Meritocracy for teachers: Evidence from Colombia

Alejandro Ome*

September 9, 2012

Abstract

In 2002 the Colombian government issued a new code for public school teachers, the *Estatuto de Profesionalización Docente* (EPD). Among its most important features is that in order to get hired and promoted, teachers must pass a series of examinations, a requirement absent in the previous code where teachers got promoted basically as they aged and through further training. I implement a school-fixed effects model to analyze the effect of EPD on dropout rates and students' test scores. I find a negative correlation between dropout rates and EPD teachers for Elementary and Secondary schools, and no effects for dropout rates at High school. For 5th grade test scores the effects are positive for both math and Spanish, but only significant for math. For 9th grade there are positive and significant effects for both math and Spanish. No effects are found for test scores at 11th grade.

1. Introduction

Introducing meritocracy and accountability in the teachers' career constitute a promising area for education policy. In this spirit, the Colombian government issued in 2002 a new code for public school teachers, the *Estatuto de Profesionalización Docente* (EPD). Among its most important features is that in order to get hired and promoted, teachers must pass a series of examinations, a requirement absent in the previous code where teachers got promoted basically as they aged and through further training. By providing teachers with this new accountability system, it is expected that students' outcomes will improve. This paper evaluates the effect of EPD on students' dropout rates and test scores.

*Email: aome@uchicago.com. I thank Jeffrey Grogger, Alicia Menedez, Kerwin Charles, Andrea Flores and seminar participants at Fedesarrollo, Universidad de los Andes and the University of Chicago for valuable suggestions and discussions. Financial support from Fedesarrollo and ICFES is gratefully acknowledged.

Broadly speaking, programs that try to incorporate accountability in the teaching profession come in two flavors; output-based accountability systems, where teachers' compensation (or tenure decisions) are linked to students' outcomes; and input-based systems, where school administrators try to select more qualified teachers, and promotions or tenure decisions are granted on the basis of some performance measure not directly linked to students' outcomes, such as peers or principals evaluations.

Output-based systems are very attractive because incentives seem well-aligned with the outcomes of interest (for a discussion see Kane and Staiger, 2002). The risk this type of systems entail is that teachers may focus too much on the outcomes over which they are being evaluated, which could lead to problems like 'teach to the test' or even cheating.

Studies that analyze output-based accountability systems provide at best mixed evidence on the effects of merit pay. Muralidharan and Sundararaman [2011] study the effect of monetary incentives for teachers linked to students' outcomes in India, where a number of schools were randomly assigned to treatment (being under a performance-pay scheme) and control groups; they find that treated schools observed significant increases in test scores. Glewwe and Kremer [2003] analyze a similar experiment in Kenya, they also found higher test scores in treated schools, but after conducting a follow up a couple of years later the difference in outcomes vanished, suggesting that treated teachers limited to 'teach to the test' rather than increasing the cognitive skills of the students. In the US, Springer et al. [2010] evaluate a large randomized experiment conducted in Nashville finding no significant results of teacher incentives on students' performance.

Input-based accountability systems, on the other hand, do not rely on students' outcomes to evaluate teachers, but in other types of strategies such as stricter protocols to hire teachers, and third party evaluations to grant promotions or tenure. As accountability is not directly linked to students' performance it is less of a threat that teachers will 'teach to the test', the question in this context is whether this type of accountability systems will affect students' outcomes at all.

Duflo et al. [2012] analyze a program in Kenya where parent-teacher associations were in charge of hiring teachers under temporary contracts; the authors find that these contract teachers observed a better performance than traditional teachers, measured by their corresponding students' test scores. Dee and Keys [2004] study the implementation of a merit pay system in Tennessee in the mid 80's in which the state government launched a career ladder for teachers based on comprehensive evaluations

made mainly by tenured teachers. To evaluate the effect of the introduction of the ladder the authors exploit a parallel school program, the widely known STAR project, that randomly assigned pupils to classrooms of different sizes to evaluate the effect of class size on achievement. Dee & Keys found that students that were (randomly) assigned to a laddered teacher had higher test scores for math but no significant effect was found for reading.

Aside the apparent difference between output-based and input-based accountability systems regarding the alignment between incentives and outcomes of interest, the two types of schemes differ also in how they are usually implemented, which affects the parameter that can be identified in each intervention.

Empirical designs that analyze output-based accountability systems in general tackle the selection problem by randomizing which teachers or schools get to be under a performance-pay scheme, this prevents that students select into teachers under performance pay (otherwise the parameter of interest would be biased), but it also impedes teachers' selection, which in general we would like to observe, as well as include its impact on students' outcomes as part of the treatment effect. Input-based accountability studies, on the other hand, usually analyze how novice teachers perform under new rules, compared to incumbent teachers that are governed by traditional rules. Because the regime under which teachers are is not randomly assigned, the effect of teachers' self-selection can be observed and measured.

In the context of the EPD, incumbent teachers in 2002 can remain under the 'old' code or switch to EPD so self-selection is partially observed (although no novice teacher can choose to be in the old regime). This feature of the policy causes that the percentage of teachers under the new code varies across schools. A naïve estimator will regress student's outcomes on the share of EPD teachers; the problem with this approach is that students might be sorted across schools in a way that could be correlated with the presence of EPD teachers, which could bias the estimate of EPD. To control for possible sorting of students I implement a school-fixed effects model. Providing that the characteristics of the children that are correlated with both the presence of EPD teachers and students' outcomes are time invariant, this model identifies the effect of EPD on students' outcomes.

Along these lines, I analyze two types of outcomes: permanence in the school system, measured by dropout rates; and achievement, measured by standardized test scores. To analyze the effect of EPD on dropout rates I constructed a panel of public schools in Colombia for the years 2004

(before the first call for EPD teachers was made) and 2008 (after three calls had been made) using data collected by the National Institute of Statistics (DANE), which provides not only information on dropout rates but also the number of teachers on each regime. To study the effect of EPD on standardized test scores I constructed a similar panel using data from the Colombian Institute for the Evaluation of Education (ICFES).

I find a negative correlation between dropout rates and EPD teachers for both Elementary and Secondary schools, and no effects for dropping out at high school. For 5th grade test scores I find significant effects only for math. For 9th grade there are positive and significant effects for both math and Spanish. No effects are found for test scores at 11th grade.

The reminder of this paper is as follows. The next section describes relevant aspects of the education sector in Colombia, focusing on teacher labor markets. The third section sketches the empirical framework. The fourth section describes the data and the main results. Last section concludes.

2. The Colombian Education System

Education in Colombia is mandatory and free[1] from age 5 to 15, covering 1 year of Preschool, 5 years of Elementary school and 4 years of Secondary school. Additionally, 2 years of High school (*Media Vocacional*) are required in order to attend a higher education institution[2].

Although education coverage has been increasing over the last decades, full coverage has not been reached yet, not even for children 7 to 11 years old. In 2005 (the year of the last census) the share of children in this age range attending an education institution was 90 percent. This coverage rate is low even for the region; according to UNESCO (2005) countries with similar or lower GDP per capita like Bolivia or Peru had, already in 2002, coverage rates higher than 90 percent. When considering children 12 to 17 years old, Colombia's coverage rate goes down to 76 percent[3].

Roughly 20 percent of students attend private schools, while the rest goes to public institutions. The public schools system is decentralized in the sense that municipalities administer resources

[1] Although attending a public school is in principle free, under the period I study the system allowed charging fees according to household income (Political Constitution, Art. 67). In effect, the majority of families paid for tuition and other expenses, even at the lowest deciles of income (see Barrera and Domínguez [2005]).

[2] Some schools specialized in pedagogy offer two additional grades after which the degree of associate teacher (Normalista) is granted.

[3] Source: Own calculation using Census 2005 tabulations. http://www.dane.gov.co/censo/files/cuadros%20censo%202005.xls

through their secretaries of education[4]; however, these resources come largely from the national government. The secretaries administer education institutions, an administrative unit that contains one or more schools (campuses). In addition, due to the lack of funds to provide all students with full-day schooling, many schools (especially in urban areas) serve multiple shifts during the day, in general one group of children in the morning and another in the afternoon.

Regarding quality of education and its evaluation, the Colombian Institute for the Evaluation of Education (ICFES) implements two types of exams, the SABER 5 & 9 and the SABER 11. The SABER 5 & 9 are presented once every three years by students in 5th and 9th grades (the last grade of Elementary and Secondary school respectively), these tests evaluate students in math, Spanish and science. The SABER 11 are presented yearly by all students finishing high school and it evaluates knowledge in math, Spanish, natural and social sciences, philosophy and English.

In international comparisons Colombia performs rather poorly. In standardized tests carried out by the OECD in 2009 (Program for International Student Assessment-PISA) Colombia ranked 52 in reading out of 65 participant countries; within Latin-American countries it ranks above Brazil, Argentina, Peru and Panama, but below Chile, Mexico and Uruguay[5].

Teachers Labor Market

Since 1979 school teachers were governed by the Decree 2277, which established a career ladder based exclusively on experience and schooling. A novice teacher would be placed in the ladder according to her education level, and raises would be granted as he gained years of experience or through further training. In this regard, the teaching profession in Colombia mimics most US public school systems, where schools districts rely on salary schedules based on education and experience (Podgursky and Springer, 2007).

In terms of labor stability, according to the Decree 2277 once a teacher was placed in the ladder he was basically tenured, as causes for dismissal were practically unrelated to students' performance[6].

A few authors have found suggestive evidence on the negative effects of the institutional arrangement embedded in the 1979's code on students' achievement. Gaviria and Barrientos [2001] analyze

[4]Only 'certified' territorial entities are in charge of paying teachers. To get certified, a municipality has to show that it has the institutional capacity to run the schools (see Decrees 2700 of 2004 and 3940 of 2007 of the Ministry of Education), and if a municipality is not certified, the corresponding department manage teachers' payroll.

[5]From Latin-America only these seven countries (and Colombia) participated in PISA 2009.

[6]In effect, most of the valid causes for firing a teacher corresponded to disciplinary misconducts (see Decree 2277, Art. 46).

the impact of schools' characteristics (e.g. infrastructure, teacher/pupil ratio) on 11th grade 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. Similarly, Núñez et al. [2002] compare test scores from traditional public schools with public schools managed by the church to find that the latter outperformed the former which, again, highlights the importance of the institutional arrangement that characterize each school. Barrera-Osorio [2006] compares chartered and traditional public schools in Bogota in the early 2000's, he finds that chartered schools observe lower dropout rates and higher test scores than comparable public schools.

Tackling the problems in the rules that governed teachers' careers was the main objective of the EPD[7]. The meritocratic nature of EPD commences with the hiring process. An individual that wants to become a teacher has to apply to calls made by municipalities[8], give a written evaluation and an interview and stay in probation for a year if she is selected for the job. Although there was a pretty similar merit based process for hiring teachers under the previous regime, the actual appointment of teachers was highly politicized (see Duarte, 1996), teachers were often hired thanks to a recommendation from a local political baron rather than their performance in the calls, which were quite infrequent.

Figure 1 shows the number of individuals that participate in the first three calls[9] for EPD teachers, the number of available positions and the number of candidates placed in probation. It is clear that the teaching profession is attractive to a large number of individuals relative to the demands of the system; in 2004 over 140 thousand candidates applied for 50 thousand available positions, and a similar 'oversupply' of candidates is observed in the subsequent years. On the other hand, this data also suggests that applicants' quality is not fulfilling the requirements to become a teacher, as the number of hired individuals is quite lower than the number of available positions, although this gap closed almost completely in 2006.

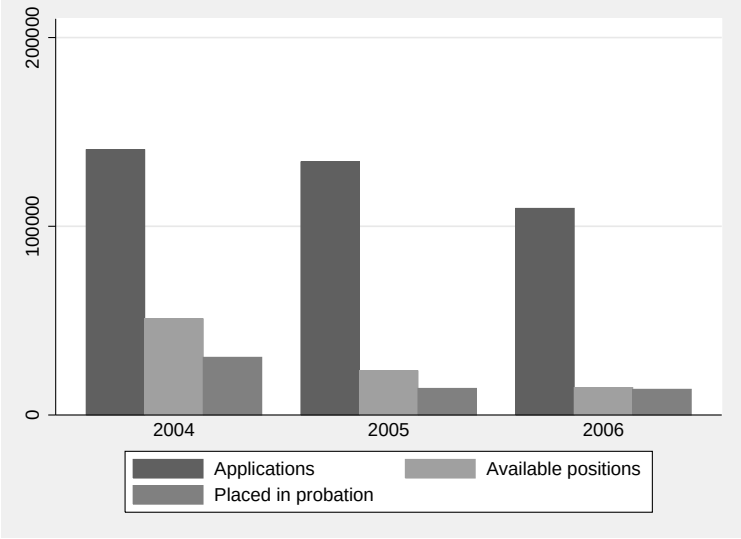
After one year in probation, teachers are subject to an evaluation that determines whether they

[7] Umaña [2004] presents an exhaustive comparison between the EPD and the Decree 2277. See also Bautista [2009].

[8] Only 'certified' territorial entities can call for teachers.

[9] The fourth call was made in 2009. To May, 2010 there were no teachers placed in probation from this call, so we do not observe any of these teachers in the present study (see: http://www.mineducacion.gov.co/1621/article-233974.html)

Figure 1: Calls for EPD teachers



Source: Ministry of Education

are placed in the ladder or let go (the probation evaluation). Depending on their education level teachers enter to different branches of the ladder, where each branch has four levels A, B, C and D. Someone with a technical degree will be placed in 1-A (short for branch 1, level A); a bachelor will be placed in 2-A[10]; and someone with a postgraduate degree in 3-A.

Once placed in the ladder, all teachers are subject to a yearly performance evaluation. This is a quite multidimensional evaluation as it can include students' work samples to parental complaints and test score evaluations (e.g. the SABER examinations). Although the Ministry of Education provides a comprehensive guideline for these evaluations that includes a system of points and a list of dimensions across which teachers should be evaluated[11] (e.g. pedagogy, knowledge of the curriculum, communication skills), its design and implementation largely depends on the principals. When a teacher does not perform well for two consecutive years his contract is terminated.

The fact that the performance evaluation depends so much on the principal raises concerns about its objectivity, these concerns can be divided in two separate issues; first, do principals have the capacity to make a reasonable assessment about the quality of the teachers? Jacob and Lefgren [2005] find that principals perform a fairly good job assessing the ability of their teachers to increase student achievement; in effect, these authors show that principal evaluations on teachers

[10] A bachelor with a degree that is not in education is required to take a program in pedagogy to enter the ladder.
[11] Ministry of Education. Guía Metodológica No. 31 'Evaluación Anual de Desempeño Laboral' (MEN).

are better predictors of student achievement than education and experience of the teacher (the usual determinants of teachers pay).

The other problem is whether principals would perform fair evaluations or exercise some favoritism; indeed, a subjective evaluation can be easily turn into an arbitrary one, and although the mentioned guideline indicates not only that the aspects to be evaluated should be clear for the teacher at the beginning of the year but that the specific goals should be discussed with him, it is impossible to discard the possibility that the evaluation might be too subjective and instead of being an adequate incentive for teachers it constitutes a source of stress for them.

The performance evaluations represent another important departure from the 1979's code. In effect, teachers under the previous regime where evaluated only every six years, and the consequences of these evaluations where far from clear. In this regard, the implementation of EPD constitutes the end of tenure in the teaching career.

The last type of evaluation introduced by EPD is the competences evaluation. After three years of being in the ladder a teacher can apply to be promoted by presenting a written evaluation aimed at establishing his skillfulness to explain problems and implement teaching strategies[12]. The teachers that approve this exam are promoted to the next level of his respective branch, which comes with a pay raise. A similar evaluation is required for teachers that acquired more education and want to be promoted to a higher branch of the ladder. The competences evaluation introduced by EPD constitutes yet another important difference with the 1979's code, considering that under this code, promotions depended solely on years of service and additional training.

Although EPD incentives are not directly tied to students' test scores (as most merit pay systems are), all the evaluations to which teachers are subject to are in one way or another aimed at rewarding teachers that perform a better job. Either by changing the behavior of teachers from what they would do under the old code, or by attracting different type of individuals to the teaching profession, the question to be answered in this study is whether EPD teachers have a different effect on student outcomes than traditional teachers.

3. Empirical Methods

To estimate the effect of EPD on students' outcomes I use the widely known fixed effects ap-

[12]Ministry of Education. Documentos Guía of the Competences Evaluation.

proach. The fixed effects model compares the change in the outcome of the population that is affected by a given intervention, which in this case is the presence of EPD teachers, with the change in the outcome of the population that is not affected by the intervention.

More specifically, I propose a model in which dropout rates are a function of the share of teachers under EPD and an additive error term that incorporates a school fixed effect and a time fixed effect. Although in this context the treatment is a continuous variable (the share of EPD teachers in a given school), the main idea behind the fixed effects framework is that it estimates how the share of EPD teachers affects the within schools change in the dependent variable. To fix ideas, a model for the dropout rate for school s at year t is described by:

$$d_{st} = \alpha s_{st} + \mathbf{x'_{st}}\beta + \phi_t + \gamma_s + \varepsilon_{st} \tag{1}$$

The parameter of interest is α, which captures the effect of the share of EPD teachers, denoted by s_{st}. $\mathbf{x_{st}}$ is a vector of time-varying control variables at the school level; ϕ_t is a year-fixed effect, γ_s bundles time-invariant observable and unobservable characteristics of the school s and the last term is an idiosyncratic error.

The underlying assumption of this model is that unobservable characteristics that affect dropout rates, both from the schools and the students, are time-invariant and therefore embedded in the school fixed effect.

Regarding test scores the estimation technique relies also on the assumption that the unobserved characteristics that affect the outcomes are time-invariant and therefore captured by the school fixed effect. Although in principle the regressions in this context could be at the student level (as each student has a test score) due to data limitations explained later, for 5th and 9th grades the empirical analysis is at the school level, following a process similar to (1), only that the dependent variable is not the dropout rate but the mean test score of each school. For 11th grade test scores, on the other hand, the specification is at the student level, and can be described by:

$$score_{est} = \alpha s_{st} + \mathbf{x'_{st}}\beta + \mathbf{z'_{est}}\theta + \phi_t + \gamma_s + \varepsilon_{st} + u_{est} \tag{2}$$

Where $score_{est}$ represents the test score of student e in school s in year t, $\mathbf{z_{st}}$is a vector of student characteristics, namely gender, age, mother's education, household income and family size,

the last term is an error at the student level, and the rest of the terms are the same as in (1).

4. Data and Main Results

Dropout Rates

The formulae C-600 is the main data source for all information related to dropout rates and number and type of teachers. These forms are to be filled every year by all schools. As mentioned before some schools have more than one shift, serving one group of children in the morning and another in the afternoon; considering that shifts within the same school can be fundamentally different (Bonilla [2011] shows that students attending the afternoon shift tend to be more disadvantaged, and have lower results in standardized test scores) and exploiting the fact that most of the C-600 information is by school-shift, the analysis is carried out at the school-shift level. I use data on 2004 as baseline given that the first call was at the end of that year, and 2008 as the follow-up year[13].

Due to the structure of the C-600 formulae all the data needed is at the school-shift level except the information related to the type of teacher. This data is available only at the institution level, which can bundle more than one school (while one school can bundle more than one shift). I impute the share of EPD teachers at the school-shift level using the share of EPD teachers at the corresponding institution.

Now, for a school to be in the panel it has to show up in both 2004 and 2008; because some schools observed in 2004 are not found in 2008 and vice versa[14], it is a source of concern that schools in the panel share characteristics that could be correlated with the presence of EPD teachers, which could bias the estimate of the effect of EPD teachers as a regression coefficient will be confounded with the effect of the characteristics that made schools select into the panel sample. To analyze this problem I compare how different are dropout rates and other variables between schools in and out of the panel sample.

Table 1 shows the number and main characteristics of schools in and out of the panel. Section A focus on Elementary schools (grades 1st to 5th). At baseline observable characteristics seem to be rather similar between schools in and out of the panel. School size and dropout rates are not

[13]It was not possible to get C-600 microdata on teachers for 2004 so I use teachers data from 2005 as proxy; as an alternative I interpolated 2004 using 2002 and 2005 data, the results (not shown) were basically the same.

[14]The reasons why a given school might show up only in one year are basically two. First, it is natural that between 2004 and 2008 some schools are opened and some are closed. Second, it is possible that data entry errors and changing codes prevents a better linking of schools across time.

Table 1: Summary statistics

	2004			2008		
	Not in Panel	Panel Sample	Difference	Not in Panel	Panel Sample	Difference
	(1)	(2)	(3)=(1)-(2)	(4)	(5)	(6)=(4)-(5)
A. Elementary						
School size	101	99	1.430	89	92	-2.252
Dropout rate	0.068	0.070	-0.003	0.055	0.056	-0.001
Teacher/pupil ratio	0.032	0.032	-0.001*	0.032	0.034	-0.003***
EPD teachers/total[(a)]	0.000	0.000	-	0.212	0.171	0.041***
School-shifts	3,139	39,212		4,156	38,915	
Schools	2,867	35,764		3,911	35,764	
B. Secondary						
School size	212	269	-56.907***	142	291	-148.564***
Dropout rate	0.067	0.062	0.006	0.062	0.059	0.003
Teacher/pupil	0.037	0.038	-0.001	0.035	0.037	-0.001**
EPD teachers/total[(a)]	0.000	0.000	-	0.204	0.154	0.050***
School-shifts	871	7,478		2,331	7,543	
Schools	763	6,383		2,158	6,383	
C. High School						
School size	106	129	-22.429***	76	146	-69.637***
Dropout rate	0.038	0.038	-0.001	0.043	0.041	0.003
Teacher/pupil	0.046	0.046	0.000	0.044	0.042	0.002
EPD teachers/total[(a)]	0.000	0.000	-	0.199	0.131	0.067***
School-shifts	545	4,641		1,795	4,642	
Schools	462	3,939		1,652	3,939	

Standard errors bootstraped (100)
Summary statistics are weighted by school size
[(a)]At the institution level
* p<0.05 ** p<0.01 *** p<0.001

statistically different between schools in and out of the panel. Teacher/pupil ratios are higher in schools in the panel, although the difference is really small relative to the mean. Of the 42,351 public elementary schools-shifts in 2004, 93 percent are in the panel.

In 2008 we can see that, in general, schools are smaller and dropout rates are lower with respect to 2004, while teacher/pupil ratios remain relatively constant. In 2008 the differences between schools in and out of the panel are either not significant (school size and dropout rates) or of the same sign than in 2004 (teacher/pupil ratio). This suggests that, although the schools not in the panel seem to be different from the schools in the panel, these differences are relatively stable over

time; and if this is the case, the fixed-effects model will control for these characteristics preventing the estimate of EPD to be biased.

Section B shows the same figures for Secondary schools (grades 6th to 9th). At baseline the differences for school size is considerably larger than for Elementary schools, but the sign of the differences is the same across years. Schools not in the panel are smaller, have higher dropout rates and lower teacher-pupil ratios than schools in the panel. Out of 8,349 public secondary schools-shifts in 2004, 90 percent are in the panel.

Lastly, Section C presents summary statistics for High schools (grades 10th to 11th). Similarly to Secondary schools, High schools in the panel have more students both in 2004 and 2008, although more so in 2008. The rest of the differences between schools in and out of the panel are not significant for either year (except for the share of EPD teachers, which only applies in 2008).

With the panel samples school-fixed effects models are estimated separately for Elementary, Secondary and High schools.

Before discussing the main results, it is important to highlight that the covariate of interest, the percentage of EPD teachers, is correlated with a number of variables that are usually included in students' outcomes analysis, such as student/teacher ratio or teachers' experience and education.

It is not clear whether these variables should be included or not in the present analysis; if the arrival of EPD teachers is changing other inputs then what happens to those variables should also be considered part of the treatment, if we want that the percentage of EPD captures the 'total' treatment effect, these variables should not be included in the regression independently. On the other hand, we could be interested on the effect of EPD teachers ceteris paribus the education, experience and other inputs related to teachers, in which case we should include these variables as controls despite they are affected by the treatment.

For the sake of illustration I present the main results with and without variables that are directly affected by the presence of EPD teachers. Also, given that the number of students by school varies substantially, I present unweighted as well as weighted regressions results, using the number of students as weights.

In Table 2 results for Elementary, Secondary and High schools are displayed. In Section A we can see unweighted results. In Column (1) only control variables that are not directly affected by the presence of EPD teachers are included, namely shift-fixed effects, a rural-specific trend term

and percentage of students with free tuition [15].

For Elementary schools the results indicate that the percentage of EPD teachers has a negative and significant effect on dropout rates, the coefficient implies that a school that goes from having 0 to 100 percent of its teachers in EPD will reduce its dropout rate in 1.3 percentage points. Considering that the average (unweighted) dropout rate is 8 percent, the estimated effect is fairly large.

In Column (2) teachers' characteristics are included as control variables, these are teacher/pupil ratio, percentage of teachers with a bachelor degree, percentage of teachers with a graduate degree, percentage of teachers with a degree in Education, percentage of contract teachers[16] and percentage of teachers with less than one years of experience. Regarding this last variable it is important to highlight that the C-600 data, unfortunately, does not include information on teachers' experience; considering the importance of this characteristic on teachers' performance (see for example Hanushek et al., 2005) and that is a variable clearly correlated with the share of EPD teachers (as most of them are novice), I constructed a proxy variable for the percentage of teachers with less than one year of experience using a secondary data source[17].

The inclusion of teachers' characteristics and the percentage of contract teachers does not seem to have a big effect on the coefficient on the share of EPD teachers in Elementary schools, which even under the richest specification remains negative and significant.

A possible explanation for these results is that EPD teachers are selecting the better pupils for their schools, so the estimated coefficients are simply picking up the better makeup of the selected students rather than a causal effect of EPD. EPD teachers can be more tempted to select better students if they are worried that low outcomes will endanger their survival as teachers.

This cream skimming can happen in two ways; first, by teachers denying admission to low ability children; and second, by teachers transferring low ability students to other schools.

Regarding the first channel, although placement exams are allowed, it is stipulated that schools

[15] The percentage of students with free tuition is calculated using data from Conpes 116 of 2008.

[16] Contract teachers are neither under the EPD nor the old code, but are governed by short-term contracts. Including this as a control variable might be important as the share of EPD teachers is negatively correlated with the share of contract teachers, which suggests that municipalities had been substituting some of their contract teachers with EPD teachers.

[17] The percentage of teachers with less than 1 year of experience is calculated using the Sistema Nacional de Educación Básica y Media (SINEB); which contains information on all public school teachers in 2009, including the school the teacher is presently working and the year his appointment started. Using this data I calculate the percentage of teachers in 2004 and 2008 with less than one year of experience by school. For this measurement to work as intended the implicit assumption is that novice teachers in 2004 and 2008 still work in the same school in 2009.

Table 2: The effect of EPD teachers on dropout rates

	Dropouts		Transfers	Drop.+Trans.
	(1)	(2)	(3)	(4)
A. Unweighted Regressions				
Elementary schools	-0.013	-0.011	0.016	0.005
	(0.003)	(0.003)	(0.002)	(0.003)
Dep_var_mean	0.080	0.080	0.031	0.111
Secondary schools	-0.010	-0.018	0.010	-0.008
	(0.006)	(0.007)	(0.004)	(0.007)
Dep_var_mean	0.074	0.074	0.024	0.098
High schools	-0.004	-0.006	0.006	0.000
	(0.006)	(0.007)	(0.006)	(0.008)
Dep_var_mean	0.045	0.045	0.016	0.061
B. Weighted Regressions				
Elementary schools	-0.007	-0.007	0.014	0.007
	(0.003)	(0.003)	(0.002)	(0.003)
Dep_var_mean	0.063	0.063	0.027	0.090
Secondary schools	-0.014	-0.017	0.009	-0.009
	(0.005)	(0.005)	(0.004)	(0.006)
Dep_var_mean	0.061	0.061	0.023	0.085
High schools	-0.009	-0.009	0.011	0.002
	(0.006)	(0.006)	(0.005)	(0.007)
Dep_var_mean	0.041	0.041	0.015	0.056
Teachers' characteristics	No	Yes	Yes	Yes

Standard errors clustered at the school level in parentheses

Note: All specifications include school-fixed effects, a year dummy, a rural-specific trend term, four dummies for shift-_xed e_ects, percentage of students with free tuition (at the school level), a dummy variable for school-shifts that report 0 teachers and dummies for item-speci_c missing data. Teachers' characteristics are teacher/student ratio, percentage of teachers with a professional degree, percentage of teachers with a graduate degree, percentage of teachers with a degree in Education, percentage of contract teachers and percentage of teachers with less than one year of experience (see text). Sample sizes are 78,127 for Elementary schools, 15,021 for Secondary schools and 9,283 for High Schools.

cannot demand exams to grant admission[18], so it is unlikely that teachers can control the quality of the students this way.

Regarding the second channel, Column (3) shows results for equation (1) using as dependent variable not dropout but transfer rates. We can see that the percentage of EPD teachers is positively correlated with transfer rates for Elementary schools, which indicates that students are more likely to be transferred out from schools where there are more EPD teachers. However, it is not possible to say if this is actually the result of cream skimming as we do not observe the quality of the students being transferred.

It is difficult to extract a clear interpretation of these findings. EPD teachers seem to be reducing dropout rates but, at the same time, their presence is positively correlated with transfer rates. This suggests that the observed reduction in dropout rates might be in part a consequence of more transfers. However, even if this is the case it is important to highlight that a transferred student is probably preferable that a student that leaves the education system altogether.

An even more conservative interpretation of these results is that transfers are actually misreported dropouts, hence the relevant dependent variable is actually dropout plus transfer rates. When we look at the results for dropout plus transfer rates (Column 4), the coefficient of interest is positive but not statistically significant. Along these lines it can be concluded that, in the extreme case that transfers are actually dropouts, EPD has had no effect on the permanence of children in the education system.

When we look at results for Secondary schools we can see that the effect on dropout rates is negative but not significant for the first specification, but when teachers' characteristics are included the coefficient is larger in absolute value and significant. By including one by one teachers' characteristics (not shown) I found that it was the percentage of contract teachers the variable which inclusion affected the most the coefficient of interest. Because the presence of EPD teachers is negatively correlated with the presence of contract teachers, omitting the percentage of contract teachers causes that the coefficient on EPD teachers picks up not only the effect of EPD but also that of the decline of the presence of contract teachers, which themselves have a negative effect on dropout rates in Secondary schools[19].

[18] Resolution 1515 of 2003, Art. 3.

[19] These regression results are available upon request.

Similarly to Elementary schools, EPD teachers in Secondary schools are positively correlated with transfer rates (Column 3), and when dropout and transfer rates are added up the effect of EPD is negative but not significant.

Finally, EPD teachers seem to have had no effect on dropout rates at High schools under any specification.

Perhaps it should not be surprising that such nil effects are observed at High schools. First, these cohorts were exposed to EPD probably too late in their development process for EPD teachers to have any effect.

Second, even in the medium term we should not expect a negative effect of EPD teachers on dropout rates in higher grades, even if EPD teachers do have a negative effect on dropouts. If EPD teachers are in fact reducing dropouts it is natural to expect that in those schools where children are staying in school longer years dropout rates will fall in lower grades, but this will cause negative selection into higher grades causing that dropout rates in higher grades will probably remain the same or even rise.

In Section B weighted results are presented. For Elementary schools the effect of EPD on dropout rates is negative and significant with and without teachers' characteristics, although the point estimate is lower in absolute value compared to the results from the unweighted regressions. This causes that when the sum of dropout and transfer rates is the dependent variable (Column 4) the effect is *positive* and significant, which implies that if transfers are really dropouts, EPD has had a negative effect on children permanence in the education system. Clearly a better understanding of the true fate of transferred students will improve our knowledge of the effects of EPD.

When we look at weighted results for Secondary schools (Section B) the main results are basically the same than for the unweighted regressions, at least for the more saturated specifications. Regarding High schools, similarly to the results from the unweighted regressions, it can be seen that EPD seem to have had no effect on dropout rates.

In sum the presented results suggest that EPD has a negative effect on dropout rates in Elementary and Secondary schools, although this is nuanced by the fact that the presence of EPD teachers seems to be also (positively) correlated with transfer rates.

Test Scores

I analyze the three main national evaluations: the exams implemented at 5th and 9th grade

(henceforth SABER5&9) and the exams implemented at 11th grade (SABER11).

The SABER5&9 have been applied three times in the last decade, first in 2002-2003, then in 2005-2006 and the last one in 2009. Since 2009 sociodemographic information is collected at the moment of the exam, but no sociodemographic information is available for the 2002-2003 and 2005-2006 examinations. Also, SABER5&9 is designed to provide a picture at the school level, so not all students take the exam.

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, 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 11th grade take it. In addition, SABER11 includes sociodemographic information at the student level, such as household income and mother's education level.

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[20].

Considering the major differences between SABER5&9 and SABER11, separate analyses are presented for these two types of exams.

5th and 9th grade test scores

To analyze the impact of EPD on SABER5&9 I use a data set assembled by ICFES that includes the last three waves and is publicly available in its website. As mentioned before, these exams evaluate students in math, Spanish and science, although historic data is available only for math and Spanish.

Before describing the panel constructed with this data, it is important to highlight that during the application of the examinations in 2002-2003 four different booklets were used, which in principle are not comparable. Similarly, in 2005-2006, 5 different booklets were implemented. To tackle this comparability problem, in 2010 ICFES implemented an 'homogenization' exercise to make the 2002-2003 and the 2005-2006 exams comparable.

However, not all booklets were homogenized. For the 2002-2003 the exercise was performed

[20] 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.

only for 465 thousand test scores, out of 672 thousand realized examinations in 5th grade (for more details see ICFES, 2011). Regarding 9th grade, out of 360 thousand exams performed in 2002-2003, only 242 thousand were made comparable. With respect to 2005-2006, for 5th graders there are 444 thousand available test scores, of a total of 714 thousand realized exams; while for 9th graders there are 393 thousand test scores out of 479 thousand[21].

To link the SABER scores to the C-600 data the latter needs to be aggregated from the school-shift level to the school level, given that the SABER data does not provide information regarding the shift in which test takers are.

To contribute to the fixed effects model a school has to have, at least for two years, data in both SABER and C600. As there are many schools that are left out of the panel, it is important to highlight the reasons for why this happens. First, and most important, the non-homogenization of test scores in 2002-2003 and 2005-2006 described before creates a missing data problem that impedes that a large number of schools in 2009 are matched with corresponding schools in previous years. Second, it is natural that over time new schools are opened, which is especially true for secondary schools as the number of children that reaches this level of education increased; because these schools did not exist in 2005 (or before) they are left out of the panel. Third, it is possible that data entry errors and changing codes prevents a better linking of schools between data sets and across time.

Table 3 shows the result of the merging processes. In Section A results for 5th grade are displayed. The first line shows that in 2002-2003 there are 14,136 schools in the panel, equivalent to 85 percent of the schools for which there is SABER data in 2002-2003. In 2005 there are 15,071 schools in the panel, equivalent to 86 percent of the schools for which there is SABER data in this period. In 2009, the 20,773 schools in the panel represent 71 percent of the schools for which there is SABER data.

Looking at the test scores means (which are negative because these are standardized means and private schools outperform public schools), we can see schools seem to be getting worse over time, as test scores means are falling over time for schools both in and out of the panel, and for math and Spanish. Also, there are no significant differences between schools in and out of the panel for math in either period. For Spanish, on the other hand, there are statistically differences in 2002-2003

[21] These figures include private schools, which are not included in the summary statistics or the regressions.

Table 3: Summary statistics - 5th and 9th grades

	2002-2003			2005			2009		
	Not in Panel (1)	Panel Sample (2)	Diff. (3)=(1)-(2)	Not in Panel (4)	Panel Sample (5)	Diff. (6)=(4)-(5)	Not in Panel (1)	Panel Sample (2)	Diff. (3)=(1)-(2)
A. 5º grade									
SABER									
Schools	2,526	14,136		2,298	15,071		8,654	20,773	
Total students[a]	94,102	289,298		83,993	359,576		173,388	454,736	
Math									
Mean score	-0.06	-0.08	0.02	-0.16	-0.15	-0.02	-0.20	-0.20	-0.004
sd	0.98	1.02		0.89	0.95		0.95	0.87	
Spanish									
Mean score	-0.01	-0.14	0.1***	-0.20	-0.18	-0.01	-0.26	-0.20	-0.05*
sd	0.95	1.01		0.86	0.91		0.91	0.85	
C600									
Schools	19,371	14,136		19,402	15,071		15,156	20,773	
Total students	362,941	297,573		304,842	426,502		223,984	495,695	
Students per grade	18.7	21.1	-2.3***	15.7	28.3	-12.6***	14.8	23.9	-9.1***
Teacher/student[b]	0.03	0.03	0.001***	0.04	0.03	0.002**	0.04	0.04	0.002***
EPD teachers/Total[c]	0.00	0.00	0	0.01	0.01	-0.006***	0.24	0.24	0.002
B. 9º grade									
SABER									
Schools	584	2,782		725	3,996		2,737	4,441	
Total students[a]	26,054	173,801		33,966	259,427		129,161	349,440	
Math									
Mean score	-0.17	-0.13	-0.04	-0.14	-0.13	0.0002	-0.39	-0.22	-0.2***
sd	1.00	0.92		0.96	0.90		0.79	0.73	
Spanish									
Mean score	-0.22	-0.15	-0.06	-0.23	-0.21	-0.01	-0.41	-0.21	-0.2***
sd	0.92	0.90		0.86	0.85		0.78	0.77	
C600									
Schools	2,618	2,782		2,297	3,996		3,261	4,441	
Total students	175,160	178,767		127,851	331,157		141,760	384,202	
Students per grade	67	64	2.6	56	83	-27.2***	43	87	-43.0***
Teacher/student[b]	0.05	0.04	0.0005	0.04	0.04	0.003***	0.04	0.04	0.002**
EPD teachers/Total[c]	0.00	0.00	0	0.01	0.01	-0.003	0.24	0.20	0.05***

Standard errors bootstraped (100)

* p<0.05 ** p<0.01 *** p<0.001

[a] In 2009 only 2 for each 3 students were supposed to take the exam, so I used this factor to impute the number of students the results are

[b]Includes students in the whole corresponding level (Elementary or Secondary)

[c]At the institution level

Note: As there are independent Saber datasets for math and language and they don't have the exact same number of observations, merging with C600 changes a little depending on which one is used, to facilitate the reading of the table results displayed for C600 data come from the merging process with the math dataset, but they are basically the same when the language dataset is used.

and 2009 years, although the differences seem to be small really small relative to their standard deviations.

When we look at the summary statistics from the C-600 data, the missing test scores problem becomes apparent. In 2002-2003 and 2005-2006 the schools in the panel represent respectively 42 and 44 percent of the schools. In 2009 the 20,773 schools in the panel represent 58 percent of the schools[22]. The main reason why so many schools are left out of the panel is that, as mentioned before, many of the exams performed in 2002-2003 and 2005-2006 were not 'homogenized', so no data on test scores is available for a large number of schools. In all three periods schools in the panel have more students on average and the teacher/student ratios are lower than in schools out of the panel.

In sum, the descriptive statistics presented in Section A of Table 3 show two important results from the merging process: i) It was possible to match and important share of the test scores both across time and type of data source. And ii) Although there are differences in teacher/student ratios and other variables between schools in and out of the panel, these differences are either consistent over time or very small with respect to their mean value (or the standard deviation in the case of test scores).

Albeit it is impossible to discard whether schools in the panel sample share unobserved characteristics correlated with the percentage of EPD teachers, the results just described suggest that selection into the panel might be correlated mainly with time-invariant characteristics, which are controlled in the fixed effects model.

An analogous merging exercise was conducted for 9th grade, the results are displayed in Section B of Table 3. Roughly 83 percent of the schools for which there is SABER data are in the panel in 2002-2003 and 2005-2006. In 2009, the number of schools with SABER data that is in the panel is 4,441, equivalent to 62 percent of the schools for which there is SABER data.

Aside the aforementioned problem of non-homogenization of test scores, an important explana-

[22]It can also be seen that in 2009 15,156 schools are left out of the panel according to C-600 data, which is almost double the number schools that are left out the panel in that year for which there is SABER data (8,654). The reason for this is that the files that are used in this work only include the test scores of the schools in 2009 that were match to some school in either 2002-2003 or 2005-2006 (for details see ICFES, 2011). In this sense, although in principle all schools in 2009 could be in the SABER data because test scores in 2009 are comparable without any 'homogenization', due to the way the data used in this work was constructed by ICFES (which, among its most appealing characteristics is that it allows identification of the same schools over time) not all schools for which there is SABER test scores show up in this data not even for 2009.

tion for the large number of schools that are left out in 2009 seems to be simply an increase in the number of schools with students in 9th grade. In effect, according to C-600 data the total number of schools with students in 9th grade increased 22 percent since 2005 (equivalent to 5.1 percent per year), which implies that even if SABER was available for all schools that existed in 2005, a large number of schools in 2009 would be left out of the panel because they are new.

When we look at the test scores data we can see that schools in the panel perform a little better in both subjects in all three years, with the differences being significant only for 2009. Also, schools in the panel in 2005-2006 and 2009 are substantially larger than those not in the panel, while the opposite pattern is observed in 2002; this can constitute a threat to the identification strategy to the extent that there might be other time-varying characteristics correlated with the probability of being in the panel, which could bias the estimate of the effect of EPD; unfortunately with the available data is little what can be done in terms of diagnose or correct this possible threat to identification; however, future research will include a more thorough analysis of the effect of EPD on students' outcomes across schools of different sizes. Finally, in all three periods schools in the panel have lower teacher/student ratios than schools out of the panel.

Table 4 presents the results for school fixed effects regressions of share of EPD teachers on mean test scores. Given that the number of students by school varies substantially, regressions results are presented with and without weights. All specifications include year dummies, rural-specific trend terms, four dummies for shift-fixed effects[23], one dummy for schools with a high probability of fraud in SABER[24] and percentage of students with free tuition at the school. Standard errors are clustered at the school level.

In Column (1) the results without using teacher characteristics are displayed. We can see that the share of EPD teachers has a negative effect on test scores, although the parameter is not significant. In Column 2 teachers' characteristics are included, these are teacher/pupil ratio, percentage of teachers with a bachelor degree, percentage of teachers with a graduate degree and percentage of

[23] Although SABER5&9 does not provide data on shifts, with C-600 data I can impute the set of shifts a given school has in a given year; fortunately, most schools actually have only one or two shifts, so I create a set of four indicator variables for schools with morning shift, afternoon shift, morning and afternoon shifts, and other (the left out category is whole day) to try to control for the type of shifts each school serves.

[24] According to an analysis performed by Martínez (2010), a massive cheating problem was detected in 2002-2003 in SABER examinations. Over 16 thousand schools with high probability of committing fraud were detected in 5th grade in 2002-2003. The same exercise was performed for 5th grade in 2009 and for both periods for 9th grade, but the incidence was much lower in these cases.

Table 4: The effect of EPD teachers on 5th and 9th grades test scores

	Math		Spanish	
	(1)	(2)	(3)	(4)
A. Unweighted Regressions				
5th grade	-0.020	-0.023	0.026	0.023
	(0.036)	(0.037)	(0.036)	(0.036)
9th grade	0.243	0.236	0.253	0.242
	(0.082)	(0.083)	(0.069)	(0.070)
B. Weighted Regressions				
5th grade	0.097	0.104	0.056	0.059
	(0.039)	(0.040)	(0.037)	(0.037)
9th grade	0.296	0.293	0.234	0.233
	(0.080)	(0.080)	(0.063)	(0.063)
Teachers' characteristics	No	Yes	No	Yes

Standard errors clustered at the school level in parentheses

Notes: All specifications include school-fixed effects, year dummies, rural-specific trend terms, four dummies for shift-fixed effects (see text), one dummy for schools with a high probability of fraud in SABER (see text), percentage of students with free tuition and one dummy for schools that report 0 teachers. The weights for the weighted regressions are the average number of test takers over the three years by school. Teachers characteristics are 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. Sample sizes are 49,980 for 5th grade math, 49,737 for 5th grade Spanish, 11,219 for 9th grade math and 11,220 for 9th grade Spanish.

teachers with a degree in Education. Unfortunately, variables included in the dropout rate analysis, namely the percentage of contract teachers and the percentage of teachers with less than one years of experience, are not available for 2002, so I cannot include them in this specification[25]. The inclusion of teachers' characteristics does not affect too much the effect of the share of EPD teachers, which remains negative but not significant. Similar results for Spanish can be seen in Columns (3) and (4). The effect of the share of EPD teachers is positive but not significant, and is not too sensitive to the introduction of other teachers' characteristics.

The results are very different when we look at 9th grade test scores. For both math and Spanish there is a positive and significant effect of EPD teachers, regardless of whether other teachers' characteristics are included or not. The results indicate that if a school goes from having 0 to 100 percent teachers under the EPD, test scores will increase approximately a fourth of a standard deviation in math and Spanish.

In Section B we can see the results for regressions where the average number of students by school are used as weights. For 5th grade math the effect of EPD is now positive and significant, while the effect for Spanish is also positive but not significant. Results for 9th grade are not sensitive to the use of weights.

11th grade test scores

The SABER11 exams are presented yearly by all students finishing high school and it collects 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[26].

Similarly to what was done for SABER5&9, to build the panel I link the C-600 and the SABER11

[25]Perhaps the main limitation of the test score analysis is the lack of data on teachers' experience. Omitting this variable is problematic because experience is for sure correlated with the covariate of interest, as EPD teachers are mostly novice novice. Although for dropout rates I imputed the percentage of teachers with less than one year of experience, this imputation was not calculated for the test score analyses as this would had implied using data from to 2009 to impute teachers' experience as back as 2002. The Ministry of Education is currently collecting more complete data on teachers, in the near future it will be possible to elaborate much richer specifications, which will allow a better understanding of the effect of EPD teachers ceteris paribus other characteristics.

[26]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.

data sets using a school-shift id number, although for some cases I only used the school id number[27].

A school is considered to be part of the panel if it has complete information (SABER11 and C-600) for at least two years, this way it contributes to the fixed effects model. In Table 5 summary statistics are presented for schools in and out of the panel.

Looking at the section on C-600 data we can see that 2,754 schools are in the panel in 2002, while 1,481 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 higher too.

Regarding SABER11 summary statistics, is not very surprising that all differences between schools in and out of the panel are significant, considering that the data is at the student level so the sample sizes are quite large. 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 small too.

With respect to 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 4,668 and 4,941 respectively. Similarly to what is observed for 2002, schools in the panel are bigger than schools out of the panel, but the teacher/student ratios are lower, although the differences are very small relative to the mean.

The share of EPD teachers in panel schools is lower than in schools that are not in the panel, both in 2008 and 2009.

With respect to SABER11 data, the displayed figures show very similar patterns to what was observed for 2002, 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 link an important

[27]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 data sets 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'.

Table 5: Summary Statistics - 11th grade

	2002			2008			2009		
	Not in Panel (1)	Panel Sample (2)	Dif.[a] (3)=(1)-(2)	Not in Panel (4)	Panel Sample (5)	Dif.[a] (6)=(4)-(5)	Not in Panel (1)	Panel Sample (2)	Dif.[a] (3)=(1)-(2)
C600									
Schools	1,481	2,754		1,533	4,668		1,420	4,941	
Total Students	69,544	170,694		72,046	294,518		66,246	308,051	
Students by grade	47	62	-15.0***	47	63	-16.1***	47	62	-15.7***
Teacher/students[b]	0.05	0.05	-0.003*	0.05	0.04	0.003***	0.05	0.04	0.003**
Share of EPD teachers[c]	0.00	0.00	-	0.17	0.13	0.04***	0.24	0.19	0.05***
SABER									
Schools	3,720	2,754		1,023	4,668		1,041	4,941	
Total students	75,674	164,182		51,995	282,323		51,722	297,077	
Math									
Mean Test Score	42.2	42.3	-0.1***	43.4	44.1	-0.7***	42.7	43.3	-0.6***
Standard Deviation	6.0	6.0		8.1	8.2		9.6	9.6	
Spanish									
Mean Test Score	47.5	47.8	-0.3***	44.9	45.4	-0.5***	45.5	45.9	-0.4***
Standard Deviation	8.4	7.8		7.0	6.9		7.2	7.2	
Females (%)	0.53	0.55	-0.02***	0.54	0.55	-0.009***	0.54	0.55	-0.01***
Education of the mother (%)			0.000			0.000			0.000
Elem. Or less	0.58	0.55		0.44	0.42		0.46	0.43	
Secon. Or HS studies	0.33	0.36		0.44	0.45		0.44	0.45	
Higher Ed. studies	0.09	0.09		0.11	0.13		0.10	0.12	
Monthly household income in minimum wages (%)			0.000			0.000			0.000
<1	0.36	0.33		0.35	0.33		0.37	0.35	
>=1 and <2	0.42	0.43		0.48	0.47		0.46	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 (%)			0.001			0.000			0.000
1-4	0.34	0.34		0.42	0.42		0.42	0.42	
5-8	0.58	0.59		0.51	0.52		0.52	0.52	
9 or more	0.08	0.07		0.06	0.06		0.06	0.06	

* p<0.05 ** p<0.01 *** p<0.001

[a]For categorical variables the statistic displayed is the p-value of a χ^2 test.

[b]Includes students in grades 6th to 11th.

[c]At the institution level

Note: The number of test scores in math and Spanish are a little different, to facilitate the reading of the table the total number of test scores corresponds to math.

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 panel sample will be controlled by the school fixed effects and therefore it will not constitute a source of bias on the effect of EPD.

Table 6 presents results for equation (2) on SABER11. In Section A results for math are displayed. The first column controls only for time-variant characteristics at the school level, so no school-fixed effects or student characteristics are included. We can see that the effect of EPD is negative and significant. In Column (2) sociodemographic characteristics 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 (Columns 3-4), 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. Furthermore, it is important to highlight that the coefficients on other teachers' characteristics are not very sensitive to the inclusion of student sociodemographic information, once school fixed effects are controlled for. A similar story described the results for Spanish (Section B), without school-fixed effects the coefficient of interest es negative and significant, but when school-fixed effects are included the coefficient is also negative but not significant.

The presented results indicate that EPD did not have much effect on SABER11, which is at odds with the results obtained before, especially for 9th grade test scores.

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

Table 6: The effect of EPD teachers on 11th grade test scores

	No FE		FE	
	(1)	(2)	(3)	(4)
A. Math				
Share of EPD teachers	-0.105	-0.064	-0.001	-0.015
	(0.025)	(0.021)	(0.023)	(0.023)
Teacher/student ratio	0.162	0.232	0.148	0.094
	(0.168)	(0.131)	(0.117)	(0.114)
Share over total teachers:				
- Teachers with professional degree	-0.009	0.001	0.016	0.020
	(0.018)	(0.015)	(0.012)	(0.011)
- Teachers with graduate degree	0.177	0.145	0.012	0.019
	(0.021)	(0.018)	(0.013)	(0.013)
- Teachers with degree in Ed.	-0.006	-0.001	-0.017	-0.013
	(0.018)	(0.015)	(0.012)	(0.012)
Sociodemographic controls	No	Yes	No	Yes
Observations	743582	743582	743582	743582
B. Spanish				
Share of EPD teachers	-0.093	-0.050	-0.020	-0.035
	(0.024)	(0.019)	(0.025)	(0.024)
Teacher/student ratio	-0.088	0.170	0.451	0.405
	(0.174)	(0.131)	(0.114)	(0.108)
Share over total teachers:				
- Teachers with professional degree	-0.010	-0.006	0.001	0.003
	(0.020)	(0.017)	(0.013)	(0.013)
- Teachers with graduate degree	0.174	0.126	-0.011	-0.006
	(0.023)	(0.020)	(0.016)	(0.016)
- Teachers with degree in Ed.	-0.009	-0.006	-0.021	-0.019
	(0.019)	(0.015)	(0.014)	(0.013)
Sociodemographic controls	No	Yes	No	Yes
Observations	743591	743591	743591	743591

Standard errors clustered at the school level in parentheses

Notes: 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 (one for secondary studies and one for higher education studies, the left out category is elementary studies); three dummies for household income (dummies indicate between one and two minimum wages, two and three minimum wages, and more than three minimum wages, the left out category is less than one minimum wage); household size and a dummy variable for whether the student works. Missing data in the covariates are accounted for with missing data dummy variables.

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. Although is not possible to say what would happen in the context of 5th and 9th grade test scores if it was possible to include students' characteristics, the evidence provided by the 11th grade test scores analysis suggests that student characteristics play a minor role if school fixed effects are included.

A couple of more plausible explanations can be borrowed from the dropout analysis. The first one is that the cohorts graduating in 2008 and 2009 were exposed to EPD teachers probably too late in their development process. To test whether this is the case further research should focus on the effects on SABER11 on cohorts graduating later than 2009.

The second is related to the negative selection associated to the effect of EPD teachers on dropout rates. If EPD teachers are reducing dropout rates this implies that they are modifying also the underlying endowment distribution in their schools, as less students are dropping out. If this change in the ability distribution is not random, then the estimates presented on the effect of EPD teachers on test scores are probably biased especially in higher grades, as the coefficients are picking up not only the direct effect of EPD teachers on the achievement of their students, but also the effect of the change in the underlying endowment distribution in schools. As students on the margin of dropping out are presumably less able than their peers, the estimated coefficients in higher grades should probably be considered underestimations of the true effects of EPD on test scores. This negative selection process could explain the zero effects in both High school dropout rates and 11th grade test scores. To test these hypothesis more information regarding the characteristics of students dropping out of school are needed. The recently collected *Encuesta Nacional de Deserción Escolar* will constitute a key input for this research agenda.

5. Final Comments

The introduction of EPD constitutes the most important policy of the last decade regarding the teaching profession in Colombia. The empirical evidence presented in this paper suggests that this input-based accountability system has had some important effects. I found that EPD teachers are negatively correlated with dropout rates in Elementary and Secondary schools, but these results are somehow nuanced by the fact that EPD teachers are also associated with higher transfer rates. Regarding test scores, I found positive effects for 5th grade math (albeit sensitive to the use of weights), positive and significant effects for 9th grade math and Spanish, and no effects for 11th grade test scores.

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 endogenous to 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, it will be possible to test some of these hypotheses regarding the effects of EPD teachers on 11th graders.

References

Felipe Barrera and Camilo Domínguez. Educación básica en Colombia: Opciones futuras de política. Technical report, Departamento Nacional de Planeación, 2005.

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

Marcela Bautista. La profesionalización docente en Colombia. 2009.

Leonardo Bonilla. Doble jornada escolar y calidad de la educación en Colombia. 2011.

Thomas S. Dee and Benjamin J. Keys. Does merit pay reward good teachers? Evidence from a randomized experiment. *Journal of Policy Analysis and Management*, 23(3):471–488, 2004. ISSN 0276-8739. doi: 10.1002/pam.20022. URL `http://doi.wiley.com/10.1002/pam.20022`.

Jesus Duarte. La debilidad del Ministerio de Educación y la politización de la educación pública en Colombia: dos problemas a enfrentar en el Plan Decenal. *Coyuntura Social*, (15), 1996.

Esther Duflo, Pascaline Dupas, and Michael Kremer. School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools. *NBER Working Paper Series*, (17939), 2012.

Alejandro Gaviria and Jorge H. Barrientos. Determinantes de la calidad de la educación en Colombia. *Archivos de Economía*, (159), 2001.

Paul Glewwe and Michael Kremer. Teacher incentives. *National Bureau of Economic Research*, (9671), 2003.

Eric A. Hanushek, Steven G. Rivkin, and John F. Kain. Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458, 2005.

Brian A Jacob and Lars Lefgren. PRINCIPALS AS AGENTS : SUBJECTIVE PERFORMANCE. 2005.

Thomas J Kane and Douglas O Staiger. The Promise and Pitfalls of Using Imprecise School Accountability Measures. *Journal of Economic Perspectives*, 16(4):91–114, December 2002. ISSN 0895-3309. doi: 10.1257/089533002320950993. URL `http://pubs.aeaweb.org/doi/abs/10.1257/089533002320950993`.

Karthik Muralidharan and Venkatesh Sundararaman. Teacher Performance Pay : Experimental Evidence from India. *Journal of Political Economy*, 119(1):39–77, 2011.

Jairo Núñez, Roberto Steiner, Ximena Cadena, and Renata Pardo. Â¿Cuáles colegios ofrecen mejor educación en Colombia? *Documento CEDE*, pages 1–56, 2002.

Michael J Podgursky and Matthew G Springer. Teacher Performance Pay : A Review. *Policy Analysis*, 26(4):909–949, 2007. doi: 10.1002/pam.

Matthew G Springer, Brian M Stecher, Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel F. McCaffrey, and Matthew Pepper. Teacher Pay for Performance. Technical report, 2010.

Claudia Marcela Umaña. Esquemas de incentivos para la carrera docente. *Archivos de Economía*, 2004.