However there is a notable exception to this trend, and it is represented by the highest-skilled new regulation teachers, who leave the public teaching career

at a particularly fast pace. In absence of data on the *reasons* for career dropout or on the destination of the individuals who leave the profession, it is hard to deliver precise policy guidance on this point, but an objective appraisal appears to be that the current merit-based incentive package is still too weak to retain top performers in the public teaching system for longer periods.

There is extensive scope for additional research on the topic. Among other avenues, it would be interesting to investigate to what extent this reform has been changing society's perception of the public teaching profession, as well as the profile of the typical high-school graduate selecting into this career. García Jaramillo et al. (2014) points our that in Latin America, students preparing for the public teaching profession tend to come from the lower end of the performance distribution. The introduction of meritocracy to the teaching career may have the power to bring about a positive change in this tendency, with high-impact and long-term repercussions on the whole education system. If such a change in perceptions is currently ongoing in Colombia, we would expect the effect of the new career regulation to be amplified over time, through the intake of progressively higher-quality applicants to the public teaching profession.

Appendix A. Evidence on teacher quality reforms in Latin America

Chile

Chile is a Latin American frontrunner in the quest for teaching quality. Well-structured and carefully planned teacher evaluation protocols have been gradually introduced since 2004, in the so-called "Framework for Good Teaching" (*Marco para la Buena Enseñanza*); targeted initially at municipal teachers only, about half of the total, it was gradually extended to the remaining categories too.[27] These protocols were the results of a decade of negotiations between teacher unions and the government, so that starting in the mid 2000s, the continuous evaluation culture started to become the accepted standard in the educational community (Avalos-Bevan, 2018). The objective of the continuous evaluation processes in this country can be summarized in three pillars: providing formative feedback for teachers; acting as a prerequisite for salary upgrades and other monetary incentives; serving as a justification for dismissal in situations beyond hope of improvement. Manzi et al. (2011) provide a comprehensive study of the Chilean public teacher evaluation system and report generally encouraging results, as measured by scores received by participating educators (Cortés and Lagos, 2011). However, two particularities of the Chilean setting should be taken into account: first, as the authors point out, composition of their samples varies substantially over the years, due to the gradual implementation of the evaluation protocols in the country, making it hard to compare one year to the next and appreciate the net effect of evaluations on teacher performance. Second, the Chilean teacher evaluation system is highly decentralized and in the hands of municipal committees: while local management has advantages in terms of suiting each community's particular needs, it poses challenges in terms of comparability of results across the nation and makes quality achievements heavily dependent on the professional capacity of each local education authority, as well as on financial support by local politicians (Cortés and Lagos, 2011).

Starting in 2016, Chile has further stepped up its efforts towards improving teacher quality by gradually introducing the 'System for Teacher Professional Development' (*Sistema de Desarrollo Profesional Docente*), whose full implementation is expected by 2026. This is a comprehensive policy with significant implications for the management of teaching workforce (OECD, 2017) and its main innovations with respect to the previous status quo are the strengthening of initial teacher preparation, from tougher selection into education programs to higher grade requirements to access the career; a reformed career ladder structure accompanied by improved financial compensation packages; additional importance given to training and continuous professional development.

While the Chilean experience on teacher quality reforms compares or even exceeds the Colombian one in terms of efforts exerted and resources dedicated, the institutional settings in the two countries differ on several aspects, so that an understanding of the Colombian setting reserves a large widening in perspective with respect to what we learned from Chile. Contrary to Chiles highly decentralized management, public education in Colombia is fully centralized, so that teacher policies and evaluation protocols are identical across the nation — which, along with its up- and downsides, means a significant addition in terms of policy lessons to be acquired. Further, the Colombian reform we study in this article was implemented sharply in 2002 at a national level and imposed upon the whole public teaching workforce hired from that year onward — in contrast to the Chilean gradual rollout, staggered across time, space and teacher categories: again, an entirely different policy implementation scheme of which to explore successes and failures. Finally, one additional ingredient distinguishes the Colombian case from the Chilean, namely the employment of nationally standardized entry exams as the main teacher candidate selection tool. Beyond representing an additional and less explored tool for teacher quality improvement in the Latin American context, the availability of standardized entry scores endows us researchers with a systematic and nationally comparable measure of candidate quality, that we use to perform solid analysis on policy impact.

[27] In 2014, the evaluation protocols had covered 83% of municipal teachers (OECD, 2017).

Mexico

Mexico gave the strongest impulse to its teacher quality policies in the late 2000s, when a whole set of new teacher evaluation instruments were introduced to the institutional framework, under the policy umbrella known as 'Alliance for Educational Quality' (*Alianza por la Calidad de la Educación*) (Guzmán Marí n, 2018). Among others, these instruments included national, exam-based competitions for teacher hiring (2008); recognition and financial incentives for high-quality teaching (2008) and continuous professional development plans for teachers (2009). However, it wasn't until a new comprehensive educational reform passed in 2013 that many of these innovations, and most notoriously the hiring through standardized testing, became compulsory and fully applied on a nationwide basis. The difficult institutional process faced by these teacher quality incentives was mainly due to heavy opposition from the national teacher union, which had held an extremely influential position in teacher hiring decisions up to that point and which was not keen on giving that up (Tatto et al., 2006; Estrada, 2019).

In fact, giving a full evaluation of the consequences of the Mexican reform experiences has revealed as a particularly challenging endeavor, mainly due to the continued political pressures surrounding the public teaching profession and the subsequent reluctancy to develop a comprehensive information system about teacher quality indicators (Tatto and Velez, 2019). The perhaps only rigorous policy evaluation study available is the one by Estrada (2019), who collected data on a sample of specific community-level schools and finds large positive impacts of teachers hired through the new standardized entry exams on student achievement. Interestingly though, the author does not find the performance of teacher candidates on the Mexican entry exam to have any predictive power on their value added in terms of student achievement. The results we will present later in this article, on Colombia, reap the benefits of the well-established, nationwide data collection processes available in the country, which offered us with the opportunity to perform policy impact analysis on the universe of public secondary schools. Moreover, our results re-open the debate on the ability of teacher test scores to predict their classroom performance (for reviews see Wayne and Youngs (2003), Hanushek and Rivkin (2006), Glewwe et al. (2013)).

Ecuador

Major changes to the public teaching profession, specifically targeting teacher quality, were introduced in Ecuador starting in 2006, following the recommendations of several national and international policy reports on the topic. The new introductions were jointly incorporated to the new comprehensive educational law and thus made effective as of 2011 (Chiriboga Montalvo and Pinto Haro, 2019). The newly introduced changes can be summarized in stricter quality-assessment of the institutions offering teacher education; tougher entry requirements and competition for teaching vacancies based on standardized testing; implementation of continuous evaluation protocols; merit-based career progression; improvement of the financial compensation for all teachers.

A recent UNESCO-sponsored case study by Chiriboga Montalvo and Pinto Haro (2019) highlights several difficulties encountered in the application of this ambitious and comprehensive reform package to the reality of public schools. The authors find that, in general, the teacher community perceived the reform intentions as positive and constructive, but are dissatisfied and demotivated by the lack of information on how to navigate the new system, the difficulty to meet the new requirements (and especially so for senior teachers), the low transparency on some evaluation criteria, the scarcity of professional development spots available, the yearlong delays in implementing the promised salary increases and in obtaining career upgrades after having successfully passed the necessary examinations. Probably due to the institutional complexity surrounding the Ecuadorian reform, rigorous impact studies on its effects remain extremely scarce. Cruz-Aguayo et al. (2017) use data on teacher evaluations and find that these do not predict how effective a teacher is at raising student achievement in math and language, concluding that "the effort that is being placed by policymakers in Latin America to design and 'improve' teacher tests is unlikely to result in large improvements in child learning".

As we describe in this article, the Colombian reform experience shares some of the Ecuadorian features, most notably the frustration experienced by teachers about a partially incomplete implementation of the reform package. On the other hand, it distinguishes itself by the fact that in Colombia the new rules applied only to newly hired teachers and were not imposed on individuals who entered the profession under a different "deal", who were in fact the ones identified as most negatively affected by the reform in Ecuador. Moreover, as noted earlier, the Colombian policy message regarding teacher testing in Latin America is much more encouraging with respect to earlier ones, given that we find the Colombian teacher exam to carry significant predictive power on teacher effectiveness, in contrast with the Ecuadorian and Mexican experiences.

Appendix B. Relationship between teacher skills and student performance

In Section 7.1 we explore the relationship between teacher scores and student scores, in order to shed light on the question of how teacher quality relates to student performance. However, the relationship we estimate is potentially subject to bias due to not observing the quality of Old Regulation teachers, who also influence student performance. Therefore, in this Section we first derive the bias analytically and then we run a bias-free robustness check to prove the robustness of our results to this type of concern.

The student score SS in a specific subject i is determined by the total quality of the teachers teaching that subject Q^T. The true population model we would like to estimate is (1), which would yield β, the impact of teacher quality on student

scores. Recall that we are operating at the within-school, within-school year level — we thus omit time and school indices. The variation used to estimate β is solely within-school, within-school year, across subject variation.

$$SS_i = \alpha + \beta Q_i^T + e_i \tag{2}$$

The total teacher quality Q^T is a weighted average of New Regulation teacher quality Q^N and of Old Regulation teacher quality Q^O, and the weights are their respective shares in subject i, S^N and $1 - S^N$:

$$Q_i^T = Q_i^N \cdot S_i^N + Q_i^O \cdot \left(1 - S_i^N\right)$$

We only observe the quality of New Regulation teachers Q^N, through their entry exam test scores. Thus, instead of (1), we are only able to estimate (2)[28]:

$$SS_i = \alpha + \beta Q_i^N + e_i \tag{3}$$

So the population regression coefficient β is $\beta = \frac{cov\left(SS_i, Q_i^T\right)}{var\left(Q_i^T\right)}$ but instead we estimate

$$\begin{aligned}
\hat{\beta} &= \frac{cov\left(SS, Q^N\right)}{var\left(Q_i^N\right)} \\
&= \frac{cov\left(\alpha + \beta Q^T + e_i, Q^N\right)}{var\left(Q^N\right)} \\
&\quad \text{assume } cov\left(e, Q^N\right) = 0 \\
&= \frac{cov\left(\beta Q^T, Q^N\right)}{var\left(Q^N\right)} \\
&= \frac{\beta cov\left(Q^T, Q^N\right)}{var\left(Q^N\right)} \\
&= \beta \cdot \frac{cov\left(Q^N S^N + Q^O\left(1 - S^N\right), Q^N\right)}{var\left(Q^N\right)} \\
&\quad \text{substituting } Q^T = S_N Q^N + \left(1 - S^N\right) Q^O \\
&= \beta \cdot \frac{cov\left(Q^N S^N, Q^N\right) + cov\left(Q^O\left(1 - S^N\right), Q^N\right)}{var\left(Q^N\right)}
\end{aligned} \tag{4}$$

The term multiplying β in (3) is the expression of the bias in which we incur when trying to pin down the relationship between teacher quality and student outcomes, and observing only the quality of New Regulation teachers. Without priors on the sign and size of the sum of covariances at the numerator, the direction and size of the bias are undefined.[29]

Given that the analytical derivation of the estimation bias leaves room for uncertainty, in Table A.13 we show results on a subsample which does not suffer of the problem of unobserved teacher quality. We retain only *[school, year, subject]* observations in which the share of NR teachers is 1 - that is, in which there are no OR colleagues of unobserved quality. Clearly the sample size drops significantly, however the results confirm a positive and significant relationship between our measure of teacher quality, the entry test score, and student performance. The magnitude of the relationship actually increases in its point estimate, from the original 4% of a standard deviation to a 6% of a standard deviation, warding off concerns about a positive bias due to unobserved OR quality in our main results.

Appendix C. Testing the main identification assumptions

C.1. No selection within-school-and-year

Our within-school-and-year identification strategy is based on performance comparisons between subjects, which implies the assumption of no performance-related selection of NR teachers into subject areas, within any given school and year. That is, when looking at school i and year t, there should be no reason to expect systematic differences in the share of

[28] We are abstracting from the presence of new teachers that we were not able to match to their entry exam scores.

[29] In fact, if $S_i^N = 1$ (i.e. if all teachers were New Regulation and we thus observed the quality of all teachers, we would obtain unbiased estimation:

$$\hat{\beta} = \beta \cdot \frac{cov\left(Q_i^N, Q_i^N\right) + 0}{var\left(Q_i^N\right)} = \beta \cdot \frac{var\left(Q_i^N\right)}{var\left(Q_i^N\right)} = \beta$$

NR teachers observed in, say, Mathematics versus Critical Reading, in a way that correlates with past student performance or with performance trends in those subjects. If we were in a context characterized by school autonomy, one may expect such differences if, for example, schools tried to improve their test scores in specific subjects by selectively hiring good teachers into those areas. Another mechanism that may lead to nonrandom allocation of teachers within schools is self-sorting of teachers through transfers: one may speculate that if, within a school, specific subjects are particularly good or bad, certain types of teachers may ask to be transferred into or out-of those subjects, creating nonrandom allocation of teachers within the school–year. Section 3.4 described how public schools in Colombia have no autonomy in the opening of teaching vacancies, which makes the opportunity for selective-hiring practices very low. Also, teachers typically ask for transfers based on the location of the school and on the quality of its students (García Jaramillo et al., 2014), which are school-level factors that do not produce within-school selection. In fact these attributes, which characterize the majority of public school systems in Latin America, substantially contribute to creating the quasi-experimental setting that we exploit in order to obtain a clean causal identification of the new teacher statute.

Beyond the assurance received from the institutional context, we wish to lift any residual doubt on within-school-and-year selection of NR teachers by performing a falsification test. The test aims at detecting subject-level correlation between earlier student performance and later shares of NR teachers, within the same school and year. More specifically, through the specifications shown in Table A.1, we look for correlation between subject performance just before the reform and the shares of NR teachers we observe entering those subjects after the reform.

We estimate our main Model (1), in its complete version featuring school–year fixed effects, with student test scores of the years 2000 and 2001 on the left-hand side and shares of NR teachers over 2008–2016 on the right hand side. If NR teachers were selectively entering subjects that 'needed' them because of bad past performance, or that 'attracted' them because of good past performance, we would expect a negative or positive correlation between earlier subject-level performance and later subject-level NR teacher shares. The results, however, confirm the absence of such correlation for any of the pre-reform years employed,[30] and thus the quasi-experimental nature of our estimation setting.

C.2. No residual experience or hiring cohort effects

We further address the potential issue of age, experience or generational differences between OR and NR teachers through robustness exercises and placebo tests specifically targeted at these dimensions.

The first two columns of Table A.2 show the results of controlling for age and experience in more flexible ways with respect to the original controls. In Column (1) and (2), age and experience controls have the shape of bins – like in the main specification – and quadratic polynomials respectively. In both cases, the parameters on each element of the bin or polynomial are allowed to differ between the Old and the New Regulation, in order to account for potential changes in the effect of these variables over time. In both cases, the performance premium of New Regulation teachers remains large and highly significant, and if anything increases with respect to our main results.[31] In Columns (3) and (4) we control for age and experience through bins or a quadratic polynomial respectively, but we account for the potential effect of generational change in a different way — namely by adding controls for hiring cohorts. For each *[school, year, subject]* observation, these controls indicate the share of teachers hired since any decade between the 1980s and 2010s, absorbing the effect of different shares of "younger generations" in each subject. Once again, results remain unchanged with respect to our main specification.

Another exercise that investigates the possibility of generational change playing an important role in our setup is to construct placebo tests in which we pretend the 2001 reform was back in 1981 or 1991. We focus on the subsample of *[school–year]*s featuring only Old Regulation teachers, leaving us with around 10,000 observations. If Colombian teachers simply improved over time *within the same career regulation* and in absence of any reform – for example due to better average human capital, nutrition, teacher training, higher pressure from a more educated society, or other factors – then we should be able to pin down a difference between OR teachers hired before and after 1981 or before and after 1991.[32] However, the corresponding specifications in Column (5) and Column (6) in Table A.2 fail to identify any relevant generational effect in this sense, adding further confidence about the performance premium we identified earlier being actually due to the new set of competitive rules introduced in 2001.

Another potential source of concern related to the time dimension is career attrition, i.e. teachers leaving their jobs for other occupations. Attrition may depend on several factors, both personal and institutional, and leads to gradual changes in the observed sample of teachers over time. This feature becomes particularly relevant in contexts such as ours, in which the 'treatment' and 'control' populations are observed at different points in their temporal evolution. Old Regulation teachers have had, on average, a longer time to drop the sample with respect to New Regulation teachers, and this a

[30] Notice that using pre-reform performance for this test is the correct way to proceed. Post-reform performance at the subject level is influenced by the shares of NR teachers in the subject, and there is over-time correlation in this share, since teachers do not turn over every year. As a consequence, in post-reform years, year t subject performance may very well be correlated to the year $t+1$, $t+2$, …, $t+n$ shares of NR teachers — since these are in turn correlated to the year t share.

[31] The estimation sample drops to around 83,000 observations, since these specifications rely on having non-zero shares of both OR and NR teachers in each *[school, year, subject]*. Observations with either 100% of OR or 100% of NR teachers are not employed for estimation.

[32] Thanks to Daniel Jones and to an anonymous referee for suggesting this point.

potential source of bias in our results. In particular, one might be worried that the positive performance premium we associate to the New Regulation is in fact driven partially or entirely by a form of attrition that is related to ability, so that the worse teachers remain in the sample for longer. In this case, we might be comparing a population of OR teachers that has been negatively selected on ability over a long time to a population of NR teachers who has not yet suffered the same degree of selection — and in this case the NR performance premium we identify might be explained by the shorter ability selection instead. Now the checks performed in Table A.2, especially those in Columns (3) to (6), already partially address this issue through cohort controls: if results were strongly driven by over-time attrition, they would be sensitive to the introduction of such controls. However, in Table A.3 we provide additional evidence for attrition not playing an important role in our results by focusing on teachers that are closer and closer in their date of hiring. Notably, the value added of the robustness exercises in Table A.3 is that they address the time and cohort dimensions without relying on parametric controls for age and experience, thus ruling out the worry that the parameters on these variables might be changing over time in a way that biases our results. On the other hand, the downside of these exercises is that they require reducing the sample down to small fractions of the original one, which results in significant loss of precision and of representativeness of the full population. However, we believe these exercises are useful in order to gain an additional impression about the role of generational change and attrition in our context.

In the first set of checks we focus on subsamples in which the average hiring cohorts across different subjects are more and more similar to each other. Specifically, in Columns (1) to (3), we show the results from estimating our model focusing on *[school, year]*s in which the largest difference in average hiring cohorts across subjects is less than 15, 10 and 5 years.[33] We can see that, as hiring cohorts become more homogeneous among each other, the point estimate of the NR performance premium becomes larger. On the 5-year sample in Column (3), it reaches almost double its size on the full sample; however, the associated statistical significance falls below the ten percent level.

The second set of checks focuses on subsamples of OR and NR teachers hired around the reform implementation, and are even more demanding than the earlier ones. These subsamples include only *[school, year]*s in which *all* teachers were hired within a narrow time window around the 2002–2004 transition period – namely four years, three years and two years on each side of this period. These are schools which were either started during those time windows, or that by chance had a full staff turnover over those years, due to retirements and transfers. Given the highly restrictive requirements, the size of these samples falls to only 2%, 1% and 0.6% of the original one, and standard errors become large. However, the point estimates of the NR performance premium keep following the pattern identified in the previous set of exercises: the more homogeneous hiring cohorts become, the larger the estimated effect of the NR tends to be. If anything, one may read this as evidence suggesting that the attrition positively related to ability is stronger in the sample of NR teachers. The earliest years of the NR teacher sample – on which Table A.3 draws – still include the highest performing individuals, who drive up the NR performance premium, and as these individuals leave the teaching profession for other opportunities, the average NR performance premium falls to the size estimated in our main results. Indeed, we find evidence of high desertion among top NR performers in our attrition and survival exercises presented in the next Section.

We are aware that the subsample analyses performed in this Section may feature schools that are not necessarily representative of the average population. However, we believe that these exercises help towards further weakening the argument that the NR performance premium we estimated may be inflated by the differences in hiring cohorts between NR and OR teachers, or a negatively selected OR sample. Overall, the results from Tables A.2 and A.3 leave us with a rather solid set of evidence against several types of time-related bias inducing overestimation of the impact of the New Regulation.

Appendix D. Career attrition after entry

Quality screening and quality-related attrition along the career path might also contribute to create a wedge between the average performance of OR and NR teachers. From the description of the various post-entry novelties introduced by the New Regulation and of the institutional setting surrounding these (Section 3), it is not easy to predict the degree of merit-based selection that we should expect among NR teachers *after* they have passed the entry exam. On one hand, we would expect a larger proportion of highly-skilled individuals to be attracted and stick to the teaching profession under the New Regulation, due to the new competitive structure given to career advancement, along with the corresponding theoretically faster ascent on the pay-scale. On the other hand, the facts that the threats of dismissal due to poor performance have been going idle and that the actual salary growth opportunities have not been as smooth as on paper, are elements which may increase desertion among the high-skilled and weaken exclusion of the low-skilled individuals. In this section we perform statistical analysis on teacher retention after entry, along the permanent-position career path.

Using the nine-year individual-level panel data at our disposal, in Table A.14 we explore the retention patterns of teachers who have reached a permanent position. We perform maximum likelihood estimations of the logit model $Y_i = \beta_0 + \beta_1 NR_i + \epsilon_i$ on the sample of permanent-position teachers, where $Y_{it} = 1$ if the teacher is still in the panel in period $t + 1$ and 0 otherwise; NR_i is a dummy taking value 1 for New Regulation teachers and 0 for Old Regulation ones. We also augment the baseline (Columns 1 and 4) with a set of individual controls for age, experience, education, subject

[33] Notice that this is a requirement on subject-level averages, and it does not imply that the largest hiring cohort difference between each pair of *individual teachers* is no larger than those cutoffs.

and geographic area (Columns 2 and 5). Finally, we allow for individual teacher effects in the random-effects specifications (Columns 3 to 6). Once accounted for individual characteristics, NR teachers indeed present lower retention in the public teaching career with respect to their OR colleagues, the yearly survival odds ratios being around 0.88 (Columns 2 and 5). Given the qualitative evidence on the very sporadic implementation of forced dismissals, we are inclined to explain our result by a higher mobility of the new type of teachers across the labor market.

Fig. A.4 shows the results of estimating the same panel retention model separately by entry test score deciles.[34] For deciles 1 to 8, the yearly survival odds of NR teachers are not significantly different from those of OR colleagues, but fall dramatically for top performers, reaching levels close to 0.5 for individuals at the 10th decile of the entry test score distribution. The higher average mobility of NR teachers thus seems driven by the individuals performing highest on the entry test. Moreover, comparing NR only among themselves, we estimate Kaplan–Meier survival functions[35] by entry test score deciles and find further support to the conclusions from our previous logit model. From the first and up to the 9th decile of the performance distributions, better entry test scores correspond to slightly higher survival rates in the system. Nevertheless, teachers at the 10th decile survive significantly less than anyone else — with more than half of these individuals leaving the public teacher career by their 10th year of tenure. Fig. A.5 shows our survival estimates for teachers at the 1st, 9th and 10th deciles.

This set of evidence, although not reflecting any causal relationship, points towards the conclusion that the merit-based incentives characterizing the NR do not seem satisfactory enough to retain individuals at the very top of the achievement distribution, who leave the profession at significantly faster pace than colleagues at any other performance level. With the data at hand at this stage, we are not able to answer the question of why exactly top performers leave early. They might be negatively surprised by the reality of the NR public teacher career, not corresponding to their fast-career expectations. Given their non-education specific degrees, they might also consciously enter the profession with a short-term horizon in mind, while awaiting more attractive professional opportunities to materialize.

Appendix E. Figure appendix

See Figs. A.1–A.4.

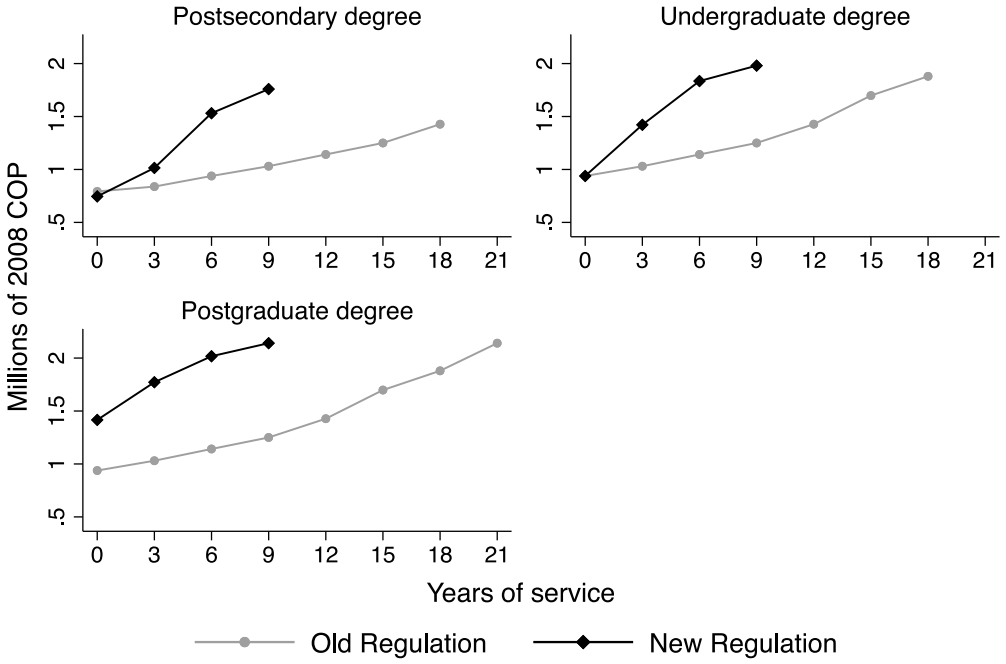

Fig. A.1. Base salary evolution by education level and type of regulation. *Note:* The figure plots the theoretical salary prospects faced by teachers of the New and the Old Regulation, depending on their achieved education level. Salary upgrades are attainable every three years, conditional on holding the education degree and having attended the training courses required for the following level. Upgrades are automatic for OR teachers, and depend on positive performance evaluations for NR teachers. Source: compiled by the authors based on Decree 2277/1979, Decree 1278/2002 Decree 259/1981, Decree 626/2008, Decree 624/2008 (*Departamento Administrativo de la Función Pública*. Salaries in 2008 Colombian Pesos.)

34 Working with the subsample of NR teachers who were associated to their scores.

35 A public teacher becomes "at risk of failing" the year in which he/she enters the teacher profession, and "fails" the year in which he/she exits it, for whichever reason. The survival analysis performed in this section is based on the individual-level information recorded over the direct observation period 2008–2016, and on the retrospective information about each teacher's initial hiring year. The survival time of teachers hired before the start of our observation window in 2008 is treated as conditional on having survived up to that year and enters the survival likelihood function with appropriate weights – i.e. we are in the presence of so called "left censoring", not to be confused with left truncation. See the excellent discussion in Wooldridge (2010).

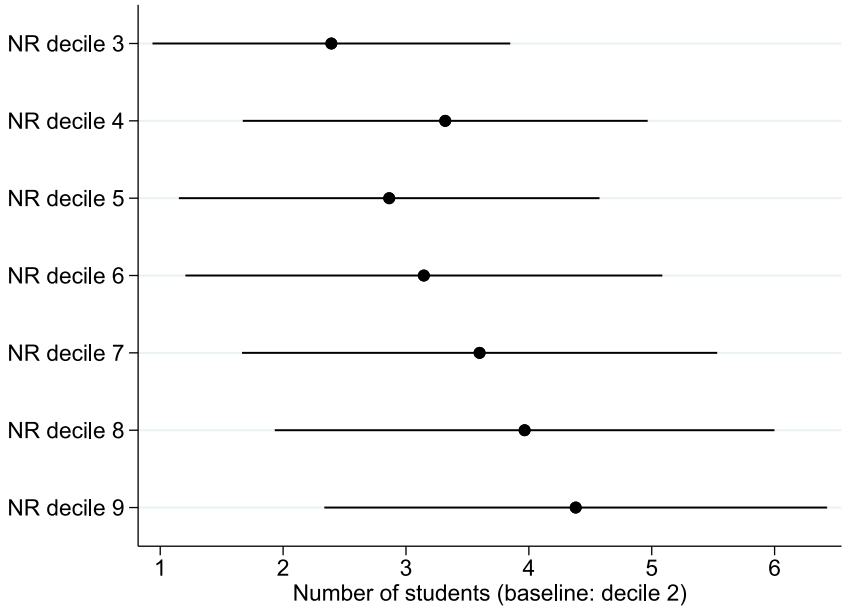

Fig. A.2. Relationship between number of test takers and NR share (2008–2016). *Note:* The figure plots the results of estimating the within-school relationship between the share of NR teachers and the number of test takers in schools, over the period 2008–2016. Point estimates indicate the difference with respect to the baseline decile 2. Deciles 1 and 10 have been omitted from the illustration. Numerical results are shown in Table A.6, Column 2.

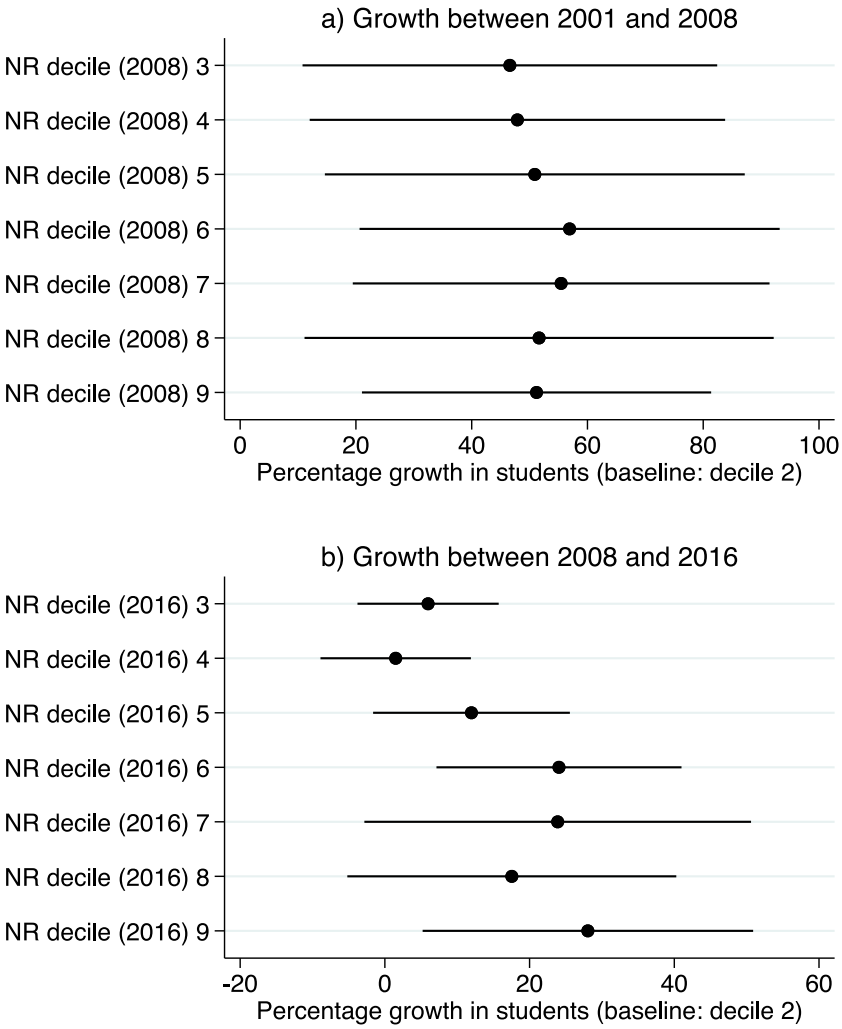

Fig. A.3. Relationship between growth in test takers and NR share. *Note:* The figure plots the results of estimating the across-school relationship between the share of NR teachers and the growth in test taker cohorts, between the years 2001 and 2008 (a) and between the years 2008 and 2016 (b). Point estimates indicate the difference with respect to the baseline decile 2. Deciles 1 and 10 have been omitted from the illustration. Numerical results are shown in Table A.6, Columns 3 and 4.

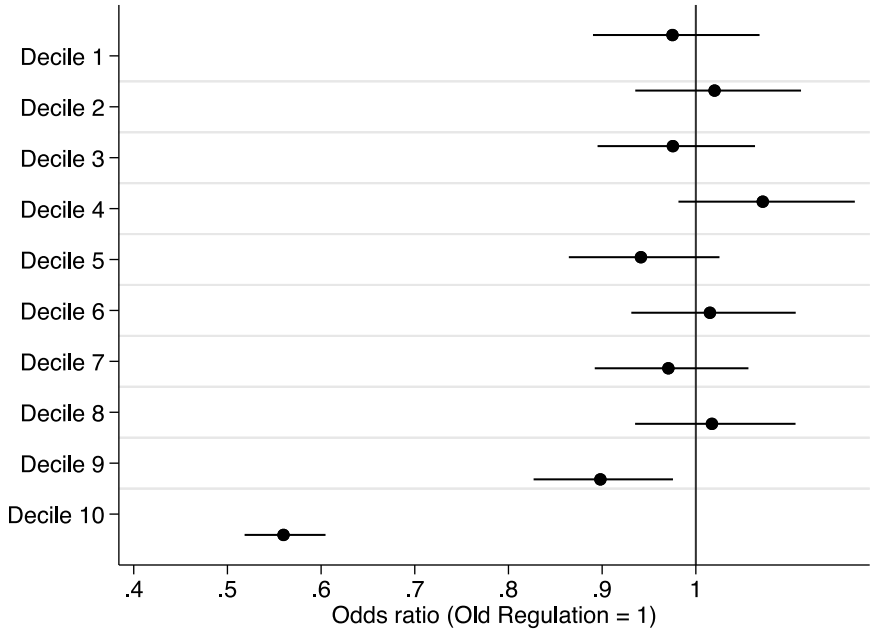

Fig. A.4. Survival odds ratios of NR teachers with respect to OR teachers. *Note:* Annual survival odds ratios of NR permanent-position teachers with respect to OR permanent-position teachers, by entry test score decile of NR teachers. Point estimates and 95% confidence spikes with Bonferroni correction.

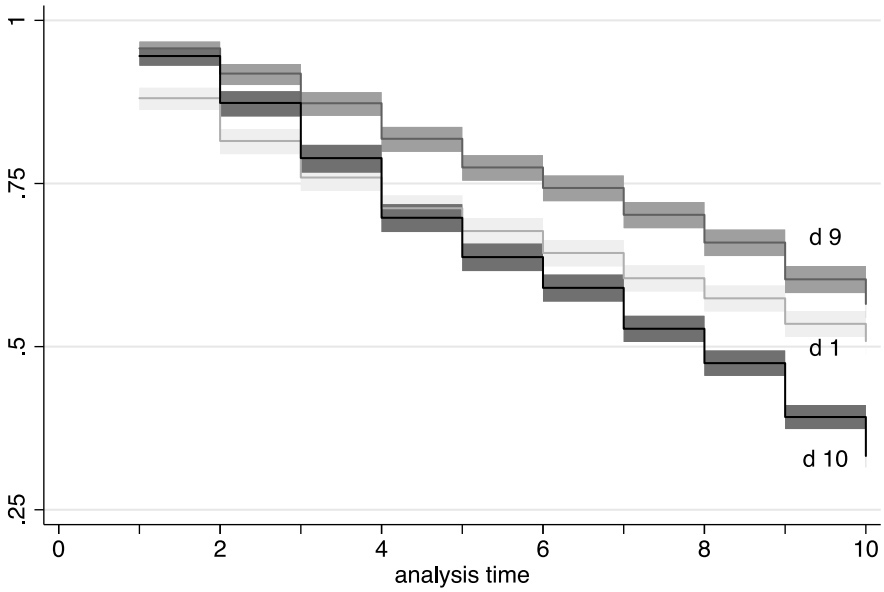

Fig. A.5. Kaplan–Meier survival functions by entry score decile. *Note:* Kaplan–Meier survival functions for New Regulation teachers, according to their entry test score decile (deciles 1, 9 and 10). Point estimates (solid lines) and 95% confidence intervals (shaded).

Appendix F. Table appendix

See Tables A.1–A.14.

Table A.1
Post-reform shares of NR teachers do not relate to pre-reform subject performance.

	2000 test scores			2001 test scores			2000–2001 change		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
New Regulation	0.011	0.033	0.033	−0.002	0.002	0.002	−0.012	−0.036	−0.036
	(0.018)	(0.029)	(0.029)	(0.019)	(0.029)	(0.029)	(0.023)	(0.035)	(0.035)
Postgraduate degree		0.023	0.023			0.011			−0.008
		(0.018)	(0.018)			(0.019)			(0.021)
Age and experience	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School–year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject-specific trends	No	No	Yes	No	No	Yes	No	No	Yes
N.obs	92,830	92,830	92,830	96,280	96,280	96,280	90,377	90,377	90,377
N.groups	25,399	25,399	25,399	26,362	26,362	26,362	24,662	24,662	24,662
R-squared	0.85	0.85	0.85	0.86	0.86	0.86	0.66	0.66	0.66

Note: Pre-reform student test scores regressed on post-reform shares of New Regulation teachers, within school and year. If there were selection of New Regulation teachers into subjects, we would expect significant relationships to appear. Each observation is the average score obtained subject 's' in school 'i' in year 2000, 2001 or the change in score between the two years. SEs clustered by schools in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2
Robustness and placebo on experience or cohort effects.

	Robustness				Placebo	
	(1)	(2)	(3)	(4)	(5)	(6)
New Regulation	0.10**	0.09**	0.07**	0.07**		
	(0.04)	(0.04)	(0.03)	(0.03)		
OR hired after 1981					−0.09	
					(0.12)	
OR hired after 1991						0.01
						(0.11)
Age and Experience (AE) bins	Yes		Yes		Yes	Yes
AE quadratic		Yes		Yes		
AE separate by regulation	Yes	Yes				
Hiring cohort dummies			Yes	Yes		
Postgraduate degree	Yes	Yes	Yes	Yes	Yes	Yes
School–year FE	Yes, all columns					
Subject FE	Yes, all columns					
Subject-specific trends	Yes, all columns					
N.obs	83,725	83,729	207,545	207,545	10,047	10,047
N.groups	32,601	32,603	47,481	47,481	2,934	2,934
R-squared	0.94	0.94	0.89	0.89	0.85	0.85

Note: In all columns, each observation is one subject 's' in school 'i' in year 'y'; all specifications correct for both selection across schools and within schools over time. Columns (1) and (2) show more flexible or alternative ways to control for age and experience with respect to our main results. In Column (1), the parameters on age and experience bins are allowed to differ between OR and NR. In Column (2), age and experience are instead controlled for through a quadratic polynomial for each Regulation. These two samples are smaller because all subject–year observations in which either OR or NR teachers are missing are dropped. In Column (3), we add controls for 'hiring cohorts', indicating the share of teachers hired in each 10-year segment between the 1980s and the 2010s. Columns(4) and (5) show placebo tests based on school–years with OR teachers only, estimating the effect of the generational change within the same teacher regulation. SEs clustered by school in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3
Analysis on similar-cohort subsamples.

	Similar average hiring cohort			Hired close to reform only		
	(1) diff < 15y	(2) diff < 10y	(3) diff < 5y	(4) 98 - 08	(5) 99 - 07	(6) 00 - 06
New Regulation	0.07***	0.10***	0.13	0.14	0.15	0.19
	(0.02)	(0.04)	(0.10)	(0.13)	(0.19)	(0.29)
Postgrad education	Yes	Yes	Yes	Yes	Yes	Yes
School–year FE	Yes, all columns					
Subject FE	Yes, all columns					
Subject-specific trends	Yes, all columns					
N.obs	129,899	75,952	25,254	4,340	2,193	1,235
N.groups	30,708	19,049	7,823	1,667	946	596
R-squared	0.89	0.89	0.86	0.77	0.79	0.81

Note: In all columns, each observation is one subject 's' in school 'i' in year 'y'; all specifications correct for both selection across schools and within schools over time. These specifications represent non-parametric alternatives to control for experience or cohort effects: age and experience controls are not included. The subsamples in columns (1)-(3) include only school–years in which the *average* experience is similar across subjects. The subsamples in columns (4)-(6) include only school–years in which *all* teachers were hired in years close to the reform. SEs clustered by school in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4
Stages of the entry contest.
Source: MEN (2006); MEN (2012b)

	Purpose	Use of score	Minimum threshold	Weight in contest		Responsibility
				T*	P*	
Exam	Teaching aptitude, subject knowledge	Eliminatory and ranking	60% (T) 70% (P)	55%	45%	ICFES[5]
Exam	Psychometric test	Ranking	–	10%	10%	ICFES
CV	Credentials evaluation	Ranking	–	20%	30%	CNSC[6] or delegate
Interview	Behavioral evaluation	Ranking	–	15%	15%	CNSC or delegate

Note: * T = teachers; P = school principals.

Table A.5
Career structure of public school teachers, and 2008 pay scales.
Source: Compiled by the authors based on Decree 2277/1979, Decree 1278/2002 Decree 259/1981, Decree 626/2008, Decree 624/2008. Salaries in 2008 Colombian Pesos. The shaded steps are the possible entry steps for first-time teachers.

Old Regulation (Dec. 2277 / 1979)			**New Regulation (Dec. 1278 / 2002)**				
Step	Education level required	2008 salary	Level	Step	Education level required	2008 salary	
A	Secondary school	525.240	1	A	Post-secondary specialization	745.624	
B		581.850		B		1.014.611	
1		652.079		C		1.531.186	
2	Postsecondary specialization	675.922		D		1.759.188	
3		717.284	2	A	Undergraduate degree	938.340	
4		745.600		B		1.421.428	
5		792.628		C		1.834.801	
6	Undergraduate degree	838.439		D		1.980.454	
7		938.315				Master deg.	PhD deg.
8		1.030.680	3	A	Postgraduate degree	1.415.933	1.721.798
9		1.141.779		B		1.772.111	2.154.919
10		1.250.166		C		2.017.127	2.452.864
11		1.427.513		D		2.140.784	2.603.232
12		1.698.112					
13		1.879.682					
14	Postgraduate degree	2.140.766					

Table A.6
New Regulation teachers and student sorting.

	Variation across years		Variation across schools	
	(1) Number of test takers	(2) Number of test takers	(3) Growth in test takers (%)	(4) Growth in test takers (%)
New Regulation	3.19***			
	(1.10)			
NR decile 1		−0.65	53.77***	12.08**
		(0.74)	(18.60)	(5.86)
NR decile 3		2.39***	46.59**	5.98
		(0.74)	(18.27)	(4.98)
NR decile 4		3.32***	47.90***	1.49
		(0.84)	(18.30)	(5.30)
NR decile 5		2.86***	50.90***	11.97*
		(0.87)	(18.50)	(6.93)
NR decile 6		3.15***	56.92***	24.06***
		(0.99)	(18.51)	(8.64)
NR decile 7		3.60***	55.46***	23.89*
		(0.99)	(18.36)	(13.63)
NR decile 8		3.97***	51.64**	17.54
		(1.04)	(20.67)	(11.60)
NR decile 9		4.38***	51.20***	28.06**
		(1.04)	(15.38)	(11.65)
NR decile 10		0.47	0.00	
		(1.68)	(.)	
Age and experience	Yes	Yes	Yes	Yes
Postgrad education	Yes	Yes	Yes	Yes
School FE	Yes	Yes		
Year FE	Yes	Yes		
School-specific trends	Yes	Yes		
N.obs	41,939	41,939	2,587	3,984
N.groups	6,431	6,431		
R-squared	0.92	0.92	0.09	0.02

Note: In columns (1) and (2), each observation is one school 'i' in a year 'y': we estimate the within-school relationship between the share of NR teachers and the number of students sitting the Saber11 exam; school fixed effects as well as school-specific time trends are included. In columns (3) and (4), each observation is one school 'i': we estimate the across-school relationship between the share of NR teachers and the growth in the test taker cohort between 2001 and 2008 (Column 3) and between 2008 and 2016 (Column 4). The NR deciles are yearly deciles for Column 2, 2008 deciles for Column 3 and 2016 deciles for Column 4. Age and experience controls are 10-year (age) and 5-year (experience) bins. SE clustered by school (Columns 1-2) or robust (Columns 3-4) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7
Individual-level teacher descriptives.

	All teachers	Old regulation	All new regulation	New regul. w/scores	New regul. certified	New regul. not certified
Age	45.66	51.26	38.28	38.38	38.52	36.02
	(10.10)	(7.42)	(8.24)	(8.15)	(7.92)	(8.67)
Experience	13.95	21.56	3.91	3.96	4.46	0.86
	(11.40)	(9.23)	(3.42)	(3.33)	(3.35)	(1.84)
Female	0.56	0.57	0.54	0.54	0.52	0.62
	(0.50)	(0.49)	(0.50)	(0.50)	(0.50)	(0.48)
Postgrad degree	0.27	0.37	0.15	0.15	0.18	0.04
	(0.45)	(0.48)	(0.35)	(0.36)	(0.39)	(0.19)
Experience 5-40yrs	0.71	0.96	0.38	0.39	0.45	0.05
	(0.45)	(0.20)	(0.49)	(0.49)	(0.50)	(0.22)
Age when hired	31.71	29.70	34.37	34.42	34.06	35.17
	(7.67)	(6.99)	(7.71)	(7.67)	(7.32)	(8.47)
Rural area	0.23	0.16	0.33	0.33	0.29	0.40
	(0.42)	(0.36)	(0.47)	(0.47)	(0.45)	(0.49)
Permanent position	0.80	0.97	0.57	0.58	0.76	0.03
	(0.40)	(0.16)	(0.50)	(0.49)	(0.43)	(0.16)
Temporary position	0.17	0.03	0.35	0.34	0.14	0.96
	(0.37)	(0.16)	(0.48)	(0.47)	(0.35)	(0.20)
Probation position			0.08	0.08	0.10	0.02
			(0.27)	(0.27)	(0.30)	(0.12)
Most recent test score					63.56	53.04
					(7.05)	(8.42)
N	719,441	409,208	310,233	298,763	224,484	22,003
N teachers	151,882	75,660	76,222	70,061	48,058	22,003

Note: Variable means and (standard deviations).

Table A.8
New Regulation teachers by entry exam result.

	Never passed (Not Certified - NC)	Passed at least once (Certified - C)	Total
Total	28,164	48,058	76,222
	36.95%	63.05%	100%
Permanent	1,138	22,970	
	4.04%	47.80%	
Temporary	26,582	12,816	
	94.38%	26.67%	
Probation	444	12,272	
	1.58%	25.54%	

Note: The table divides NR teachers who were matched with their entry exam scores by whether they have reached the minimum score threshold (60 for teachers, 70 for principals) on at least one attempt. Those who did are ''Certified'' and those who are not are ''Not certified''. Each of the two categories is further divided by the type of position we observe these teachers occupying.

Table A.9
Descriptive statistics at school–year–subject level.

	Total	Between	Within
Mean student score	0.00		
	(2.79)	(2.60)	(1.00)
Share New Regulation	0.52		
	(0.41)	(0.32)	(0.27)
Share New Regulation Qualified	0.35		
	(0.38)	(0.27)	(0.28)
Share New Regulation Not Qualified	0.17		
	(0.31)	(0.23)	(0.22)
Mean age	44.36		
	(7.95)	(5.67)	(5.84)
Mean experience	12.41		
	(8.65)	(6.60)	(5.84)
Share postgraduate degree	0.25		
	(0.34)	(0.25)	(0.24)
N	207,545		
N groups	47,481		

Note: Variable means and (standard deviations). Standard deviations are Total, Between-group and Within-group, where a group is a school–year combination.

Table A.10
Selectivity of the entry exam.
Source: MEN (2013a).

CONTESTS	1st (2004)	2nd (2005)	3rd (2006)	4th (2009)	5th (2013)
N of local authorities	69	66	49	66	92
Vacancies	50.947	23.355	14.579	25.392	NA
Candidates to exam	140.541	134.090	109.487	228.985	301.589
Passed exam stage	60.078 (43%)	32.720 (24%)	27.931 (26%)	66.687 (29%)	54.906 (18%)

Table A.11
The influence of teachers' entry exam scores, by deciles, on student test scores.

	Simple	School selection	School and time selection			
	(1)	(2)	(3)	(4)	(5)	(6)
Decile 2	0.19	0.23***	0.02	0.02	0.02	0.02
	(0.10)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Decile 3	0.48***	0.32***	0.05*	0.05	0.05	0.05
	(0.10)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Decile 4	0.75***	0.35***	0.08***	0.07**	0.07**	0.07**
	(0.10)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Decile 5	0.87***	0.37***	0.07***	0.06**	0.06**	0.06**
	(0.10)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Decile 6	1.12***	0.36***	0.09***	0.08***	0.08***	0.08***
	(0.10)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Decile 7	1.30***	0.34***	0.13***	0.12***	0.12***	0.12***
	(0.10)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Decile 8	1.46***	0.33***	0.13***	0.12***	0.12***	0.12***
	(0.10)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Decile 9	1.65***	0.28***	0.15***	0.14***	0.14***	0.14***
	(0.10)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Decile 10	1.81***	0.19***	0.16***	0.15***	0.15***	0.15***
	(0.11)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
New Regulation						0.02
						(0.03)
Postgrad education				Yes	Yes	Yes
Age and experience				Yes	Yes	Yes
School FE		Yes	Yes	Yes	Yes	Yes
Year FE		Yes	Yes	Yes	Yes	Yes
School–year FE			Yes	Yes	Yes	Yes
Subject FE			Yes	Yes	Yes	Yes
Subject-specific trends					Yes	Yes
N.obs	6,415	45,734	164,226	164,226	164,226	164,226
N.groups	.	6,412	45,733	45,733	45,733	45,733
R-squared	0.09	0.87	0.90	0.90	0.90	0.90

Note: This table reports the results underlying Fig. 2. In column (1), each observation is one school 'i'. In column (2), each observation is one school 'i' in year 'y'. In columns (3) to (6), each observation is one subject 's' in school 'i' in year 'y'. Age and experience controls are 10-year (age) and 5-year (experience) bins. SEs clustered by school in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, Bonferroni correction for multiple hypothesis testing applied.

Table A.12
The effect of NR teachers, by certification status, on student performance.

	Simple	School selection	School and time selection		
	(1)	(2)	(3)	(4)	(5)
NR Certified	−0.078	0.089*	0.073***	0.084***	0.084***
	(0.106)	(0.051)	(0.016)	(0.023)	(0.023)
NR Non-certified	−2.583***	−0.452***	−0.018	0.017	0.017
	(0.113)	(0.064)	(0.019)	(0.027)	(0.027)
Postgrad education				Yes	Yes
Age and experience				Yes	Yes
School FE		Yes	Yes	Yes	Yes
Year FE		Yes	Yes	Yes	Yes
School–year FE			Yes	Yes	Yes
Subject FE			Yes	Yes	Yes
Subject-specific trends					Yes
N.obs	6,471	47,481	207,545	207,545	207,545
N.groups	.	6,471	47,481	47,481	47,481
R-squared	0.08	0.87	0.89	0.89	0.89

Note: In column (1), each observation is one school 'i'. In column (2), each observation is one school 'i' in year 'y'. In columns (3) to (5), each observation is one subject 's' in school 'i' in year 'y'. Age and experience controls are 10-year (age) and 5-year (experience) bins. SEs clustered by school in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.13
Teacher's scores and student scores — New Regulation teachers only.

	Simple	School selection	School and time selection		
	(1)	(2)	(3)	(4)	(5)
Teachers' score	0.762***	0.080***	0.061***	0.058***	0.058***
	(0.050)	(0.024)	(0.011)	(0.012)	(0.012)
Postgrad education				Yes	Yes
Age and experience				Yes	Yes
School FE		Yes	Yes	Yes	Yes
Year FE		Yes	Yes	Yes	Yes
School–year FE			Yes	Yes	Yes
Subject FE			Yes	Yes	Yes
Subject-specific trends					Yes
N.obs	5,274	28,536	65,349	65,349	65,349
N.groups	.	5,274	28,536	28,536	28,536
R-squared	0.07	0.85	0.89	0.89	0.89

Note: In column (1), each observation is one school 'i'. In column (2), each observation is one school 'i' in year 'y'. In columns (3) to (6), each observation is one subject 's' in school 'i' in year 'y'. Age and experience controls are 10-year (age) and 5-year (experience) bins. SEs clustered by school in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.14
Panel retention for permanent-position NR teachers.

	Logit			RE Logit		
	(1)	(2)	(3)	(4)	(5)	(6)
New Regulation	1.00	0.87***		1.03***	0.88***	
	(0.01)	(0.01)		(0.01)	(0.01)	
Decile 2			1.11*			1.12**
			(0.04)			(0.05)
Decile 3			1.09			1.09
			(0.04)			(0.04)
Decile 4			1.21***			1.22***
			(0.05)			(0.05)
Decile 5			1.09			1.09
			(0.04)			(0.04)
Decile 6			1.22***			1.23***
			(0.05)			(0.05)
Decile 7			1.20***			1.21***
			(0.05)			(0.05)
Decile 8			1.30***			1.31***
			(0.05)			(0.05)
Decile 9			1.17***			1.17***
			(0.05)			(0.05)
Decile 10			0.72***			0.70***
			(0.03)			(0.03)
Age and experience		Yes	Yes		Yes	Yes
Education		Yes	Yes		Yes	Yes
Geographic location		Yes	Yes		Yes	Yes
N.obs	468,246	468,246	151,163	468,246	468,246	151,163
N.groups				100,779	100,779	33,592

Note: The table displays survival odds ratios of permanent-position New Regulation teachers relative to permanent-position Old Regulation colleagues, as well as the survival odds ratios of New Regulation teachers at different entry-score deciles. The outcome variable is Y=1 if the teacher is still in the panel the following year, Y=0 otherwise. In columns (1), (2), (4) and (5), the sample includes permanent-position NR and OR teachers. In columns (3) and (6) only NR teachers are used, and the Bonferroni correction for multiple hypothesis testing is applied. For columns (1) to (3) SEs clustered by individual in parentheses; for (4) to (6) Observed Information Matrix SEs in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Araujo, M.D., 2019. Measuring the effect of competitive teacher recruitment on student achievement: Evidence from Ecuador. Technical Report, BERG Working Paper Series.

Arizmendi Domí nguez, A., Saa García, B., 2018. Teacher evaluation in Colombia. Revista Educación, PolíTica Y Sociedad 3 (2), 60–75.

Ashraf, N., Bandiera, O., Lee, S., 2018. Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. Technical Report 65, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE, London, UK.

Avalos-Bevan, B., 2018. Teacher evaluation in Chile: highlights and complexities in 13 years of experience. Teachers Teaching 24 (3), 297–331.

Bau, N., Das, J., 2017. The misallocation of pay and productivity in the public sector: Evidence from the labor market for teachers. Technical Report, The World Bank.

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

Boyd, D., Goldhaber, D., Lankford, H., Wyckoff, J., 2007. The effect of certification and preparation on teacher quality. Future Child. 45–68.

Boyd, D., Lankford, H., Loeb, S., Wyckoff, J., 2005. Explaining the short careers of high-achieving teachers in schools with low-performing students. Amer. Econ. Rev. 95 (2), 166–171.

Bruns, B., Luque, J., 2014. Great Teachers: how To Raise Student Learning in Latin America and the Caribbean. The World Bank.

Chetty, R., Friedman, J.N., Rockoff, J.E., 2014. Measuring the impacts of teacher I: Evaluating bias in teacher value-added estimates. Amer. Econ. Rev. 104 (9), 2633–2679.

Chiriboga Montalvo, C., Pinto Haro, J., 2019. Reforma de la carrera docente en Ecuador: Estudio de caso. Technical Report, Instituto Internacional de Planificación de la Educación, UNESCO, Paris, France.

Cifuentes Cubillos, C., 2014. Impacto del Nuevo Estatuto de Profesionalización en la función docente en Colombia. Análisis de los dos estatutos vigentes: Decreto 2277 de 1979 y Decreto 1278 del 2002. Revista Colombiana de SociologíA 37 (2), 213–250.

Congreso de Colombia, 1994. Ley general de Educación, 115/1994.

Consejo de Estado de Colombia, 2004. Radicación No. 1603/2004.

Cortés, F., Lagos, M.J., 2011. Consecuencias de la evaluación docente. In: La Evaluación Docente En Chile. Centro UC Medición - MIDE, Santiago.

Cowan, J., Goldhaber, D., 2016. National board certification and teacher effectiveness: Evidence from Washington State. J. Res. Edu. Effect. 9 (3), 233–258.

Cruz-Aguayo, Y., Ibarrarán, P., Schady, N., 2017. Do tests applied to teachers predict their effectiveness? Econ. Lett. 159, 108–111.

Dal Bó, E., Finan, F., Rossi, M.n.A., 2013. Strengthening state capabilities: the role of financial incentives in the call to public service. Q. J. Econ. 128 (3), 1169–1218.

Deserranno, E., 2019. Financial incentives as signals: experimental evidence from the recruitment of village promoters in Uganda. Am. Econ. J. Appl. Econ. 11 (1), 277–317.

Duarte, J., 2001. Política y educación: Tentaciones particularistas en la educación Latinoamericana. EconomíA PolíTica de Las Reformas Educativas En América Latina.

Duarte, J., 2003. Educación Pública Y Clientelismo En Colombia. Universidad de Antioquia, Medellín.

ECE, 2017. Analisis y consolidado de la evaluación de desempeño de docentes y directivos docentes del Departamento del Magdalena 2016. Technical Report, Equipo Calidad Educativa, Secretaria de Educación del Magdalena, Santa Marta.

Echeverri Álvarez, J.C., Martí nez, J.D., Pérez Ramí rez, F.A., Rí os Osorio, E.A., Zuluaga Avendaño, D., Delgado Paz, O.A., 2017. La evaluación anual de desempeño de docentes y directivos docentes en Colombia: Interpretaciones de una lectura local. In: Colección ECDOC, ASCOFADE, Colombia.

Eide, E., Goldhaber, D., Brewer, D., 2004. The teacher labour market and teacher quality. Oxf. Rev. Econ. Policy 20 (2), 230–244.

Elacqua, G., Hincapie, D., Vegas, E., Alfonso, M., 2018. Profesión: Profesor en América Latina. šPor qué se perdió el prestigio docente y cómo recuperarlo?. Interamerican Development Bank, Washington, D.C..

Estrada, R., 2019. Rules versus discretion in public service: Teacher hiring in Mexico. J. Labor Econ. 37 (2), 545–579.

Figlio, D.N., 2001. Can public schools buy better-qualified teachers. Indus. Labor Relations Rev. 55, 686.

Finan, F., Olken, B.A., Pande, R., 2017. The personnel economics of the developing state. In: Handbook of Economic Field Experiments, vol. 2. Elsevier, pp. 467–514.

Gallego Henao, A.M.a., Vásquez Jaramillo, O.C., 2015. Promotion and wage categorization in teachers. a dignification issue regarding teacher's work. Revista Virtual Universidad CatÓlica Del Norte 46, 60–70.

García Jaramillo, S., Maldonado, D., Perry Rubio, G., Rodríguez, C., Saavedra Calvo, J.E., 2014. Tras la excelencia docente: Cómo mejorar la calidad de la educación para todos los colombianos. Technical Report, Fundación Compartir, Bogotá, D.C..

Glewwe, P., Hanushek, E.A., Humpage, S., Ravina, R., 2013. School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010. University of Chicago Press.

Goldhaber, D., Anthony, E., 2007. Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. Rev. Econ. Stat. 89 (1), 134–150.

Goldhaber, D., Cowan, J., Theobald, R., 2017b. Evaluating prospective teachers: Testing the predictive validity of the edTPA. J. Teacher Edu. 68 (4), 377–393.

Goldhaber, D., Grout, C., Huntington-Klein, N., 2017a. Screen twice, cut once: Assessing the predictive validity of applicant selection tools. Edu. Finance Policy 12 (2), 197–223.

Goldhaber, D., Liu, A., 2003. Occupational choices and the academic proficiency of the teacher workforce. Developments in School Finance 2001 2, 53–75.

Gordon, R.J., Kane, T.J., Staiger, D.O., 2006. Identifying Effective Teachers using Performance on the Job. Brookings Institution Washington, DC.

Guzmán Marí n, F., 2018. La experiencia de la evaluación docente en México: Análisis crítico de la imposición del servicio profesional docente. Revista Iberoamericana de Evaluación Educativa 11 (1), 135–158.

Hanushek, E.A., 2009. School policy: implications of recent research for human capital investments in South Asia and other developing countries. Educ. Econ. 17 (3), 291–313.

Hanushek, E.A., Haycock, K., 2010. An effective teacher in every classroom. Education Next 10 (3).

Hanushek, E.A., Kain, J.F., Rivkin, S.G., 2004. Why public schools lose teachers. J. Hum. Resour. 39 (2), 326–354.

Hanushek, E.A., Rivkin, S.G., 2006. Teacher quality. Handbook Econ. Edu. 2, 1051–1078.

Hanushek, E.A., Rivkin, S.G., 2012. The distribution of teacher quality and implications for policy. Annu. Rev. Econ. 4, 131–157.

Hanushek, E.A., Woessmann, L., 2012. Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. J. Econ. Growth 17 (4), 267–321.

Hendricks, M., 2016. Differential teacher attrition: Do high-ability teachers exit at higher rates? Available At SSRN 2824586.

Kane, T.J., Rockoff, J.E., Staiger, D.O., 2008. What does certification tell us about teacher effectiveness? Evidence from New York City. Econ. Educ. Rev. 27 (6), 615–631.

Larsen, B., 2013. Occupational licensing and quality: Distributional and heterogeneous effects in the teaching profession. Available At SSRN 2387096.

Lavy, V., 2009. Performance pay and teachers' effort, productivity, and grading ethics. Amer. Econ. Rev. 99 (5), 1979–2011.
Leaver, C., Ozier, O., Serneels, P., Zeitlin, A., 2021. Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools. Amer. Econ. Rev. 111 (7), 2213–2246.
Legarralde, M., Palamidessi, M., 2006. Sindicalismo docente, gobiernos y reformas educativas en América Latina y el Caribe: Condiciones para el diálogo. Technical Report, Interamerican Development Bank.
Manzi, J., Gonzalez, R., Sun, Y., 2011. La Evaluación Docente En Chile. Centro UC Medición - MIDE, Santiago.
Maussa Díaz, E., 2012. Evaluación de docentes, escalafones salariales e inversión en educación. Económicas CUC 33 (1), 225–238.
MEN, 2006. Decreto 3982 de 2006, vol. 46449. Ministerio de Educación Nacional; Diario Oficial.
MEN, 2012a. Guia para la administración de los recursos financieros del sector educativo. Technical Report 8, Ministerio de Educación Nacional, Bogotá, D.C..
MEN, 2012b. Por meritocracia se seleccionarán 20.610 docentes, directivos docentes y etnoeducadores. Technical Report, Ministerio de Educación Nacional, http://www.mineducacion.gov.co/cvn/1665/w3-article-312669.html [21 Jan 2015].
MEN, 2013a. Estadísticas concursos 2004 - 2009. Technical Report, Ministerio de Educación Nacional.
MEN, 2013b. šQuiénes se reubican salarialmente o ascienden?. Technical Report, Ministerio de Educación Nacional, http://www.mineducacion.gov.co/proyectos/1737/w3-article-309814.html [22 Jan 2015].
MEN, 2014. Informe de resultados de la evaluación de competencias para el ascenso o la reubicación de nivel salarial en el escalafón de docentes y directivos docentes regidos por el Decreto Ley 1278 de 2002 - Agregado Nacional.
Mizala, A., Schneider, B., 2019. Promoting quality education in Chile: the politics of reforming teacher careers. J. Education Policy 1–27.
Muralidharan, K., Sundararaman, V., 2011. Teacher performance pay: Experimental evidence from India. J. Polit. Econ. 119 (1), 39.
Murillo, F., González, A., Rizo, M., 2007. Evaluación del desempeño y carrera profesional docente. Una panorámica de América y Europa, second ed. Santiago de Chile: UNESCO.
OECD, 2016. Education in Colombia. Technical Report, OECD Publishing, Paris.
OECD, 2017. The teaching workforce in Chile. In: OECD Reviews of School Resources: Chile 2017. OECD, Paris, France.
Ome, A., 2012. The Effects of Meritocracy for Teachers in Colombia. Technical Report 010260, Fedesarrollo.
Ome, A., 2013. El estatuto de profesionalización docente: una primera evaluación. Fedesarrollo.
Rivkin, S.G., Hanushek, E.A., Kain, J.F., 2005. Teachers, schools, and academic achievement. Econometrica 73 (2), 417–458.
Rockoff, J.E., 2005. The impact of individual teachers on student achievement: Evidence from panel data. Amer. Econ. Rev. 5 (1), 247–252.
Rockoff, J.E., Jacob, B.A., Kane, T.J., Staiger, D.O., 2011. Can you recognize an effective teacher when you recruit one? Education 6 (1), 43–74.
Saavedra, J., Maldonado, D., Santibanez, L., Prada, L.O.H., 2017. Premium or Penalty? Labor Market Returns to Novice Public Sector Teachers. Technical Report, National Bureau of Economic Research.
Sánchez Cotrino, O.A., 2014. Evaluación de desempeño y proletarización docente. Análisis de contenido de las políticas de evaluación docente. (Ph.D. thesis). Universidad Pedagógica Nacional, Bogotá, D.C..
Subsecretaria de Calidad Educativa, 2015. Evaluación de desempeño de directivos docentes y docentes del Departamento del Valle del Cauca - Documento de aplicación, análisis y uso de resultados año 2014. Technical Report, Gobernación del Valle del Cauca - Secretaria de Educación, Santiago de Cali.
Tatto, M.T., Schmelkes, S., Guevara, M.d.R., Tapia, M., 2006. Implementing reform amidst resistance: The regulation of teacher education and work in Mexico. Int. J. Edu. Res. 45 (4–5), 267–278.
Tatto, M.T., Velez, E., 2019. Teacher education reform initiatives: The case of Mexico. In: Latin American Education: Comparative Perspectives. Taylor and Francis, New York, pp. 165–2016.
Tiempo, R.E., 2003. Docentes Desobedecen evaluación. ElTiempo.
Wayne, A.J., Youngs, P., 2003. Teacher characteristics and student achievement gains: A review. Rev. Edu. Res. 73 (1), 89–122.
Wiswall, M., 2007. Licensing and Occupational Sorting in the Market for Teachers. New York University.
Wooldridge, J., 2010. Econometric Analysis of Cross Section and Panel Data. MIT Press.