THE EFFECTS OF MERITOCRACY FOR TEACHERS IN COLOMBIA

Final Draft

Prepared for

Centro de Investigación Económica y Social - Fedesarrollo

Fondo Germán Botero Ríos

by

Alejandro Ome V.

Harris School of Public Policy Studies

University of Chicago

July 30, 2012

Contents

Executive Summary

In 2002 the Colombian government issued the *Estatuto de Profesionalización Docente* (henceforth EPD), introducing a new code for public school teachers. The EPD implements two main strategies in order to improve the quality of education; first, attract capable individuals to the teaching profession; and second, implement recurring evaluations so teachers that are not meeting the expectations are let go and accomplished teachers are promoted. In this report I ask three main questions:

- Has the EPD succeeded in attracting qualified individuals to the teaching profession?
- Has the EPD improved students' outcomes?
- Is the new system to grant promotions awarding the best teachers?

To tackle these issues I exploit the fact that when the reform was issued, incumbent teachers were allowed to remain under the old regime. This old regime provides much more labor stability, and its promotion policy is based on a fixed schedule. The gradualism of EPD's implementation causes that even by 2009 we observe both EPD and non-EPD teachers in schools, allowing comparisons between these two types of teachers over different outcomes of interest.

The main results can be summarized as follows:

1. EPD teachers have more education than comparable non-EPD teachers. Although on average EPD teachers seem to have *less* education than non-EPD teachers, I show that this is mostly due to the differences in the age distributions of teachers in the two regimes, being EPD teachers substantially younger than non-EPD teachers. When young EPD teachers are compared with young non-EPD teachers, I show that EPD teachers have more education than their counterparts in the old regime.
2. EPD teachers had had little effect on students graduating from high school in 2009, perhaps because the new policy impact students too late in their development. This contrast with findings of a companion paper where I study the effect of EPD on 5^{th} and 9^{th} graders' test scores, in this study I find positive effects especially for 9^{th} graders.

3. The system of evaluations to grant promotions seems to be awarding the best teachers in the context of 5^{th} grade evaluations. Regarding 9^{th} and 11^{th} grade evaluations I found no evidence that the new system is awarding the best teachers.

General Introduction

High quality education constitutes a fundamental milestone for development. In Colombia, despite significant improvements in terms of access to elementary and secondary education, its quality continues to be a front that needs much work. To improve the quality of teaching in public schools, in the early 2000's the Colombian government introduced a new code for teachers, the *Estatuto de Profesionalización Docente* (henceforth EPD). Among its most important features is that, in order to get hired and promoted, teachers must pass a series of examinations, a requisite absent in the previous code.

To analyze the extent to which this reform changed the status quo, it is useful to highlight the main differences between the EPD and the previous teacher's code, the Decree 2277 of 1979[1].

The first difference is the entry process. According to the EPD an individual that wants to become a teacher has to apply to calls made by departments or municipalities, give written examinations and an interview. If the candidate passes this process, he is put in probation for one year, after which another evaluation (the probation evaluation) is performed, the result of this evaluation determines whether he is placed in the ladder or let go. Although in the previous regime there was also a merit based procedure to hire teachers, in practice appointments were quite politicized and calls were infrequent (Duarte, 1996).

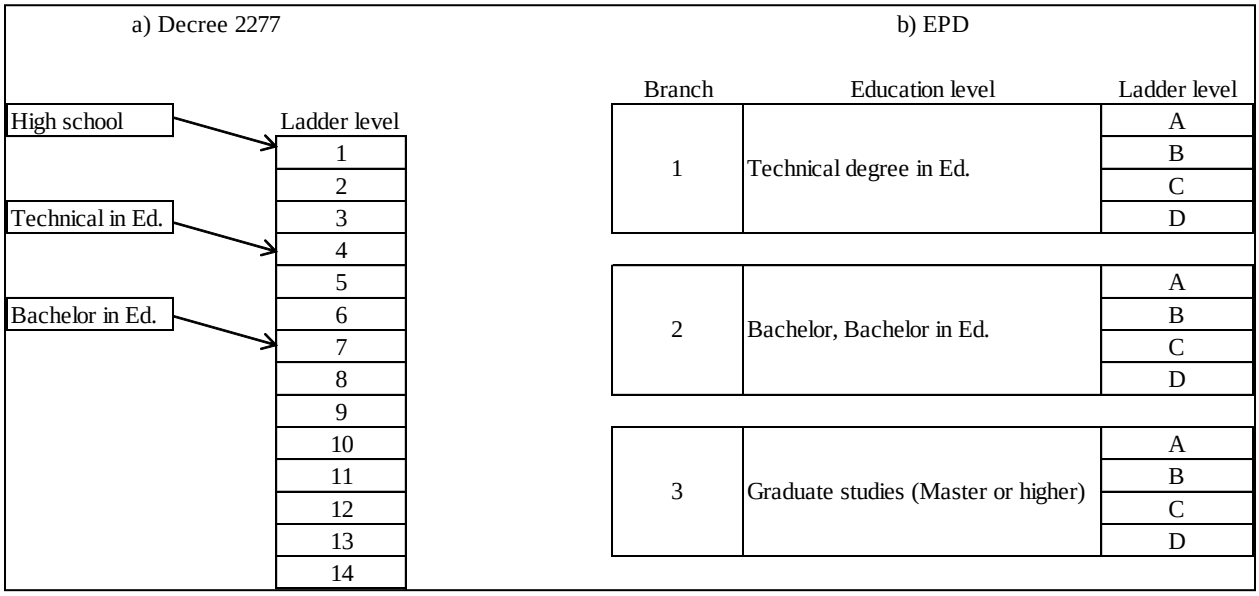
The structures of the two ladders are also quite different between the two regimes, albeit this difference is more formal than substantial. In the old regime the ladder is divided in 14 levels, and teachers enter at different levels according to her education; for example a teacher with a High School degree enters at the lowest level, while a teacher with a professional degree in Education (*Licenciado*) enters at level 7.

The EPD ladder, on the other hand, is divided in three branches, and each branch has four levels. The education level of teachers determine the entering branch; a teacher with a technical degree in Education goes to branch 1, a Bachelor goes to branch 2 and a Master or PhD goes to

[1] For a comprehensive comparison between the two regimes see Bautista (2009) and also Umaña (2004).

branch 3. To fix ideas, Figure 1 displays the basic structure of the two ladders and the corresponding education level to each entering position.

Figure 1



The third major difference regards contract termination. Teachers under the old regime enjoy great labor stability as termination can only be caused by disciplinary misconducts. The EPD, on the other hand, established a yearly performance evaluation, and if a given teacher does not pass this evaluation for two consecutive years her contract is terminated. It is important to highlight that the design and execution of this evaluation largely depends on the principals; however, the Ministry of Education provides guidelines and at the beginning of each year the evaluation needs to be discussed with each teacher. This is the most multidimensional evaluation as it can include from students' work samples to parental complaints and students' test scores.

Finally, the fourth difference has to do with promotions. Under the old regime a teacher gets promoted accumulating years of experience and additional training, and to some extent these two factors are interchangeable. Under the EPD, education and experience play a rather different role. As explained before the education level of a teacher determines at which branch is ascribed; within each branch promotions are granted if teachers pass Skills Evaluations, which are standardized written exams that evaluate teachers in pedagogy and teaching strategies. Therefore, in the context of EPS years of experience alone is not enough to get promoted, as it is under the old regime.

The described changes introduced meritocratic components to the hiring, termination and promotion policies in the teachers' career, with the ultimate objective of improving the quality of public education. Regarding the impacts of such comprehensive reform, two essential questions arise; first, has the EPD succeeded in attracting high quality teachers? and second, what have been the effects of the reform on students' outcomes? This study tries to answer these two questions.

One of the particularities of the EPD is that when the reform was issued, back in 2002, incumbent teachers were allowed to remain under the old regime, which causes that even by 2009 we observe both EPD and non-EPD teachers in schools. This situation is exploited throughout this study to estimate different constructs of interest, such as the impact of EPD on test scores.

This report contains two chapters. The first chapter analyzes the effects that EPD has had on teachers; I study to what extent EPD has succeeded in attracting teachers with more education; and also, for a given level of education, whether EPD is providing attractive wages, in comparison with what is offered to teachers under the old regime and with the wages of other workers.

The second chapter analyzes the effect of EPD on students' outcomes. In previous work I used a school-fixed effects model to estimate the effect on math and Spanish test scores in 5^{th} and 9^{th} grade, finding positive and significant effects especially for 9^{th} grade. In the present study I extend the school-fixed effects framework to analyze the impact of EPD on 11^{th} grade test scores. In addition, I analyze to what extent the instrument used to award promotions to teachers, the Skills Evaluations, had succeeded in distinguishing the better teachers, in terms of their effects on students' test scores.

1. Effects on teachers' education and wages

a) Introduction

One of the main objectives of the EPD is to attract highly qualified individuals to the teaching profession. To achieve this goal EPD demands higher educational requirements (compared to the previous regime) to become a teacher and offers substantially higher wages for teachers with graduate degrees. Keeping everything else constant, these provisions should improve the level of education of teachers. On the other hand, EPD introduced a series of teachers examinations that condition job stability and promotions; the effect of this aspect of EPD on the level of education is uncertain, it could be that more educated individuals are attracted by the meritocratic nature of EPD, but it could also be that better trained individuals will try to seek more stable jobs. This chapter analyzes whether EPD has improved or not the level of education of teachers. I also analyze to what extent the teaching profession is offering attractive wages; to do this I compare wages among teachers (EPD and non-EPD), and also teachers in general with other workers.

To analyze the effect of EPD on teachers' education I use the *Sistema Nacional de Educación Básica y Media* (SINEB), from the Ministry of Education; this dataset allows a direct comparison of teachers under the two regimes. I find that the effect of EPD on teachers' education is heterogeneous across age groups. EPD seems to have a positive effect on the education level of young teachers and very little, or even a negative effect, on the education level of older teachers. Explanations for these seemingly contradictory findings are discussed.

The SINEB is also used to analyze whether EPD had provided attractive wage profiles compared to the previous regime. I find a negative correlation between EPD and teachers wages, the main explanation for this result is that EPD teachers are just starting their careers so I only observe wages for novice teachers; hence, the only reasonable comparison that can be made at this point is between novice teachers in the two regimes, and these wages are found to be basically the same.

To compare teachers' wages with other workers' I use the National Household Survey (NHS). Because this data does not allow identification of the regime to which each teacher is ascribed to, the analysis refers to teachers in the public sector in general. I estimate the premium

(or penalty) that teachers have in terms of wages relative to comparable workers, and how this differential has evolved over time. Previous work has found that the wage premium enjoyed by public school teachers rose during the late 90's. Borjas & Acosta (2000) estimate that public school teachers earn roughly 5 percent less than workers with similar education and age in 1994; but this premium grew to 11 percent in 1998. Similarly, Gaviria & Umaña (2004) using a slightly different methodology found that public school teachers went from earning 10 percent less than comparable workers in 1994 to 8 percent more in 2000.

I contribute to this literature extending the analysis across the 2000s; furthermore, I explore a non-parametric approach that avoids imposing functional form restrictions that were present in previous work. I show that these restrictions were driving some of the findings of the cited studies. Although my results are consistent with previous work in the sense that I also find a positive trend in teachers' wage gap, my preferred estimates suggest that teachers have not received a premium really, but just that teachers were taking a penalty in the 90s and now the gap between teachers and other workers is negligible.

This chapter has five sections including this introduction. The next section reviews aspects of EPD relevant to teachers' education and wages; the third section describes the datasets employed; the fourth section presents the methods and the main results. The last section concludes.

b) EPD dispositions regarding teachers' education and wages

The introduction of EPD imposed a series of modifications to several aspects of the teaching profession, with respect to what the old regime provided. First, to apply for a teaching position under EPD a minimum of *Normalista* degree is needed, while in the previous code (issued in 1979) it was possible to become a teacher with just a high school degree. This implies that a minimum of 13 years of education are needed to have a teaching position in the EPD whereas in the previous code individuals with 11 years of education could join the profession.

Another important difference is that EPD provides substantial premiums for graduate studies, especially for Master and PhD degrees. In effect, under the EPD a novice teacher with a

Master's degree earns 60 percent more than a novice teacher with a Bachelor degree[2], while under the old regime the premium is roughly 10 percent[3].

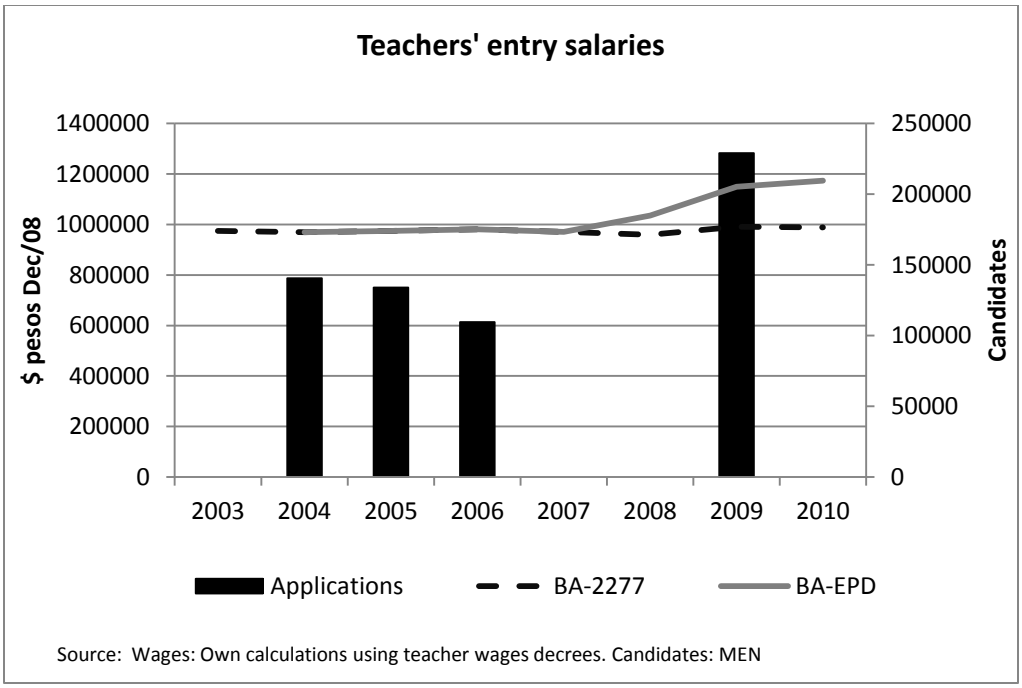
In addition to these general dispositions, over the last few years the entry level salaries in the EPD had raised substantially, which in principle should attract more individuals to the profession. With a larger pool of candidates from which to draw teachers, it will be easier to fill a given number of positions with highly qualified individuals (assuming that the distribution of the education level of applicants remains relatively constant across calls, and the demand for teachers is relatively constant too).

Figure 1 shows entry salaries for teachers with a Bachelor degree in the two regimes in Colombian pesos (henceforth COP) of 2008. The continuous line represents the entry salary for EPD teachers while the dashed line represents the entry salary for non-EPD teachers. We can see that from 2004 to 2007 the wages were basically the same but since 2008 there is a pronounced increase in the starting salary of EPD teachers. Figure 1 also plots the number of candidates of each call up to 2009, we can see that in 2009 the number of applicants rose substantially, which might in part be a consequence of the observed increase in entry salaries, although it could also simply be a consequence of the lack of calls in 2007 and 2008.

Figure 1

[2] Decree 702 of 2009.

[3] According to Decree 259 of 1981, Art. 13th, a graduate degree is equivalent to three years of experience, which corresponds to the time needed to advance one position in the ladder, which grants roughly 10 percent wage increase.

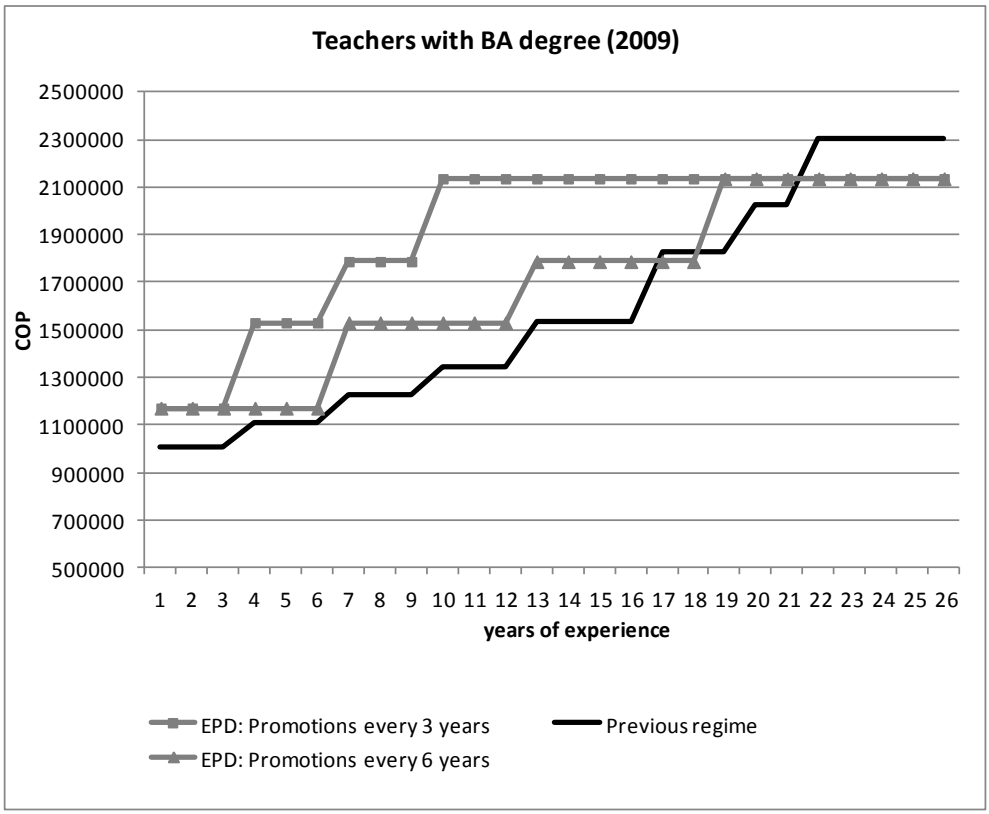


Aside starting salaries, the EPD also introduced important changes in the wage profiles of teachers across their careers. To study the differences in earnings trajectories implied by the two regimes, Figure 2 plots wages for a teacher with a Bachelor degree as a function of years of experience under the two regimes, following the wage schedules for 2009. However, because each regime provides a series of contingencies that affect the timing of promotions, it is important to describe in some detail what assumptions are behind these trajectories.

In the case of non-EPD teachers, a novice teacher with a Bachelor degree in Education (*Licenciado*) enters at level 7 of the ladder, where he earns COP 1 million per month. To get promoted this teacher can simply accumulate years of experience or, in addition to that, get supplementary training which he can substitute for years of experience. If this teacher does not substitute any training for experience, every three years he gets promoted to the next level up to 13th, where the wage is COP 2 million; if he wants to be promoted to level 14 (the last one), he will have to either get a graduate degree or write a scientific volume. This trajectory is described by the plain continuous line in Figure 2.

Now, if this teacher decides to get additional training, he will move up in the ladder faster; for example, a graduate degree will count as three years of experience, which will move him one position up; and additional courses can also be used to substitute training for experience.

Figure 2



Source: Own calculations and Decrees 1278/2002, 1238/2009, 2277/1979 and 700/2009.

Regarding EPD, the specific provisions of this code instead of accelerating promotions, in some cases could delay them. In effect, it is possible that for some teachers getting promotions will take a long time because they are required to pass skills evaluations. Clearly the extent to which skills evaluations will delay promotions will depend on how likely teachers are to pass them; considering that only a fourth passed the skills evaluations performed in 2010, it is reasonable to assume that for an important share of teachers these evaluations will be challenging.

To provide an idea of different wage scenarios that an EPD teacher can have, in Figure 2 I plotted two possible trajectories. The line with the square markers corresponds to a trajectory

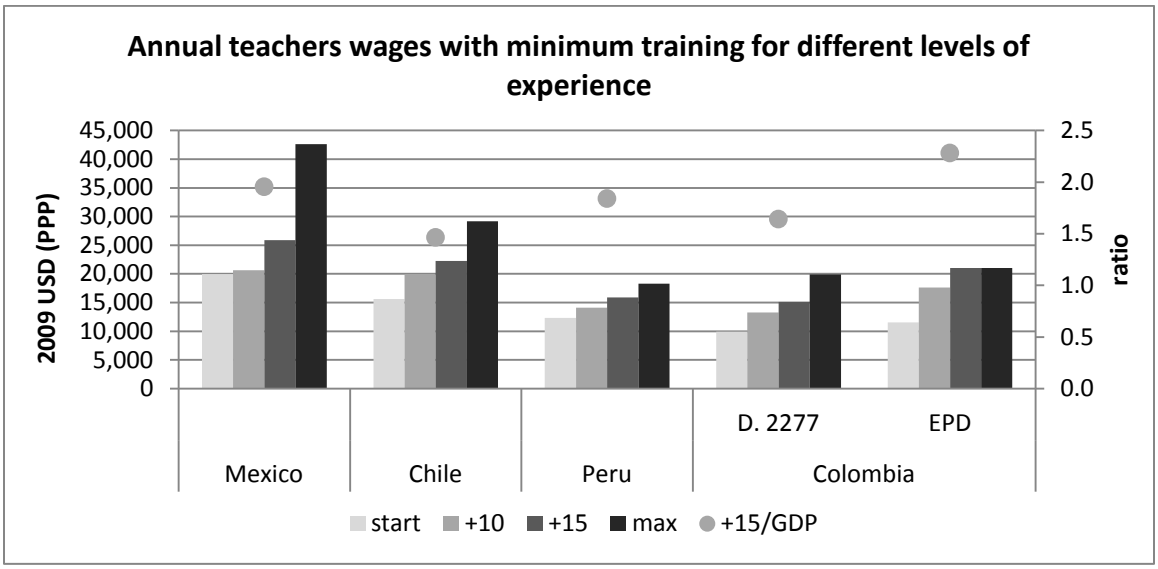
where the teacher is promoted every three years; it is apparent how EPD offers quite generous wages relative to the trajectory observed for teachers under the old regime; although both types of teachers start at a similar wage rate, the raises associated to EPD are substantially larger than those observed for the old regime, in effect, a teacher that is promoted every three years will reach the top of branch 2 in 10 years, when he will earn COP 2.1 million; while a teacher under the old regime will have to wait a lot longer to get at this wage level.

As an alternative scenario, the line with triangle markers plots the trajectory for an EPD teacher that gets promoted every seven years, this scenario provides a wage trajectory relatively similar to the one observed for the old regime, and it's probably relevant for a significant share of teachers.

In sum, the comparison of wage trajectories provides two main conclusions; first, the wage schedule seems to be substantially more attractive under EPD than under the old regime, albeit the difference largely depends on the ability of teachers to pass skills evaluations. Second, although in the medium and long-term wages could be higher for EPD teacher than for teachers under the old regime, at the beginning of their careers wages are quite similar, especially if we consider that non-EPD teachers can substitute training for experience. For this reason we should not expect that for novice teachers there should be a large and positive effect of EPD on wages. Moreover, as long as all EPD teachers are novice teachers by 2009 (in terms of their position in the ladder as the first promotion evaluation were performed in 2009-2010), we should not expect to find a positive effect of EPD on wages so far.

Finally, although the raises associated with EPD seem quite attractive, one important caveat is that the new ladder is relatively 'short', in the sense that after only three promotions a teacher would had reached the top position for a fixed level of education. Offering wage profiles with steep enough ladders is a challenge faced by many school systems. To have an idea of how the EPD looks in this front compared to other school systems, Figure 3 displays statutory wages for public school teachers in Mexico, Chile, Peru and Colombia. The bars display statutory annual wages for teachers with a professional degree across different levels of experience, namely: no experience, 10 years of experience, 15 years of experience and at the top of the scale. The circles refer to the ratio of the wage for a teacher with 15 years of experience over GDP per capita.

Figure 3



Sources: Mexico and Chile: OECD (2011); Peru: Decree 079 of 2009 and own calculations; the wage schedule corresponds to the recently launched *Carrera Pública Magisterial* for a teacher working 30 hours a week. Colombia: Decree 1278 of 2002, Decree 2277 of 1979 and own calculations; raises for teachers under the Decree 2277 might happen faster because training courses can substitute experience. Displayed raises for EPD teachers imply passing Skills Evaluations every five years.

Although Colombia's EPD offers fairly attractive mid-career wages, at least in terms of GDP per capita, an EPD teacher might end up in the top of his branch relatively early in his career, and only acquiring more education will allow him to progress more. In effect, all of the other wage schedules plotted, including the one stipulated by Colombia's old regime, offer raises even after 15 years of experience for the same level of education. Although this EPD's feature could incentivize teachers to get further education, providing that this can be very costly (not only financially but also due to psychic costs), this type of 'short' wage schedule might be less attractive for teachers than a more 'long' one.

c) Data

To analyze the effect of EPD on the education level and wages of teachers I use two datasets, the *Sistema Nacional de Educación Básica y Media*, from the Ministry of Education; and National Household Survey, from the National Department of Statistics.

The SINEB dataset contains information on all public school teachers, the variables available include age, gender, education level and whether they are under the new or the old regime.

Table 1 displays summary statistics for SINEB in 2009, separately for teachers under the old regime and the EPD. According to this data, in 2009 there were 302 thousand public school teachers[4], out of which 90 thousand (30 percent) are EPD teachers. The age distribution is quite different between the two regimes, as teachers under the old regime are older than EPD teachers. Only 2.5 percent of non-EPD teachers are younger than 35, while 41.8 percent of EPD teachers are younger than 35; on the other hand, 30 percent of non-EPD teachers are older than 55, while the equivalent figure for EPD teachers is just 2.5 percent. Although these differences in the age distribution were expected due to the fact that there should be no non-EPD teachers hired after 2002, the differences are substantial.

[4] From the original data provided by the Ministry of Education, a total of 7,517 observations are dropped mostly due to missing position in the ladder.

Table 1. Summary statistics - SINEB

	Dec. 2277	EPD	Total
Teachers	212,586	89,612	302,198
Age			
19<=age<35	2.5	41.8	14.2
35<=age<45	26.0	39.5	30
45<=age<55	41.5	16.3	34
55<=age	30.0	2.5	21.8
Total	100.0	100.0	100
Education			
HS or less	16.0	12.6	15
Technical	8.7	15.9	10.9
Professional	53.3	67.0	57.4
Postgraduate	22.0	4.5	16.8
Female	67.0	63.4	65.9
Rural	23.8	41.8	29.2
Monthly wage (pesos Dec/08)*			
<1M	7.7	16.1	10.2
>=1M <1.5M	14.3	82.7	34.6
>=1.5M <2M	45.0	1.0	32.0
>=2M	32.9	0.1	23.2

Source: Own calculations using SINEB data.

*Monthly wage is imputed using the ladder and reported bonuses.

With respect to education level there seems to be no univocal pattern that differentiates EPD and non-EPD teachers. Most EPD teachers have a bachelor degree or a technical degree[5], while teachers under the old regime are more likely to have graduate degrees, but also more likely to have only a high school diploma or less. The fact that non-EPD teachers are considerably older than their EPD counterparts might explain part of the large difference between the likelihood of having graduate degrees among the two regimes, as older teachers have had more chance to get a graduate education than younger teachers.

[5] Technical degrees include *Normalista*, a degree that special high schools grant after two years of training in education (in addition to the standard eleven years needed to get a high school diploma).

Another important difference between teachers in the two regimes is urbanicity; EPD teachers are substantially overrepresented in rural areas, which could be in part a consequence of the fact the old regime stipulates that vacant positions in urban areas will preferably be granted to more senior teachers[6].

Regarding wages the differences between non-EPD and EPD teachers are also pretty apparent. Most EPD teachers earn less than COP 1.5 million. Among non-EPD teachers, on the other hand, 45 percent earn between COP 1.5 and 2 millions, and 32.9 percent earn more than COP 2 millions.

The other dataset employed is the National Household Survey. This survey is representative at the national level and is the main instrument to measure the unemployment rate in Colombia. The survey provides information on education level, gender, age, earnings and other demographic characteristics. I use years 1995, 2000, 2005 and 2010; previous work by Gaviria & Umaña (2004) use data for September; to provide more comparable results I also use data for this month[7].

Table 2 presents summary statistics for each year, discriminating between teachers in the public sector and other workers. The sample includes private and public sector employees[8], ages 18 to 65 with valid data on monthly wages and hours worked. All summary statistics and regression results are weighted using the survey weights. Teachers in the public sector are defined as workers whose main profession is teaching and work in the public sector; although this definition includes professors at public Universities, these constitute a minority and should not affect the main results.

[6] Decree 2277 of 1979, Article 38.
[7] Starting in 2001, DANE changed the periodicity and other aspects of the NHS, turning it into a continuous survey. Although the new survey is monthly representative at the national level, the sample sizes for September are quite smaller than before 2001 because then the whole national rural survey was collected in September. For this reason, for 2005 and 2010 instead of using only the September data I included August and October, which allows having a larger sample but with low risk that any seasonal pattern will affect the main results.
[8] Workers excluded are Self-employed, employers and domestic employees.

Table 2. Summary statistics - National Household Survey. September survey for various years*.

	1995		2000		2005		2010	
	Teachers - Public Sc.	Other Workers	Teachers - Public Sc.	Other Workers	Teachers - Public Sc.	Other Workers	Teachers - Public Sc.	Other Workers
N	981	23,534	1,223	25,042	1,324	25,616	1,719	34,112
Age								
y18_24	8.6	24.6	5.4	23.0	1.7	21.3	1.9	19.8
y25_34	26.6	35.5	23.1	34.5	16.2	33.3	15.3	33.8
y35_44	35.9	23.1	33.7	25.3	39.2	26.2	28.7	24.8
y45_54	23.7	11.5	30.8	12.5	34.6	14.4	35.1	15.9
y55 and over	5.2	5.2	7.1	4.8	8.4	4.8	19.0	5.7
Years of Education								
1:<11	2.1	65.2	0.7	55.1	0.3	49.4	0.0	43.0
2:>=11 <16	58.1	28.2	36.0	34.6	21.1	40.2	12.9	46.3
3:16	35.2	5.7	40.3	7.6	53.8	7.7	42.7	6.5
4:>16	4.6	0.9	23.0	2.8	24.9	2.7	44.4	4.2
Female	67.4	35.6	66.0	40.0	63.2	41.4	64.3	40.6
Head of household	37.5	44.5	42.9	42.7	49.7	45.8	53.6	47.6
Rural	20.9	25.2	20.6	21.0	7.9	13.8	9.7	16.6
Household characteristics								
Presence of Children (0-12yo)	63.4	66.0	56.7	64.3	59.2	63.6	49.0	58.2
Presence of elder (>64yo)	15.3	15.7	16.2	16.2	14.5	13.8	15.5	13.5
Presence of other breadwinner	79.4	72.5	73.2	71.3	70.1	70.5	63.4	69.5
Monthly wage (pesos Dec/08)								
<1M	68.7	87.7	42.1	86.3	41.5	87.7	13.8	85.4
>=1M <1.5M	20.3	5.8	27.9	6.6	31.1	6.3	39.5	7.9
>=1.5M <2M	7.2	3.2	15.0	3.2	16.8	2.7	24.6	3.1
>=2M	3.8	3.4	15.0	3.8	10.6	3.2	22.1	3.6

Source: Own calculations using NHS.

*Starting in 2001 DANE changed the periodicity and other aspects of the NHS, turning it into a continuous survey. Although the new survey is monthly representative at the national level, the sample sizes for September are quite smaller than before 2001 because then the whole national rural survey was collected in September. For this reason, for 2005 and 2010 instead of using only the September data I included August and October, which allows having a larger sample but with low risk that any seasonal pattern will affect the main results.

We can see that teachers are relatively older than other workers, a pattern that seems to be reinforcing over time, probably due to the very little hiring of teachers that took place in the early 2000s, after the EPD was issued (in 2002) but before the first call for teachers was made (late 2004).

Regarding education level, NHS does not provide data on the last level approved but only the number of years of education[9]; to analyze the distribution I divided individuals by using the thresholds relevant to High School and Bachelor degrees (in Colombia a High School diploma is earned after completing eleven years of education, and most Bachelor programs last five years). Teachers have substantially more education than other workers, most teachers have at least 11 years of education, and an increasing fraction has 16 or more years of education. Most of the rest of workers, on the other hand, have less than 16 years of education, and even as late as 2010 near half has less than 11 years of education.

Females are overrepresented in the teachers group, and the fraction of teachers that are also household head has been roughly 45 percent over the analyzed years. Household characteristics, such as presence of children under 12 years old, presence of elderly people and presence of a second breadwinner in the household seem to be relatively stable over time and fairly similar between teachers and other workers.

Finally, teachers have higher wage profiles than the rest of the workers, which is consistent with the substantial differences in educational attainment and the age distribution.

d) Empirical methods and results

Education level

The SINEB data allows direct comparison of education level between EPD and the rest of teachers. The basic objective is to estimate the effect of EPD on years of education, controlling for age and other characteristics. The proposed model is described by:

$$years\_educ_i = \alpha + \gamma EPD_i + \mathbf{x_i'}\boldsymbol{\beta} + \varepsilon_i \quad (1)$$

Where $years\_educ_i$ represents the years of education of teacher *i*; EPD_i indicates whether the teacher is ascribed to the new code or not; $\mathbf{x_i}$ is a vector of control variables; ε_{it} is an error term and α, γ and $\boldsymbol{\beta}$ are parameters to be estimated. The coefficient of interest is γ, which captures the correlation of EPD on years of education.

[9] Since 2008 the survey actually provides both last level approved and years of education, to keep consistency with previous years of the survey I use only years of education.

Because SINEB reports last degree achieved rather than years of education, the dependent variable is categorical rather than continuous; as this type of variables do not have a straightforward cardinal meaning is better to use multinomial models rather than standard OLS analysis. In Table 3 the first column presents results of an ordered probit model where the underlying latent variable is assumed to be generated by the process described in (1); the control variables included are age, age squared and dummies for rural areas, gender and the four main geographic regions in the country. We can see that the coefficient of EPD is negative but not significant, which indicates that there seems to be no differences in the education levels between EPD and non-EPD teachers. This is rather surprising as EPD was especially designed to professionalize teachers, as its very name indicates.

Table 3. Education level on EPD-Ordered probit

	All	Age groups			
		19-35	35-45	45-55	>55
	(1)	(2)	(3)	(4)	(5)
EPD	-0.039	0.380	-0.006	-0.210	-0.280
	(0.060)	(0.105)	(0.071)	(0.059)	(0.095)
Obs. EPD	89,612	37,417	35,369	14,596	2,230
Obs. Total	302,198	42,788	90,606	102,798	66,006

Standard errors clustered at the department level in parentheses

Note: All regressions include age, age squared and dummies for rural areas, gender and the four main geographic regions in the country.

Source: Own calculations using SINEB data.

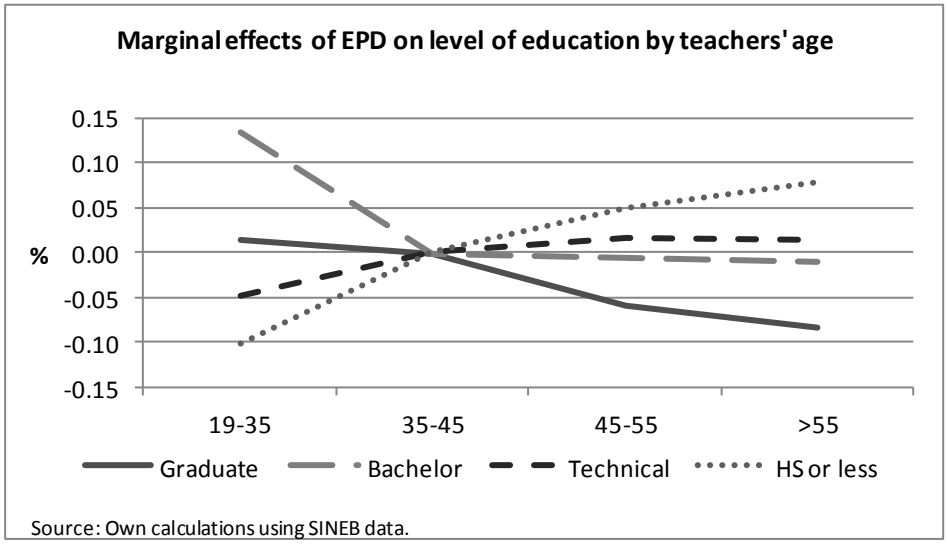
It is possible that this is due to the substantially different age distribution that teachers have under the two regimes. Although in principle this should be controlled by the inclusion of age as a covariate in the model, including this variable in the regression does not completely offset its impact if EPD has heterogeneous effects across age profiles.

To analyze this possibility, columns 2 to 5 present results of the same model than in column 1, splitting the sample in four age groups. We can see that the effect of EPD on education level is positive and significant for teachers 18 to 35 years old, negative but not significant for 35-45, and negative and significant for the remaining two age groups. This suggests that the effect of EPD is heterogeneous across age groups; among young teachers those

that are ascribed to the EPD have more education. However, for older teachers the effect has the opposite sign.

To provide an idea of the size of these estimates, Figure 4 presents marginal effects evaluated at the mean of the covariates. EPD teachers under 35 are 13 percentage points more likely to have a Bachelor degree than non-EPD teachers on the same age group, but this coefficient declines with age and is basically zero for teachers older than 35. Regarding graduate studies, the effect of EPD is equal to zero for teachers under 45, but negative for the rest.

Figure 4



The negative effect of EPD found for older teachers can be explained by the fact that most teachers that have graduate studies are already in high positions of the old ladder, and if they switch to the new code they will have to start from the bottom of the respective branch, which might not be very attractive. For example, a teacher in level 12 of the old ladder with 1-year graduate degree earns COP 1.8 million in 2009; if he switches to the new code he will fall in branch 2, where, with a 1-year graduate degree (*Especialización)* he will earn COP 1.2 million; unless this teacher is very confident that he will get promoted quickly (which is conditioned on passing the skills evaluations), is not going to be very attractive to switch to the EPD.

Also, it is important to highlight that the EPD teachers older than 45 are a minority among the EPD teachers (18.8 percent). These teachers are most likely individuals that worked

many years as teachers under temporary contracts that had finally being appointed through the EPD calls, or individuals that come from different professions. In the long run it is probably more valuable to focus on the education level of young EPD teachers than on the minority of older EPD teachers.

The prevalence of graduate studies among teachers under the old regime observed in the summary statistics, suggests that the results from the ordered probit model might be driven by this alone; along this line of thought, it is interesting to analyze simply the effect of EPD on the probability of having a Bachelor and graduate studies independently. Table 4 shows results for equation (1) run separately for two outcomes: whether the teacher has a Bachelor degree or higher (first column) and whether the teacher has a graduate degree (second column). The results correspond to a Linear Probability Model, so the coefficients can be directly interpreted in terms of percentage points.

Table 4. Education level on EPD - LPM

	Bachelor	Grad. Degree
EPD	0.061	-0.098
	(0.027)	(0.022)
Dep. Var. Mean	0.74	0.17
N	302,198	302,198

Standard errors clustered at the department level in parentheses

Note: All regressions include age, age squared and dummies for rural areas, gender and the four main geographic regions in the country.

Source: Own calculations using SINEB data.

We can see that EPD has positive effect on the probability of having a Bachelor degree, EPD teachers are 6.1 percentage points more likely to have a Bachelor degree or higher than non-EPD teachers. At the same time, EPD teachers are 9.8 percentage points less likely to have a graduate degree.

The effect of EPD seems to be heterogeneous across two dimensions, age and type of degree. In effect, these two forms of heterogeneity might be two sides of the same coin. Teachers under the old regime are older than EPD teachers, hence the former had had more time to get a graduate education, for this reason the majority of teachers with a graduate degree are under the

old regime, and for this EPD seems to have a negative effect on graduate education, but among novice teachers the effect on education is positive.

Bonilla & Galvis (2011) reach a rather different conclusion finding that EPD is negatively correlated with teachers' education, these authors use a dataset that does not allow analysis by age which makes it impossible to compare the effect of EPD across different age profiles.

Wages

As noted in the SINEB summary statistics, teachers under the old regime have higher wages than EPD teachers. This might be driven by the fact that non-EPD teachers are older so they have more experience and education. To analyze to what extent wage differentials are associated to education and experience, Table 6 presents regression results of the logarithm of monthly wages on education level, experience[10], experience squared, gender, four dummies for education level, one dummy for whether the degree is in Education, a dummy variable for rural areas, and 32 dummies for departments. In column 1 the results for the whole sample are displayed, we can see that non-EPD teachers have higher wages even after controlling for education and experience; the results indicate that non-EPD teachers earn 13 percent more than EPD teachers.

The coefficients of the displayed control variables (education and experience) have the expected sign and are statistically significant. For education level the left out category is having a Bachelor degree, so a Technical degree or lower has a negative effect on wages, while a graduate degree has a positive effect. Experience has a positive and decreasing effect.

[10] Experience is calculated using the date when the teacher was ascribed to the ladder. Using potential experience instead (age-years_of_education-6) does not change the main results.

Table 6. Log-wages on EPD

		Age groups				Future
	All	18-35	35-45	45-55	>55	wage*
	(1)	(2)	(3)	(4)	(5)	(6)
EPD	-0.129	0.0002	-0.083	-0.208	-0.346	0.027
	(0.024)	(0.033)	(0.025)	(0.025)	(0.031)	(0.024)
HS or less	-0.258	-0.213	-0.300	-0.234	-0.266	-0.273
	(0.043)	(0.014)	(0.046)	(0.050)	(0.050)	(0.043)
Technical degree	-0.208	-0.195	-0.188	-0.191	-0.289	-0.217
	(0.025)	(0.009)	(0.028)	(0.033)	(0.036)	(0.024)
Grad. Studies	0.128	0.078	0.137	0.121	0.111	0.073
	(0.012)	(0.012)	(0.013)	(0.011)	(0.012)	(0.011)
Experience	0.026	0.009	0.022	0.020	0.012	0.026
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Experience sq.	-0.00035	0.00019	-0.00011	-0.00024	-0.00018	-0.00039
	(0.00006)	(0.00015)	(0.00010)	(0.00007)	(0.00003)	(0.00005)
Obs. EPD	89,612	37,417	35,369	14,596	2,230	89,612
Obs. Total	302,198	42,788	90,606	102,798	66,006	302,198

Standard errors clustered at the department level in parentheses.
Note: All specifications include dummies for whether the degree is in Education, gender, rural areas and 32 departments.
*The dependent variable is the wage of each teacher, assuming that each teacher is moved one step up in his ladder.

There are a couple of considerations that need to be taken into account to properly analyze these results. Once again, while EPD teachers are just starting their careers, non-EPD teachers have had the opportunity to progress for many years; so the age and experience distributions are quite different. This creates a similar problem to the one observed in the education analysis, because the majority of EPD teachers are young and the majority of non-EPD teachers are relatively old.

To study to what extent EPD has different effects across different age profiles, columns 2 to 5 of Table 6 present regression results for four age groups. We can see that the effect for teachers between 19 and 35 years old is positive but not significant, while the effect for the rest of the age groups is negative, significant, and increasing (in absolute value) with age.

The institutional context described in section 2 also helps explaining these heterogeneous results. With respect to the nil effects for young teachers, it is important to remember that

although the entry salary of teachers with a Bachelor degree is higher for EPD teachers than for a non-EPD teachers, non-EPD teachers are able to substitute training for experience in a way that EPD teachers are not, so it is probably not too surprising that the effect on young teachers is so close to zero.

Furthermore, although statutory salaries for EPD teachers up in their ladder are in general higher than salaries for non-EPD teachers associated to similar levels of experience, the fact that skills evaluations for EPD teachers started as late as 2009 (and the evaluation itself was fielded in 2010) implies that all the wages we see in this data (which corresponds to 2009) consider only EPD teachers in the bottom of the ladder, regardless of their level of experience. For this reason, when we compare older teachers the effect of EPD is negative.

It is also interesting to notice how the coefficients on the control variables change across age groups. While the effects of education level remain relatively stable, the effect of experience is substantially lower for young teachers; considering that by 2009 no promotions had been granted to EPD teachers, and that most young teachers are under the EPD, it is straightforward to see that the returns to experience for young teachers up to 2009 should be zero.

It is apparent that the presented results on wages are highly driven by the youngness of EPD itself. To have an idea of what could be the scenario in a few years where EPD teachers had had the chance to progress in their ladder, the last column of Table 6 presents results of a regression where the dependent variable is the log-wage, assuming that all teachers (EPD and non-EPD) have been moved to the next position in their respective ladder. We can see that the coefficient on EPD is now positive although not significant. Considering that this back of the envelope simulation moved only one position (so for each level of education, teachers in EPD have two more levels to get promoted), it is clear that the effect of EPD will change substantially when we start to see more EPD teachers progressing in their careers.

Teachers' wages compared to other professions

Previous work on teachers' wage differentials in Colombia had employed full parametric approaches; in particular, Borjas & Acosta (2000) and Gaviria & Umaña (2004) estimate Mincer type equations including dummy variables for teachers. The main limitation of this method is that in a context where the distributions of observables characteristics are as different as they are

between teachers and other workers, assuming a specific functional form for the data generating process is probably not appropriate.

Although the problem is similar to the one described in the previous section when I compare education and wages between EPD and non-EPD teachers, in that context it was probably easier to get around it by splitting the regression across different age groups, as the major difference was in the age distributions.

In the context of comparing teachers and other workers, however, the differences are salient not only across the age distributions, but also in education level and gender. To avoid placing functional form assumptions I follow a methodology proposed by Ñopo (2008) to decompose wage gaps that extends the well known Oaxaca-Binder decomposition to a non-parametric approach. This methodology is applied by Ñopo & Mizala (2011) to analyze teachers' wage gaps in various countries in Latin America, which facilitates comparisons with other countries in the region.

The basic idea behind this decomposition is to calculate wage differentials only over comparable individuals. In the context of this work, each teacher is exactly matched (across sociodemographic characteristics) to a cell of other workers, and the differences of wages are integrated across the distribution of teachers' characteristics.

An important limitation of this approach is that the comparison is going to be restricted to those teachers within the common support area. In other words, the estimated results are going to be relevant only for the teachers that can be exactly matched. To have an idea of the magnitude of this caveat, the percentage of teachers that are actually matched is displayed for each of the specifications.

The other issue that needs to be addressed is the comparison group. Considering that the wage gap estimations can be sensible to the use of different comparison groups (which, as I show later, is true for some specifications), I present results for two different sets of workers. The first one includes all other workers earning a wage, the second group includes only Professionals and Technicians.

Table 6 shows separate results for years 1995, 2000, 2005 and 2010. The first row presents estimates for earnings equations over the whole set of workers. The control variables included are education, experience and experience squared, gender, and dummy variables for whether the teacher is the household head, presence of children younger than 12 years old, presence of adults older than 65 years old, presence of a second earner in the household and part-time work (less than 30 hours per week). According to this first specification teachers in 1995 were receiving a premium of 4 percent on their hourly wage, and this premium has been rising consistently over the years.

Table 6. Public school teacher's premium in hourly wages

		1995	2000	2005	2010
Earnings equation					
All workers	Coef.	0.04	0.26	0.32	0.51
	se	(0.02)	(0.02)	(0.02)	(0.02)
	N	24,515	26,265	26,940	35,831
Professionals and	Coef.	-0.16	0.03	0.02	0.12
Technical workers	se	(0.02)	(0.03)	(0.02)	(0.02)
	N	2,742	3,267	3,653	5,360
Oaxaca-Ñopo					
All workers	Unexplained Gap	-0.22	0.03	-0.01	0.11
	se	(0.07)	(0.07)	(0.03)	(0.04)
	Matched teachers (%)	82	81	88	78
	N	11,536	9,408	11,308	13,460
Professionals and	Unexplained Gap	-0.28	-0.17	-0.13	0.03
Technical workers	se	(0.06)	(0.09)	(0.04)	(0.04)
	Matched teachers (%)	66	70	82	70
	N	2,106	2,335	2,805	3,671

Notes: Standard errors for the Oaxaca-Ñopo estimator are bootstrapped (100).

When we focus on the more selective comparison group of Professionals and Technicians, we can see that back in 1995 teachers were actually penalized, as they earned 16 percent less than comparable workers. However, consistent with the previous specification, a clear trend can be observed in favor of teachers, and in 2010 we can see that teachers seem to be earning 12 percent more than comparable workers.

The third and fourth row show results for the Oaxaca-Ñopo estimator. In this context the results are more similar between the different comparison groups. When the comparison group includes all other workers the wage gap is negative in 1995, not significant in 2000 and 2005 and positive and significant in 2010. According to these results teachers earned 22 percent less than comparable workers in 1995 but in 2010 they earned 11 percent more. Finally, using the Professional and Technical workers as a comparison group, we can see that the wage gap has been decreasing consistently over time, in 1995 teachers were earning 28 percent less than comparable workers, but by 2010 the difference is positive although very small and not significant.

Although the different methodologies over the different comparison groups portrait rather different results in terms of levels, the trends on the teachers' wage gap is pretty stable; in short, in the mid 90's teachers were taking a penalty in their wages' compared to other workers, but this penalty had been reducing consistently over the last 15 years and, for the non-parametric approach using the most selective comparison group, the difference is basically negligible in 2010.

Comparing the results under the two specifications it is apparent how the non-parametric decomposition is substantially less sensitive to the comparison group employed than the earnings equation approach. The reason for this is that the non-parametric method only uses as counterfactuals exact matches, while the earnings equation uses all the observations in the comparison group. In other words, even under the first comparison group the second approach is already discarding a lot of observations that are not comparable (less educated individuals, for example) so when the most selective comparison group is employed the changes in the results are not as dramatic as when the earnings equation method is used.

Was there ever a wage gap?

The documented decline in the wage differential suggests that labor conditions had improved for teachers; however, this interpretation overlooks other important changes that had taken place in the teachers' career in the last few years. In particular, perks in the social security regime (e.g. early retirement at age 50 and exceptions to social security contributions) are now

fading away or not in place at all. In this sense it can be argue that while wages have been improving the social security regime has gradually been homogenized with that of other workers.

Reducing the distortions inherent to the social security system that teachers had and at the same time making their wages more competitive will probably make more clear the costs and benefits of the teachers' career.

Wage gaps in other countries

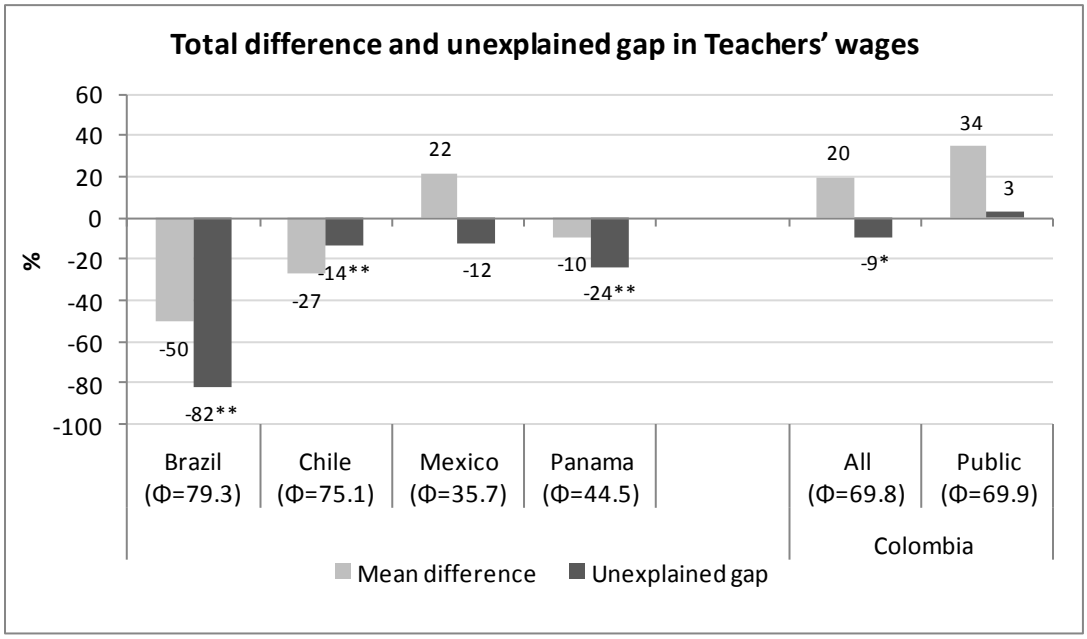
To put in context the wage differentials calculated in the previous section with those observed in other countries, I compare some of my estimates with the results from Mizala & Ñopo (2011), who also follow Ñopo (2008) methodology to decompose wage gaps.

Figure 4 presents results for Brazil, Chile, Mexico and Panama, the four countries with highest percentage of matched teachers in the mentioned study. Household surveys for each country were performed in the late 2000s, and sample sizes were roughly 25 thousand for Brazil, 10 thousand for Chile, 35 hundred for Mexico and 23 hundred for Panama. Ñopo & Mizala (henceforth ÑM) restricted the comparison group to Professionals and Technicians.

An important difference between the estimates from MÑ and those presented in the previous section is that MÑ's correspond to wage gaps for school teachers in general, while I focus on public school teachers. For this reason for the case of Colombia Figure 4 displays results using teachers in general (labeled 'All'), and another focusing on public school teachers (labeled 'Public'). To display the most comparable results, the estimates I display correspond to 2010, using the Professionals and Other technical workers as comparison group[11].

Figure 5

[11] Unfortunately the Colombian NHS does not use the same classification that is used in the countries that ÑM analyzed, so the comparison groups are not exactly the same. ÑM use as counterfactual groups 2 and 3 of the *Clasificación Internacional Única de Ocupaciones* (CIUO), which correspond to Professionals and Scientists and Technicians and Mid-level Professionals respectively. The Colombian NHS uses an older classification and the most similar category is called Professionals, Technicians and Other workers.



Sources: Brazil, Chile, Mexico and Panama: Mizala & Ñopo (2011). Colombia: Author's calculations. Φ denotes the percentage of teachers in the sample that are matched. Significance levels are displayed only for the unexplained gaps. (* $p<0.05$ ** $p<0.01$)

Except for Mexico and Colombia, teachers earn less than other workers when we look at the difference in unconditional means. With the exception of Chile, controlling for teachers' characteristics deepens the wage gap. Although in Colombia teachers in general seem to be taking a penalty on their wages of 9 percent, when we focus on public school teachers the difference is actually positive but not significant. This suggests that most of the underpayment that teachers are receiving is being driven by private school teachers, but the phenomenon is not affecting public schools teachers too much.

e) Summary and final comments

The evidence described in this chapter suggests that EPD has succeeded in attracting more qualified teachers to the teaching profession. Although at face value it seemed that EPD teachers have less education than teachers under the previous regime, when the analysis is performed by age groups I find that among novice teachers, EPD teachers are more likely to have a Bachelor degree. Along these lines I argue that the reason why teachers under the

previous regime seem to have more education is that older teachers have had more time to complete graduate studies, something that EPD teachers probably will start doing soon.

Regarding wages the comparison between EPD and the rest of teachers is complicated by the fact that the former are all in the first step of the ladder in 2009, while most of the latter are at advanced stages of their careers. When the comparison is made between novice teachers in both regimes there seems to be no difference in wages; considering that statutory wages for EPD teachers have a steeper trend than the schedules of the previous regime (at least in the early stages of the career), it can be expected that in the near future wages for EPD teachers will start looking higher than those of the previous regime.

With respect to the comparisons between teachers' wages and other workers' the sensibility to the different comparison groups employed reduces dramatically when the Oaxaca-Ñopo estimator is employed. I found no robust evidence that suggests that public school teachers are either over or under paid. Using as a comparison group all other workers and the Oaxaca-Ñopo estimator I found a wage gap in favor of teachers of 11 percent, however, the difference is negligible when the more selective comparison group is employed.

It is important to highlight that the estimated wage gaps are only relevant across the common support; further analysis of the implications of this limitation will improve our understanding of how competitive are teachers' wages in Colombia.

2. Effects on students' test scores

a) Introduction

The ultimate purpose of EPD is to improve the education of students. To achieve this goal EPD put in place a series of merit based components in the teachers' career. In particular, the selection process tries to select the most capable individuals; the Performance Evaluations prevents that teachers that are not meeting the expectations are kept in the payroll; and the Skills Evaluations rewards teachers' abilities.

Clearly, all these evaluations are far from being perfect instruments. A few written evaluations and an interview might not be enough to properly value the quality of a given candidate; the Performance Evaluations can be subject to favoritisms or discrimination; and the Skills Evaluations may not measure the ability of a teacher to educate but just his ability to give exams. At the end of the day, the important issue is whether EPD had led to better students' outcomes or not. This chapter analyzes the effect that EPD has had on 11^{th} grade test scores.

To perform this evaluation, I exploit the fact that teachers that were laddered before 2002 were allowed to remain under the rules of the old teacher's code, the Decree 2277 of 1979. This causes that the percentage of EPD teachers varies across schools. This study uses this source of variation to identify the effect of EPD on students' outcomes.

A naïve approach would be to estimate the correlation between the percentage of EPD teachers with test scores at a given year. The problem with this method is that unless students or teachers are randomly distributed, a correlation between the share of EPD and test scores will be a biased estimate of the effect of EPD. For example, if EPD teachers are more likely to be sent to low performing schools, then even if EPD has zero or positive effect, the coefficient will probably be negative because it will pick up the fact that the schools to which EPD are being sent are low performing schools per se.

To tackle this identification problem I propose a school-fixed effects model, which controls for time invariant characteristics at the school level. Providing that the students' characteristics that are correlated with the presence of EPD teachers are time invariant, this model identifies the causal effect of EPD on students' outcomes.

In previous work I used this method to analyze the effect of EPD on dropout rates in Elementary and Secondary schools (Ome, forthcoming). I found that EPD teachers are negatively correlated with dropout rates in elementary and secondary schools. In the same study I also analyzed the effect of EPD on test scores for 5^{th} and 9^{th} grade math and Spanish (henceforth SABER5&9), where I found positive effects, especially for 9^{th} graders. In the present study I extend the analysis to 11^{th} grade test scores (SABER11).

SABER11 has some advantages and some disadvantages with respect to SABER5&9. First, SABER11 had been for many years the primary source of data regarding students' achievement (for example, all papers that analyze achievement using Colombian data that are referenced in this study use SABER11), the importance of this exam resides in that is needed to apply to higher education, so the exam is performed every year and most students that reach 11^{th} grade take it; SABER5&9, in turn, is designed to provide a picture at the school level, so not all students take the exam, and the exams are performed every three years (for more details see Ome (forthcoming) and the references listed there). In addition, SABER11 includes sociodemographic information at the student level that is not available for SABER5&9.

On the down side, SABER11 is preformed at the end of high school which means that the results are relevant only for the individuals that reach this point, which in Colombia may be less than 62 percent[12].

I find no effect of EPD on SABER11. One possible explanation for this result is that the underlying distribution of ability of students is changing thanks to the very EPD. As mentioned before, in previous work I found that EPD had a negative effect on dropout rates in Elementary and Secondary schools, which implies that in schools where there are more EPD teachers there are also students that in the absence of EPD teachers would had dropout from school; if these students that stay in school thanks to the presence of EPD teachers are less well endowed than the average student in the school, then the fixed effects estimate of EPD on SABER11 is going to be biased because is going to pick up the change in the distribution of ability caused by the effect of EPD on dropout rates.

[12] According to the National Household Survey of 2009 (September), as many as 38 percent of individuals between 20 and 25 years old reported to have less than 11 years of education.

Another explanation can be that 11th graders in 2009 had little exposure to EPD, and were exposed in a developmental age when probably any treatment is less likely to have an important impact (compared to 5th and 9th graders).

Finally, I analyze to what extent the Skills Evaluation, the instrument designed in the EPD to grant promotions to teachers, is effectively distinguishing the teachers that add more value to their students. For this, I extend the fixed effects model presented before to incorporate an additional effect for schools where EPD teachers achieve a high score on the Skills Evaluations.

I implement this method across all three SABER examinations (5th, 9th and 11th grades). For 5th grade I find that students in schools with EPD teachers that do well in the Skills Evaluation observe better results than students in schools with low performing EPD teachers. For 9th and 11th grade, however, I found no evidence that the Skills Evaluation is helping detecting the better teachers.

This chapter has six sections including this introduction. The next section discuss relevant studies on the effect of meritocracy for teachers; the third describes the data employed; the fourth section presents the empirical framework and the main results, the fifth section extends the analyses to study to what extent Skills Evaluations are rewarding better teachers or not. The last section presents some final comments.

b) Literature review

The introduction of a new teachers' code can affect their outcomes through two main channels. The first one is changing the behavior of a given teacher (behavioral effect), the second one is by attracting a different type of professional to the teaching profession (sorting effect)[13]. It is important to distinguish these two mechanisms as the empirical literature is not always specific about which type of effect is being estimated.

In particular, studies that randomized students or teachers into merit pay schemes in general do not estimate the sorting effect because selection is determined randomly. The purpose of using random assignment is to avoid selection of students; however, by randomizing the effect

[13] For a discussion on this see Lazear (2003) and also Woessmann (2010).

of self-selection of teachers is also offset which impedes its estimation. These experiments provide mixed evidence at best of the effects of merit pay on students' outcomes. Muralidharan et al. (2009) find some positive effects of merit pay for teachers on students' test scores in an experiment run in India. Glewwe & Kremer (2003) analyze a similar intervention in Kenya; although these authors found positive effects right after the intervention, the effects vanished just a few years later. In Nashville (US) Springer et al. (2010) analyze a large randomized intervention finding no effects of performance pay on students' test scores.

Studies that do not rely on random assignment can potentially identified the sorting effect; the challenge these studies face is how to control for selection of students. For example, if abler students select into schools under merit pay, then the effect of merit pay will be confounded with the fact that abler children (or their families) tend to select themselves into schools with this type of pay scheme.

Dee & Keys (2004) study the implementation of a merit pay system in Tennessee in the mid 80's relatively similar to the EPD. The authors exploit a parallel school program in Tennessee that randomly assigned pupils to classrooms of different sizes to evaluate the effect of class size over achievement. The fortunate design of this study allowed controlling for selection of students (as they were randomly assigned to classrooms) but teachers were not randomly assigned to merit pay. They found that students that had a teacher on performance pay had higher test scores for math but no significant effect was found for reading.

Regarding the Colombian context, in a previous study I implemented a school-fixed effects model to estimate the impact of EPD on dropout rates in Elementary and Secondary schools. The estimated effects bundled the behavioral and sorting effect as selection of teachers was not randomized; regarding selection of students, the fixed effects model controls for it assuming that the characteristics that determine this selection process are time invariant. I found that EPD reduces dropout rates in elementary and secondary schools. Using the same framework, I also analyzed the effect of EPD on standardized test scores for 5^{th} and 9^{th} grade math and Spanish, where I found a positive effect 9^{th} grade test scores.

Although no other study has addressed the impact of EPD itself, in Colombia a few studies had analyzed the effects of the previous teachers' code in comparison to other

institutional arrangements. Barrera-Osorio (2006) compares chartered and traditional public schools in Bogotá in the early 2000's, he finds that chartered schools observe lower dropout rates and higher test scores than comparable public schools.

Núñez et al. (2002) compare test scores from traditional public schools with public schools managed by the church. These authors find that schools managed by the church outperformed the standard public schools, highlighting the importance of the institutional arrangement.

Furthermore, Gaviria & Barrientos (2001) analyze the impact of schools' characteristics (e.g. infrastructure, teacher/pupil ratio) on test scores. They find that while in private schools better characteristics are associated with better outcomes, this correlation is not observed in public schools, which suggests that the incentive structure plays a major role in determining the productivity of school inputs.

c) Data

To analyze the effect of EPD on 11^{th} grade test scores I use two datasets, the C-600 forms and the SABER11 data. All the descriptive statistics and regression results refer to data on public schools.

The C-600 dataset contains administrative information of all schools, such as number of students and teachers' education level; this information is collected every year through forms filled by each school in the country. Part of this data is at the Institution level, which is an administrative unit that can encompass one or more schools or campuses; in addition, due to infrastructure restrictions many schools serve multiple shifts during the day, in general one group of children in the morning and another in the afternoon. Therefore, one institution can have more than one school, and one school can have more than one shift. As aforementioned, some data in the C-600 files are at the Institution level, but most of it is at the school-shift level. To avoid bundling shifts into schools, the C-600 is used at the school-shift level.

The SABER 11 exams are presented yearly by all students finishing high school and it evaluates knowledge in math, language, natural and social sciences, philosophy and English.

Aside test scores this data contains sociodemographic information of each student presenting the exam, including gender, age, education of the parents and household income.

The years analyzed are 2002, the baseline, and two years of follow-up, 2008 and 2009. Although there is SABER11 data for all years between 2002 and 2009, the selection of these years obeys mostly to data availability issues[14].

To build the panel I link the C-600 and the SABER11 datasets using a school-shift id number, although for some cases I only used the school number[15]. A school is considered to be part of the panel if it has complete information (SABER 11 and C-600) for at least two years, this way it contributes to the fixed effects model. Because not all schools are in the panel (some schools are observed only in one year, or are observed in various years but not across datasets)[16], it is a source of concern that the schools in the panel share characteristics that could be correlated with the presence of EPD teachers, this could bias the estimate of the effect of EPD teachers as the regression coefficient will confound the effect of EPD with the effect of the characteristics that made schools select into the panel sample. To analyze this problem in Table 1 I compare to what extent test scores and other variables are different between the panel sample and the rest of the schools.

[14] In particular, most years between 2002 and 2008 do not have all the sociodemographic data that I need for the specification; also, C-600 data is not available in the 2003-2004 years.

[15] The merging process that uses a school-shift id can be characterized as a 'hard match', as the probability that a pair of schools is erroneously matched is low. However, this hard match also leaves a large number of schools unmatched in both the C-600 and SABER11 data. An important share of schools that are not 'hard matched' can be matched when only school id is used and shift conflicts of the same school within datasets are ignored; it is possible that data entry errors of the shift cause this problem, which has been also documented by Caballero (2010). To reduce the number of schools that would had to be dropped because they do not match through the school-shift id, after performing the 'hard match' I also matched some of the remaining schools using only the school number, taking the shift from C-600 as 'the right one'. Although this adjustment can raise concerns on its own, it is important to highlight that the main results do not change when the regressions are restricted to the schools that are 'hard matched'.

[16] The reasons why not all schools are part of the panel are basically two. First, it is natural that between 2002 and 2008 new schools were opened, because these schools did not exist in 2002 they are left out of the panel as they have no counterfactual at baseline. Second, it is possible that data entry errors and changing codes prevents a better linking of schools across time and data sources.

Table 1. Differences between matched and unmatched schools

	2002		2008		2009	
	Not in Panel	Panel	Not in Panel	Panel	Not in Panel	Panel
C600						
Schools	1,750	2,769	1,963	4,698	1,754	4,959
Total Students	69,544	170,694	72,046	294,518	66,246	308,051
Students by grade	93	135	88	138	91	138
Teacher/students[(1)]	0.08	0.06	0.06	0.05	0.06	0.05
Share of EPD teachers[(2)]	-	-	0.22	0.17	0.30	0.25
SABER						
Schools	3,705	2,769	993	4,698	1,023	4,959
Total students	74,529	165,327	49,719	284,599	49,587	299,212
Math						
Mean Test Score	42.2	42.3	43.4	44.1	42.7	43.3
Standard Deviation	6.0	6.0	8.1	8.2	9.6	9.6
Spanish						
Mean Test Score	47.5	47.8	44.9	45.4	45.5	45.9
Standard Deviation	8.4	7.8	7.0	6.9	7.2	7.2
Females (%)	0.53	0.55	0.54	0.55	0.54	0.55
Education of the mother (%)						
Elem. Or less	0.59	0.55	0.45	0.42	0.46	0.43
Secon. Or HS studies	0.33	0.36	0.44	0.45	0.43	0.45
Higher Ed. studies	0.09	0.09	0.11	0.13	0.10	0.12
Monthly household income in minimum wages (%)						
<1	0.36	0.33	0.36	0.33	0.38	0.35
>=1 and <2	0.42	0.43	0.47	0.47	0.45	0.46
>=2 and <3	0.14	0.15	0.12	0.14	0.12	0.13
>=3	0.08	0.08	0.05	0.06	0.05	0.05
Household size (%)						
1-4	0.34	0.34	0.42	0.42	0.41	0.42
5-8	0.58	0.59	0.52	0.52	0.53	0.52
9 or more	0.08	0.07	0.06	0.06	0.06	0.06

[(1)]Includes students in all six secondary grades

[(2)]At the institution level

Looking at the section on C-600 data we can see that 2,769 schools are in the panel in 2002, while 1,750 are not in the panel; in this year, schools in the panel tend to be bigger than schools out of the panel, and their student/teacher ratios lower. When we look at test scores for

2002, the differences between the means of the two groups are quite small (relative to the standard deviations); and the difference in the proportion of females is very similar too. Regarding household characteristics, education of the mother in schools in the panel is relatively higher, and household income is higher too. With respect to household size, the groups seem to have relatively similar distribution in 2002.

In 2008 and 2009 the number of schools in the panel is 4,698 and 4,959 respectively. Similarly to what is observed for 2002, schools in the panel are bigger than schools out of the panel, and the teacher/student ratios are lower.

The share of EPD teachers in panel schools is lower than in schools that are not in the panel, both in 2008 and 2009. It is worth saying that the data that refers to the number of teachers under each regime is at the institution level in the C-600 files; because the linking with the SABER11 data is at the school-shift level I imputed the percentage of EPD teachers in each school-shift using the percentage of EPD teachers in the corresponding institution.

With respect to SABER11 data, the displayed figures show very similar patterns to what was observed for 2002, the differences in test scores are small relative to the standard deviations, the mothers in panel schools are more educated and the households are wealthier; while there seem to be no large differences in household size.

These summary statistics show two important results. First, it was possible to match an important share of the data; in 2002 69 percent of test scores are in the panel, in 2008 this figure is 85 percent and in 2009 is 86 percent. Not surprisingly, 2002 is the year when more data is lost, it is possible that some schools were closed or were bundled into bigger, new schools, so I cannot find them in the follow-up years.

Second, the differences in the analyzed variables are small and, more important, relatively similar across time. In all three years schools in the panel are bigger, the mothers of the corresponding students are more educated and their household income are higher than schools that are not in the panel; this consistency of the differences suggests that whatever that makes a school to take part in the panel is relatively constant over time, in which case selection into the sample will be controlled by the school fixed effects and therefore it will not constitute a source of bias on the effect of EPD. Although is not possible to discard that selection into the sample

will bias the estimate of EPD on students' outcomes, the described summary statistics suggests that this selection process is probably driven by time invariant factors that are offset with the proposed fixed effects approach.

d) Methods and results

To evaluate the effect of EPD on test scores I use a school-fixed effects model. To fix ideas, I assume that the data generating process can be described by:

$$saber_{est} = \alpha s_{st} + x_{st}^{'}\beta + z_{est}^{'}\theta + \phi_t + \gamma_s + \varepsilon_{st} + u_{est} \quad (2)$$

Where $saber_{est}$ represents the standardized[17] test score of the student *e* in school *s* and year *t*; s_{st} is the percentage of teachers under EPD; x_{st} is a vector of time-varying control variables at the school level; z_{est} represent observed characteristics at the student level, ϕ_t is a year-fixed effect, γ_s bundles time-invariant observable and unobservable characteristics at the school level, ε_{st} is an idiosyncratic error at the school level, u_{est} represent unobserved effects at the student level and α, β and θ are parameters to be estimated. The coefficient of interest is α, which captures the effect of the percentage of teachers under EPD.

Table 2 presents results for math. The first column controls only for time-variant characteristics at the school level. We can see that the effect of EPD is negative and significant. In column 2 sociodemographic characteristics[18] are included; in this case the effect of EPD declines in absolute value, although it remains negative and significant. The introduction of school-fixed effects changes the estimates dramatically, the effect of EPD is still negative but very small and not statistically different from zero, this suggests that the negative effect found initially was picking up the fact that EPD teachers tend to be placed in low performing schools, which highlights the importance of using a school-fixed effects approach to evaluate this policy.

[17] Using the population mean and standard deviation of the corresponding year and calendar.

[18] The included variables are gender, age, two dummies for mother's education (the left out category is Elementary studies, the two dummies indicate mothers with Secondary studies and higher education studies); three dummies for household income (the left out category is less than one minimum wage, included dummies indicate between one and two minimum wages, between two and three minimum wages, and the third dummy for more than three minimum wages); household size and a dummy variable for whether the student works or not.

Table 2. Share of EPD teachers on SABER11 test scores - Math

	(1)	(2)	(3)	(4)
Share of EPD teachers	-0.107	-0.067	-0.004	-0.017
	0.025	0.021	0.023	0.023
Teacher/students	0.103	0.183	0.137	0.092
	0.164	0.129	0.111	0.108
Teachers with prof. degree / total	-0.012	-0.001	0.014	0.018
	0.018	0.015	0.011	0.011
Teachers with grad. degree / total	0.175	0.143	0.010	0.017
	0.021	0.018	0.013	0.012
Teachers pedag. studies / total	-0.006	-0.001	-0.017	-0.013
	0.018	0.015	0.012	0.011
Sociodemographic controls		X		X
School Fixed Effects			X	X
N	749,138	749,138	749,138	749,138

Standard errors clustered at the school level in parentheses

Note: All specifications include year and shift dummies, rural-specific trend terms, an indicator variable for schools that report 0 teachers and a dummy variable for type of schedule (A or B). Sociodemographic variables are gender, age, two dummies for mother's education (the left out category is elementary studies, the two dummies indicate mothers with secondary studies and higher education studies); three dummies for household income (the left out category is less than one minimum wage, included dummies indicate between one and two minimum wages, between two and three minimum wages, and the third dummy for more than three minimum wages); household size and a dummy variable for whether the student works. Missing data in the covariates are accounted for with missing data dummy variables.

Furthermore, it is interesting to analyze what happens to other school-level variables when we include the school-fixed effects. In particular, the coefficient on the percentage of teachers with graduate studies is positive and significant in the first two columns, but when the school-fixed effects are included is substantially smaller and not significant. This suggests that the positive effect initially observed was probably due in part to the fact that better students tend to attend schools where teachers have more education. It is important to say these results do not necessarily mean that there is no causal effect of the percentage of teachers with a graduate degree on test scores. It could be the case that the variable has a positive effect but does not have enough within school variation for its coefficient to remain significant after including the school-fixed effects. Clearly, further research is needed to analyze which of these possibilities explain these results.

In Table 3 we can see results for Spanish test scores. The effect of EPD shows the same pattern than for math, the coefficient is negative and significant for the first two specifications, but when school-fixed effects are included the coefficient remains negative but is not significant. The teacher/student ratio exhibits the most dramatic changes, the coefficient is not significant under the first two specifications but when school-fixed effects are incorporated the coefficient is positive and significant, which suggests that although some schools with high teacher/student ratios perform relatively poorly, when the within-school comparison is made there seems to be a positive effect of teacher/student ratios. The effect of the percentage of teachers with a graduate degree observes the same pattern than in the analysis for math, the coefficient is positive and significant for the first two models, but when school-fixed effects are included the effect washes out.

Table 3. Share of EPD teachers on SABER11 test scores - Spanish

	(1)	(2)	(3)	(4)
Share of EPD teachers	-0.095	-0.052	-0.023	-0.038
	0.024	0.019	0.025	0.024
Teacher/students	-0.144	0.120	0.400	0.363
	0.172	0.131	0.111	0.105
Teachers with prof. degree / total	-0.012	-0.008	-0.001	0.002
	0.020	0.017	0.013	0.013
Teachers with grad. degree / total	0.172	0.126	-0.012	-0.007
	0.023	0.020	0.016	0.016
Teachers pedag. studies / total	-0.009	-0.006	-0.019	-0.017
	0.019	0.015	0.013	0.013
Sociodemographic controls		X		X
School Fixed Effects			X	X
N	749148	749148	749148	749148

Standard errors clustered at the school level in parentheses

Note: All specifications include year and shift dummies, rural-specific trend terms, an indicator variable for schools that report 0 teachers and a dummy variable for type of schedule (A or B). Sociodemographic variables are gender, age, two dummies for mother's education (the left out category is elementary studies, the two dummies indicate mothers with secondary studies and higher education studies); three dummies for household income (the left out category is less than one minimum wage, included dummies indicate between one and two minimum wages, between two and three minimum wages, and the third dummy for more than three minimum wages); household size and a dummy variable for whether the student works. Missing data in the covariates are accounted for with missing data dummy variables.

The presented results indicate that EPD did not have much effect on SABER11, which is at odds with the results that I obtained for 5^{th} and 9^{th} grade test scores (Ome, forthcoming).

One explanation for this is that in SABER11 I have controls that I do not have in SABER5&9, so my estimates in SABERE5&9 may be actually biased and the 'true' causal parameter is zero for all grades. In effect, one major difference between the analysis performed for SABER5&9 and SABER11 is that in the latter I have access to sociodemographic characteristics at the student level, while in the former I do not. If students with better sociodemographic characteristics tend to go to schools where more EPD teachers are being hired, the share of EPD teachers in SABER5&9 will be picking up not the causal effect of EPD but simply the better makeup of the students going to schools with EPD teachers; this could explain why I find positive (and biased) effects for EPD teachers in SABER5&9 (where I cannot control for sociodemographic characteristics) but no effects for SABER11 (where I can). I do not believe this is the case because, as shown before, the use of sociodemographic characteristics in SABER11 does not really affect the estimates of EPD or any other school-level variable once school-fixed effects are controlled for. Along these lines it seems that once school-fixed effects are controlled for, omitting sociodemographic characteristics has only minor effects.

An alternative explanation is that the cohort graduating in 2009 was exposed to EPD teachers too late in their development process. To test whether this is the case further research should focus on the effects on SABER11 of cohorts graduating later than 2009; this will surely improve our understating of the effect of EPD on students' outcomes.

e) Heterogeneous effects of EPD and the Skills Evaluation

One of the most important components of the EPD is its promotion policy. In the context of the previous regime, promotions are granted in accordance to a wage schedule that maps experience and education into wages. In the EPD, education and experience also matter for promotions, but in addition to that, teachers also need to pass the Skills Evaluation (SE).

The SE is a written exam implemented by the Ministry of Education. The SE is divided in three sections, the behavioral section, the pedagogy section and the academic section. Each of these sections weights roughly one third of the total score[19]. A minimum score of 80 out of 100 points is required to be considered for promotion, which is ultimately granted if budgetary restrictions at the corresponding territorial level allow it.

[19] The weights vary a little depending on the type of promotion a teacher is applying to, and whether he is a regular teacher or a principal.

The purpose of the behavioral section is to measure how teachers react to different circumstances in the school and the classroom. In general, questions are design to portrait an everyday situation in the school, such as a lecture or a parents meeting.

The pedagogy section attempts to grade teachers in their ability to communicate to the students the different contents of the curricula, as well as its inclination for establishing associations between different cognitive areas.

The academic section evaluates teachers' knowledge and mastery of their specific area of expertise. The academic section for elementary school teachers encompass four 'basic skills', namely: science, math, communication and citizenship. In the case of secondary and high school teachers, the academic section is specific to the subject each teacher is in charge of (e.g. Spanish, math)[20].

The nature of the SE, specifically the fact that is a standardized exam, has some positive and some negative features. On the upside is that is an impartial and objective evaluation. Because is designed and graded centrally it is unlikely that the results are subject to favoritisms or discrimination. In this sense it can be argued that the SE is fairly transparent and neutral.

On the downside, the SE tries to measure constructs that are probably not easy to assess through a written exam, such as responsibility or propensity to plan activities ahead. Another potential problem with the SE is that a standardized exam may not be the best instrument to measure the value that teachers add to their students, considering the very different challenges that teachers have to face in varying geographic and socioeconomic contexts.

Along these lines, the key question to answer is whether the SE helps distinguishing the teachers whose students progress the most. To analyze this issue I extend the fixed effects model presented before to analyze whether the effect of EPD is different between schools with high scoring teachers and schools with low scoring teachers.

To perform this analysis I use the results from the first SE, fielded in January 2010. This evaluation was presented by 33 thousand EPD teachers and principals, and was passed by a little

[20] For sample questions on each of these sections see the Ministry's guidelines for the Skills Evaluations, available at: http://www.mineducacion.gov.co/proyectos/1737/article-210992.html

more than a fourth of them. I drop roughly 4 thousand exam results that correspond to principals or preschool teachers. Table 4 presents descriptive statistics of participants in SE, discriminating by those that score lower and higher than 80.

Table 4. Skills evaluations - Summary statistics

	Score		All
	<80	>=80	
Education level (%)			
Technical	6.9	6	6.6
Professional	60.7	56.2	59.5
Professional Ed.	16.1	16.3	16.1
Grad. (1 year)	15.1	15.7	15.2
Grad. (2+ year)	1.2	5.9	2.5
Age (%)			
19<=age<35	39.3	50.3	42.3
35<=age<45	41.1	34.9	39.4
45<=age<55	14	10	13
55<=age	5.6	4.8	5.4
Female (%)	58.6	58.4	58.6
Level of teaching (%)			
Elementary	42.4	42.7	42.5
Secondary	42.5	40.3	42
High School	15	16.9	15.5
N	21,765	8,011	29,776

Source: Ministery of Education

We can see that only 8,011 teachers (27 percent of total examinees) passed the exam, in this regard it is apparent that SE is not just a mere formality, but a quite difficult requisite to be promoted. Teachers with graduate studies seem to do better in the SE, especially those with a Master or PhD. Interestingly, younger teachers seem to perform substantially better than their older counterparts. Finally, neither gender nor the level of teaching seem to be correlated with performance in the SE.

To analyze to what extent the SE help distinguishing the best teachers, I calculate the mean SE score by school and merge these means to the panel on SABER test scores. The analysis is limited to the schools where there is at least one teacher that took the SE. Furthermore, because not all EPD teachers gave the exam in 2010, the mean score that is calculated for each school is based only on the scores of the teachers that took the exam.

For elementary schools 46 percent of schools in the panel have at least one SE score in 2009. In the case of secondary and high schools these percentages are substantially lower as only schools with scores in subject-specific SE are included in the analysis. For example, for the regression on 9th grade math only the secondary schools with a test score in math are analyzed, which implies that only 31 percent of schools in the panel are included in the analysis for this exam. The regression for 9th grade Spanish includes 19 percent of the original panel sample.

For 11th grade it is important to highlight two additional considerations. First, the analysis presented in the previous section included the years 2002, 2008 and 2009, for this section I dropped 2008 data as I do not observe any measure of quality of that year's EPD teachers. Second, high schools in Colombia have only two grades (10th and 11th) so most high schools are attached to a secondary school, for this reason to calculate the mean SE score for each high school I use secondary and high school teachers.

Using the calculated means as a measure for the quality of EPD I run a slightly modified version of equation (1) on the explained subsamples:

$$saber_{est} = \alpha s_{st} + \varphi \mathbf{1}(mean\_SE_s > 80) + x_{st}^{'}\beta + z_{est}^{'}\theta + \phi_t + \gamma_s + \varepsilon_{st} + u_{est} \quad (3)$$

Where **1(.)** is an indicator variable that takes the value of 1 if the expression in parentheses is true. The parameter of interest is φ that captures the effect that the SE average score in school *s* is greater than 80 points. An important assumption underlying this approach is that the school to which each teacher is affiliated when he presented the SE is the same school to which he was affiliated through 2009, when the students outcomes are observed; although there might be some teachers that changed schools at the end of 2009, it is unlikely that switching schools in such short period is affecting the presented results in a major way.

Table 5 shows results for equation (2). The baseline results corresponding to 5th and 9th grades are discussed extensively somewhere else (Ome, forthcoming). As mentioned before, one of the major differences between SABER11 and SABER5&9 is that the former has sociodemographic information, while the latter does not; also, SABER5&9 is designed to rank schools rather than students, so the unit of observation is the school, while the unit of observation in SABER11 is the student (this is why the sample sizes for SABER11 are so much larger than those of SABER5&9).

Table 5. Heterogeneity on the effect of EPD on students' test scores according to teachers' performance in SE.

	Math			Spanish		
	All	Schools with SE score		All	Schools with SE score	
A. 5th grade						
% of EPD	0.094	0.099	0.100	0.053	-0.042	-0.041
	(0.039)	(0.058)	(0.058)	(0.037)	(0.053)	(0.053)
High SE			0.086			0.087
			(0.038)			(0.036)
N	49,980	22,811	22,811	49,737	22,648	22,648
B. 9th grade						
% of EPD	0.299	0.059	0.061	0.237	0.043	0.040
	(0.080)	(0.106)	(0.106)	(0.063)	(0.115)	(0.115)
High SE			0.102			0.023
			(0.067)			(0.060)
N	11,219	3,494	3,494	11,220	2,097	2,097
C. 11th grade						
% of EPD	-0.033	-0.026	-0.027	-0.047	-0.080	-0.091
	(0.036)	(0.087)	(0.087)	(0.038)	(0.121)	(0.116)
High SE			-0.018			-0.026
			(0.038)			(0.061)
N	464,539	99,093	99,093	464,549	39,743	39,743

Standard errors clustered at the school level in parentheses

Notes:
5th and 9th grade: Regressions are weighted using the average number of students in years 2002, 2005 and 2009. The regressions include school-fixed effects, year dummies, rural-specific trend terms, four dummies for shift-fixed effects, one dummy for schools with a high probability of fraud in SABER, percentage of students with free tuition, teacher/pupil ratios, percentage of teachers with a bachelor degree or higher, percentage of teachers with formal training in pedagogy and one dummy variable for schools that report 0 teachers. For more details see Ome (forthcoming).
11th grade: See Table 3.

Section A shows results for 5th grade test scores. The first column shows the coefficient for the baseline specification, we can see that the effect of the share of EPD teachers on math is positive and significant. In the second column the same regression is run but over the subsample of schools with SE score, in this case the coefficient on the share of EPD teachers remains stable despite the sample reduction. In the third column I include the indicator variable for schools where the average SE score is greater that 80; we can see that the effect is positive and significant, the coefficient implies that in schools where EPD teachers scored more than 80 on average, the 5th grade students will have math scores higher by 9 percent of a standard deviation. A similar result arises when we look at 5th grade Spanish; as the coefficient on the indicator variable is positive and significant.

Section B presents results for 9th grade test scores. The baseline effect of EPD is substantially higher for 9th graders than for 5th graders; however, when the regressions are restricted to the subsample of schools with SE scores, the effect of EPD teachers is substantially smaller and no longer significant for either math or Spanish. The reason for this might be that, in the context of the restricted subsample, the share of EPD teachers has much less variation than when all the schools are included in the regression. With respect to the effect of having high scoring teachers, the coefficient is positive but not significant for math and Spanish.

In the case of the 11th grade results (displayed in Section C), the coefficients on having high scoring teachers is negative but not significant.

The fact that not all schools have EPD teachers (and that not all EPD teachers took the exam in 2010) causes that the sample sizes reduce dramatically when I try to analyze the effect of SE, especially for 9th and 11th grades. This could be causing the nil results founded for the effect of having high scoring teachers for these grades, especially if we consider the positive and significant effects for 5th graders, where the sample sizes are not as reduced with respect to the baseline regressions.

f) Final comments

Overall it seems that the EPD did not do much for the cohort that graduated in 2009 from high school. A very different story seems to be occurring with 5th and 9th graders. In previous work I showed that EPD teachers are correlated with higher test scores, especially for 9th graders.

The heterogeneity of the effect of EPD across levels of education can have different explanations. The two that were already mentioned before are: i) changes in the underlying composition of ability correlated with the presence of EPD teachers can be down biasing the estimates of EPD on test scores; and ii) that the cohort graduating in 2009 might have had little or too late exposure to the EPD teachers.

Another, less optimistic explanation (for EPD teachers) is that positive effects of EPD might fade away in high school. In the same flavor it could be the case that more experimented teachers are better at teaching high school students than the novice EPD teachers.

In the upcoming years, when the cohorts where an effect of EPD has been picked up graduate, we will be able to test some of these hypotheses regarding the effects of EPD teachers on 11th graders.

Regarding the SE, I found that schools with high scoring EPD teachers outperform schools with low scoring EPD teachers for 5th grade math and Spanish, while no differential effects were found for either subject in 9th or 11th grades.

References

Barrera-Osorio, F. (2006). The impact of private provision of public education: Empirical evidence from Bogotá's concession schools. *Impact evaluation series*, (10), 1-30.

Bautista, M. (2009). La profesionalización docente en Colombia.

Bonilla, L., & Galvis, L. A. (2011). Profesionalización docente y la calidad de la educación en Colombia.

Borjas, G. J., & Acosta, O. L. (2000). Colombia EDUCATION REFORM IN COLOMBIA * George J . Borjas , Harvard University Olga Lucía Acosta , Fedesarrollo. *Education*, (19), 1-27.

Caballero, G. (2010). Base de datos consolidada, 2005-2008.

Dee, T. S., & Keys, B. J. (2004). Does merit pay reward good teachers? Evidence from a randomized experiment. *Journal of Policy Analysis and Management*, *23*(3), 471-488. doi:10.1002/pam.20022

Duarte, J. (1996). La debilidad del Ministerio de Educación y la politización de la educación: dos problemas a enfrentar en el plan decenal. *Coyuntura Social 14: 145-167.*

Gaviria, A., & Barrientos, J. H. (2001). Determinantes de la calidad de la educación en Colombia. *Archivos de Economía*, (159).

Gaviria, A., & Umaña, C. M. (2004). Estructura salarial de los docentes públicos en Colombia.

Glewwe, P., & Kremer, M. (2003). Teacher incentives. *National Bureau of Economic Research*, (9671).

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review*, *10*, 179-214.

Ministerio de Educación Nacional. Decreto 1278 de 2002. *Por el cual se expide el Estatuto de Profesionalización Docente*. Meritocracia para Docentes 23

Ministerio de Educación Nacional. Decreto 2277 of 1979. *Por el cual se adoptan normas sobre el ejercicio de la profesión docente*.

Ministerio de Educación Nacional. Decretos 230 de 2002, 2100 de 2004, 3982 de 2006, 3986 de 2006, 3940 de 2007, 789 de 2009 y 1290 de 2009.

Mizala, A., & Ñopo, H. (2011). Teachers ' Salaries in Latin America : How Much Are They (Under or Over) Paid ? *Development*, (5947).

Muralidharan, K., Sundararaman, V., Cullen, J., Dahl, G., Das, J., Feldstein, M., Freeman, R., et al. (2009). Teacher Performance Pay : Experimental Evidence from India. *Human Development*, (August).

Núñez, J., Steiner, R., Cadena, X., & Pardo, R. (2002). ¿Cuáles colegios ofrecen mejor educación en Colombia? *Documento CEDE*, 1-56.

Ñopo, H. (2008). Matching as a tool to decompose wage gaps ˜. *The Review of Economics and Statistics*, *90*(May), 290-299.

OECD. (2011). *Education at a Glance 2011 OECD inDiCatOrs*. *Education*.

Ome, Alejandro. Meritocracy for Teachers: Evidence from Colombia.

Springer, M. G., Stecher, B. M., Ballou, D., Hamilton, L., Le, V.-N., Lockwood, J. R., McCaffrey, D. F., et al. (2010). Teacher Pay for Performance. *National Center on Performance Ince2ntives*.

Umaña, C. M. (2004). Esquemas de incentivos para la carrera docente. *Archivos de Economía*.

Woessmann, L. (2010). Cross-country evidence on teacher performance pay. *Prepared for the PEPG Conference*.