DISCUSSION PAPER SERIES

IZA DP No. 8477

Broken Gears: The Value Added of Higher Education on Teachers' Academic Achievement

Carlos Felipe Balcázar
Hugo Ñopo

September 2014

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Broken Gears: The Value Added of Higher Education on Teachers' Academic Achievement

Carlos Felipe Balcázar
World Bank Group

Hugo Ñopo
Inter-American Development Bank and IZA

Discussion Paper No. 8477
September 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 8477
September 2014

ABSTRACT

Broken Gears: The Value Added of Higher Education on Teachers' Academic Achievement*

A growing literature establishes that good teachers are essential for high quality educational systems. However, little is known about teachers' skills formation during their college years. In this paper we use a novel panel data set combining two standardized tests for Colombian students: one that is taken at the end of senior year in high school and the other when students are near graduation from college. Accounting for selection into majors we test for the extent to which education majors relatively improve or deteriorate their skills in comparison to students in other programs. We analyze three sets of skills: quantitative reasoning, native language (Spanish) and foreign language (English). After around 5 years of college, teachers' skills vis-à-vis those in other majors deteriorate in quantitative reasoning, although they deteriorate less for those in math-oriented programs. For native and foreign language we do not find evidence of robust changes in relative learning.

JEL Classification: I2, I21, J24

Keywords: teacher performance, career choice, self-selection, relative learning mobility

Corresponding author:

Carlos Felipe Balcázar
Poverty Global Practice
The World Bank Group
1818 H St. NW
Washington, D.C. 20433
USA
E-mail: cbalcazarsalazar@worldbank.org

* The comments of Mariana Alfonso, Fernando Fernández, Catherine Rodríguez and Emiliana Vegas are specially acknowledged. Daniel Alonso provided valuable research assistance. The findings herein do not necessarily represent the views of the Inter-American Development Bank, The World Bank Group, or their Boards of Directors.

1. Introduction

This document builds upon two literatures of high policy relevance, not only in the developing world but also in the developed one: *teacher quality* and *value added of higher education*. It does it by providing evidence for Colombia on the difference in skill-acquisition between students in education majors and students in other majors. To do so we compare test score-rankings between these two groups of students using a novel panel data set that combines two standardized tests, one before and one near the end of college studies. We measure test score-rankings using the z-scores of subjects assessed in both tests and compare changes in the relative position of students in the distributions of skills.

A growing body of literature establishes that teachers are essential for a high-quality educational system (Goldhaber and Brewer, 1997; Hanushek, 2002; Eide et al., 2004; Rockoff, 2004; Rivkin et al., 2005; Hanushek and Rivkin, 2006; Clotfelter et. al., 2007; Kukla-Acevedo, 2009; De Paola, 2009). More recently, Chetty et al. (2013a and 2013b) find that, after controlling for socio-demographic characteristics, those students who had high quality teachers during their school years are more likely to attend higher education, to go to a better university and to obtain higher labor earnings. Nonetheless, despite that the literature has extensively addressed the role of the quality of teachers on social and academic outcomes, we know little about the role of higher education on the production of those teachers.

Evidence from successful educational systems (Finland, Singapore, Korea, China) highlights the importance of teachers and their formation. In Finland, for instance, only the best and brightest manage to become teachers on that system, after demanding undergrad and master's programs (OECD, 2011; Sahlberg, 2011). This has been claimed as an important stepping stone for their impressive improvement between the 60s and the early 21st century: from having educational outcomes comparable to those of developing economies, Finland now ranks amongst the top performers in educational achievement and attainment in the world (Sahlberg, 2009, 2011).

In most educational systems, however, the best and brightest do not choose education majors. Teachers' academic performance and potential is lower than that of their peers who choose to attend other programs (Giesen and Gold, 1993; Hanushek and Pace, 1995; Podgursky et al., 2004; Denzler and Wolter, 2008). In this context, competitive salaries and benefits, as well as a meaningful and challenging career path for teachers are crucial for attracting and retaining high-quality teachers (Murnane et al., 1991; Ballou and Podgursky, 1997; Odden and Kelley, 1997; Dolton, 1990, 2006; Dolton and van der Klaauw 1999). Countries that pay low salaries to their teachers (relative to the earnings distribution of the country's population) show lower pupil performance as measured by international student assessments (Dolton and Marcenaro-Gutierrez, 2011).

In Latin America, and particularly in Colombia, the situation is no different. Talented people might get discouraged from choosing a career in education to avoid getting trapped in a profession with low social prestige, bad academic reputation and low salaries (Mizala and Ñopo, 2011; García et al., 2014). As it has been shown, in Colombia, those who graduate as school teachers from university or technical programs are more likely to have scored at the bottom quintiles of the Colombian college admission test, Saber 11 (Barón and Bonilla, 2011; Barón et al., 2013).

There is solid evidence that self-selection plays a role on teachers' quality. Granted, but, what about the role of higher education? Even if prospective teachers start at disadvantage compared to their peers in other careers, does higher education level the playing field? To address these questions we analyze *relative learning mobility*, understood as the change in score-rankings between two periods, for students who are close to graduate from a university program in education (which we will also call throughout the paper: *teachers*) vis-à-vis students that are also about to graduate, but from other university academic programs (which we will also call throughout the paper: *other professionals*). We use data from two standardized tests: Saber 11 and Saber Pro, which provide information on educational outcomes and socio-demographic characteristics for students in Colombia at their senior year of high school and near the end of their college education. In this sense this paper not only deals with teachers' academic competence but also contributes to the *value added of higher education* literature (Klein et al, 2005; Liu, 2011; Saavedra, 2009; Saavedra and Saavedra, 2011; Cunha and Miller, 2014).

Our results indicate that after around 5 years of academic training, teachers' skills vis-à-vis those in other professions deteriorate in quantitative reasoning, although they deteriorate less for those in math oriented programs. We do not find evidence of a statistically significant change in relative learning for native language (Spanish) and foreign language (English). The rest of the paper proceeds as follows. In the next section we introduce the data sources and present some descriptive statistics. Section three introduces the concept of relative learning mobility. In the fourth section we present the main empirical analysis. In the fifth section we conclude.

2. Data and descriptive statistics

The data comes from *Instituto Colombiano para la Evaluación de la Educación Superior* (ICFES). We use data from two national standardized tests: Saber 11 (2002-2007) and Saber Pro (2011). *Saber 11* assesses senior year high school students' academic competences in language (Spanish), mathematics, biology, chemistry, physics, social sciences, philosophy, and foreign language (English). *Saber Pro* assesses the academic competences of higher education students who have completed at least 75% of their academic programs, on program specific areas (e.g., engineers are assessed in engineering; economists in economics; biologists in biology, and so on). The novelty that makes this paper feasible is that since the second semester of 2011 Saber Pro also assesses all students, regardless of their major, in quantitative reasoning, reading, writing and English as foreign language.[2]

Saber 11 is mandatory for all senior year high school students who wish to obtain their school degree, and serves as an input for college entrance. In fact, in many universities this is the sole admission criterion (Saavedra and Saavedra, 2011). Saber Pro is mandatory for those students who wish to earn their college degrees. Furthermore, employers might use Saber Pro scores to screen applicants for a (professional) job position (Saavedra, 2009). Nonetheless, schools and universities cannot retain students based on their scores.

[2] See: http://www.icfes.gov.co/examenes/.

Saber 11 and Saber Pro share three components, which allow us to compare the "change" in students' learning: native language, quantitative reasoning and foreign language. "Language" in Saber 11 is comparable to "Reading" in Saber Pro as both evaluate academic competences in Spanish (reading skills). In this paper we use "Native Language" to refer to both. "Mathematics" in Saber 11 is comparable to "Quantitative Reasoning" in Saber Pro. In this paper we use "Quantitative Reasoning" to refer to both. The foreign language component in both tests evaluates proficiency in English. In all three cases, the level of difficulty is higher in Saber Pro, capturing the fact that students should develop their academic skills during higher education. We will not compare scores in both tests in a direct way. Instead, we will use the individuals' position on the distribution of scores (z-scores), analyzing the changes from one test to the other.

ICFES provided researchers the possibility of linking both tests using a unique key that identifies individuals. These keys are available for Saber 11 from the first semester of 2002 to 2010, and for Saber Pro from the first semester of 2007 to 2011. However, we do not use data from Saber 11 from the second semester of 2007 on due to changes in the survey structure, and from Saber Pro from the first semester of 2011 backwards given that quantitative reasoning, native language and foreign language were not assessed.

We restrict ourselves to students attending on-campus programs at their 4th year of college or above, who at the time of the test had completed at least 75% of their academic programs, whose identification key allows tracing them in both Saber 11 and Saber Pro, and with no missing values in the variables of interest. This comprises 67% of the full sample of college students who took Saber Pro in the second semester of 2011 (60% of the education majors and 68% of those in other majors). Those students for which tracing is possible and show no missing values on the variables of interest, that is, those in the 67%, scored above the remaining 33% in the three subjects (on average): 0.26 standard deviations in quantitative reasoning, 0.26 standard deviations in native language, and 0.30 standard deviations in foreign language. The sample of interest is then biased towards higher performing students but such bias is similar for both education and other majors.

Table 1 shows the sample distribution of the 67% of students by the semester in which they took Saber 11, split by group: teachers (students majoring in education, 8% of the sample) and other professionals (students majoring in other programs). The bulk of students in the second semester of each year is explained by the fact that most students at the national level (85%) are enrolled in the *A school calendar*, which goes from February to November. Students in the *B school calendar* (2% of the population of interest), which goes from August/September to June, take the test in the first semester of the year. The remaining 13% corresponds to *flexible calendar* students (sabbatine, fast-tracks, etc.). From these, 77% usually take the test on the second semester of each year and the remaining 23% on the first one.

Table 1. Sample size

Test	Year	Semester	Teachers		Other professionals	
			Number of observations	Percentage (%)	Number of observations	Percentage (%)
Saber 11	2002	I	36	0.87	393	0.84
		II	300	7.22	3038	6.52
	2003	I	55	1.32	585	1.25
		II	485	11.67	4362	9.36
	2004	I	58	1.40	830	1.78
		II	700	16.85	6773	14.53
	2005	I	101	2.43	1310	2.81
		II	958	23.06	10288	22.07
	2006	I	103	2.48	2133	4.58
		II	1278	30.76	14649	31.42
	2007	I	81	1.95	2256	4.84
Saber Pro	**2011**	**II**	**4155**	**100.00**	**46617**	**100.00**

Source: Authors' compilations based on ICFES data.

2.1. Differences in socio-demographic characteristics between teachers and other professionals

Table 2 presents socio-demographic characteristics for students of education majors and students of other programs. The share of females studying to become teachers is higher than in other programs. Also, students of education majors are more likely to come from larger families; they have less educated parents/guardians; they are less likely to migrate to another administrative unit to pursue their studies; they are more likely to have studied in public schools; and also they are more likely to enroll in public universities and in programs with lower tuition fees. This reaffirms the idea that teachers are more likely to come from a disadvantaged socioeconomic position compared to their peers (Mizala and Ñopo, 2011).

Table 2. Descriptive statistics

a. Characteristics at the student level

Variables	Other professionals	Teachers	Difference between groups
Socio-demographic characteristics (%)			
Gender (Female, as measured by Saber 11 and Saber Pro)	58.10	66.60	***
	(0.23)	(0.73)	
Familiy size (more than 5 persons, as measured by Saber Pro)	14.30	22.30	***
	(0.16)	(0.65)	
Max education of the parents/guardians (as measured by Saber Pro)			
Secondary incomplete or less	19.70	39.70	***
	(0.18)	(0.76)	
Secondary complete or tertiary incomplete	28.10	34.20	***
	(0.21)	(0.74)	
Technical or technician education complete	12.10	10.30	***
	(0.15)	(0.47)	
Universitary education complete	40.00	15.80	***
	(0.23)	(0.57)	
The student moved to another administrative unit for his/her higher education (as measured by a difference in the administrative unit of residence in Saber 11 and Saber Pro)	37.90	27.10	***
	(0.22)	(0.69)	
Semester of study in current university program (as measured by Saber Pro)			
7 or 8	11.20	9.10	***
	(0.15)	(0.45)	
9 or 10	77.70	81.80	***
	(0.19)	(0.6)	
11 or 12	11.10	9.10	***
	(0.15)	(0.45)	

(*Continues on next page*)

b. Characteristics at the high school level

Variables	Other professionals	Teachers	Difference between groups
High school characteristics (%), as measured by Saber 11 unless otherwise noted			
School administration (public)	48.00	74.80	***
	(0.23)	(0.67)	
School type (mixed gender)	79.1	87.7	***
	(0.19)	(0.51)	
School day			
Complete	32.30	23.90	***
	(0.22)	(0.66)	
Morning	51.70	51.80	
	(0.23)	(0.78)	
Afternoon	14.50	21.10	***
	(0.06)	(0.26)	
Night	1.40	2.90	***
	(0.02)	(0.08)	
Weekend	0.10	0.30	
	(0.16)	(0.63)	
School Calendar (A calendar)†	82.10	87.30	***
	(0.18)	(0.52)	
Degree type (as measured by Saber Pro)			
Academic	78.40	69.40	***
	(0.15)	(0.45)	
Technical	19.10	21.40	***
	(0.19)	(0.6)	
Superior normal school	2.60	9.20	***
	(0.18)	(0.64)	

c. Characteristics at the university level

Variables	Other professionals	Teachers	Difference between groups
University characteristics (%), as measured by Saber Pro			
University administration (public)	34.60	73.80	***
	(0.18)	(0.56)	
University fee (academic semester)††			
None	1.00	1.40	*
	(0.05)	(0.19)	
Less than 1,000,000 COP	27.30	72.90	***
	(0.21)	(0.69)	
Between 1,000,000 and 3,000,000 COP	30.90	23.20	***
	(0.21)	(0.66)	
Between 3,000,000 and 5,000,000 COP	20.80	2.30	***
	(0.19)	(0.23)	
More than 5,000,000 COP	20.00	0.20	***
	(0.19)	(0.08)	

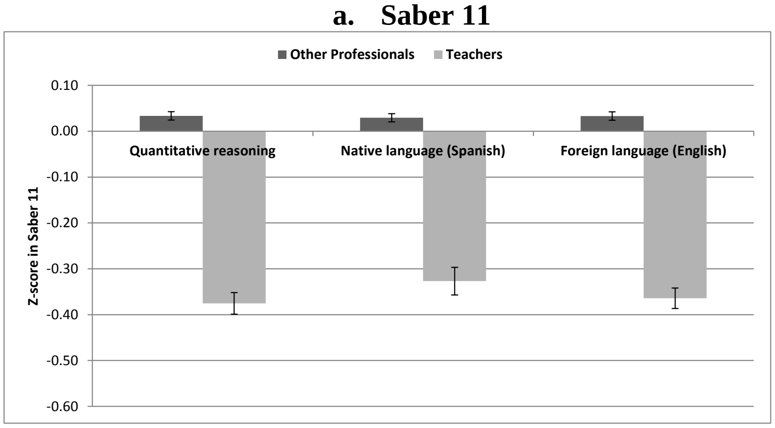
Standard errors in parentheses. * Significant at ten percent; ** significant at five percent; *** significant at one percent.
† Colombia has two regular school calendars: The "A calendar", which goes from February to November; and the "B calendar", which goes from August to June.
†† COP: Colombian Peso. The average exchange rate for the second semester of 2011 was 1856 COP/USD.
Source: Authors' calculations based on ICFES data.

As mentioned above, the common components in Saber 11 and Saber Pro are not directly comparable. Although they evaluate the same subjects, the scores have different metrics. To perform the comparisons we use standardized test scores (z-scores), analyzing the changes in the students' rankings between the two tests. Figure 1 shows the average standardized scores (and confidence intervals for the means) for teachers and other professionals in both tests and the three subjects assessed. As it is clear from the figure, teachers underperform in comparison to their peers in all the three subjects in both tests.

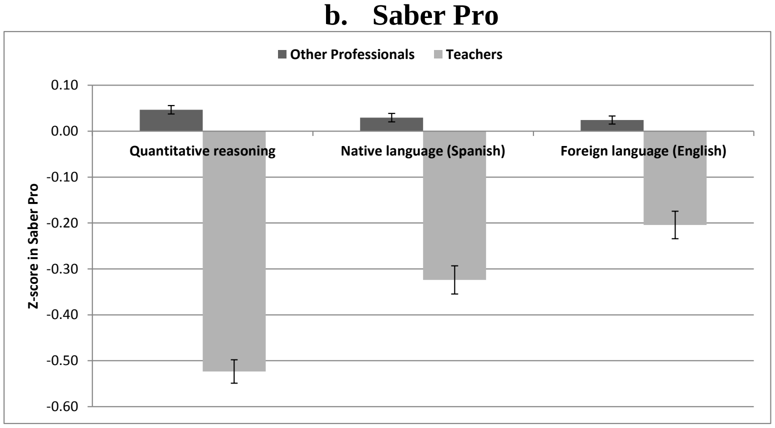
Low scores in Saber 11 might also work as incentives for underachievers to choose the teaching profession, as admittance cut-off scores for teaching careers are low. Indeed, Figure A1 in the appendix shows that the probability of being a teacher is lower the higher the score attained in Saber 11 for our three common subjects. Therefore, low scores and socioeconomic disadvantage might narrow some students' possibilities when choosing a college program. This is the non-random selection into teaching that we will control for in our empirical strategy.

Figure 1. Average scores of teachers and non-teachers in Saber 11 and Saber Pro

a. Saber 11



b. Saber Pro



Source: Authors' calculations based on ICFES data.
Note: Confidence intervals at 95% added.

3. Relative learning mobility

Relative learning mobility is assessed by the change in z-scores in quantitative reasoning, native language and foreign language between Saber 11 and Saber PRO. Figure 2 illustrates such *relative learning mobility* for some selected university programs. The education majors are grouped into four categories, according to their emphasis: quantitative reasoning (mathematics, physics, biology, etc.), native language (Spanish, social sciences, humanities, philology), foreign language (mainly English, although there are students of some other foreign languages), and others (preschool, arts, sports and others).[3] Among the four groups of education majors, only those emphasizing native language and foreign languages show a sizeable improvement in the English and Spanish skills of their students. For all other groups of teaching careers their average z-scores fall. Out of the 53 non-teaching careers, in contrast, many improved their z-scores: 30 in quantitative reasoning, 24 in native language and 30 in foreign language (not shown but available upon request to the authors).

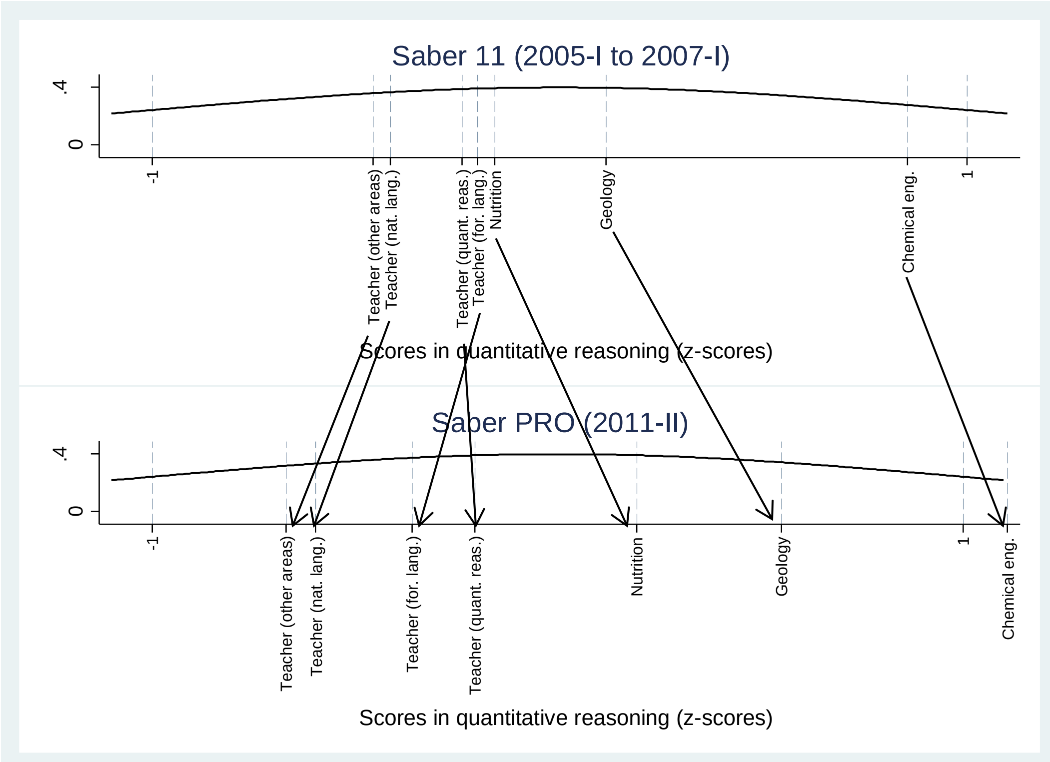
For the selected majors shown, the figure makes evident that after around five years of college, gaps in quantitative reasoning between school teaching programs and other (selected) academic programs have widened.[4] That is, teachers' skills in quantitative reasoning either deteriorate or do not improve as much as those of professionals in other areas of knowledge.

[3] See the online appendix for information on how we classified each major in education into each of these four categories: https://sites.google.com/site/cfbalcazars/misc

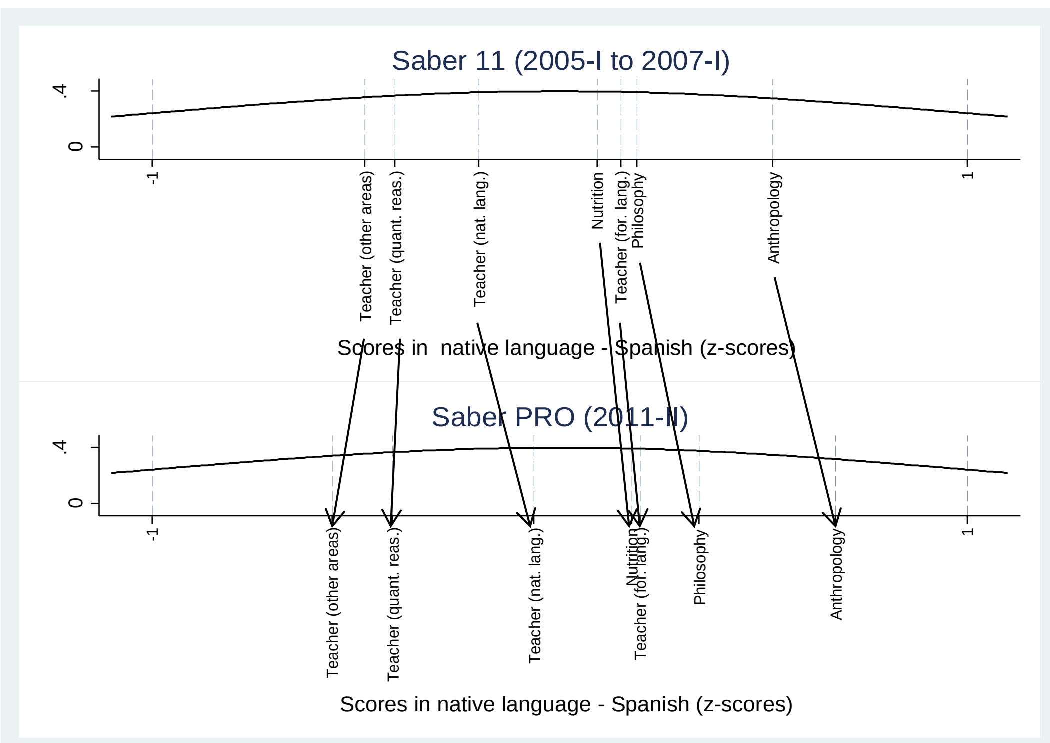
[4] We selected those programs with the highest changes in their z-scores between Saber 11 and Saber Pro for our illustrative purposes.

Figure 2. Distribution of standardized test scores in Saber 11 and Saber Pro, selected programs highlighted

a. Quantitative reasoning

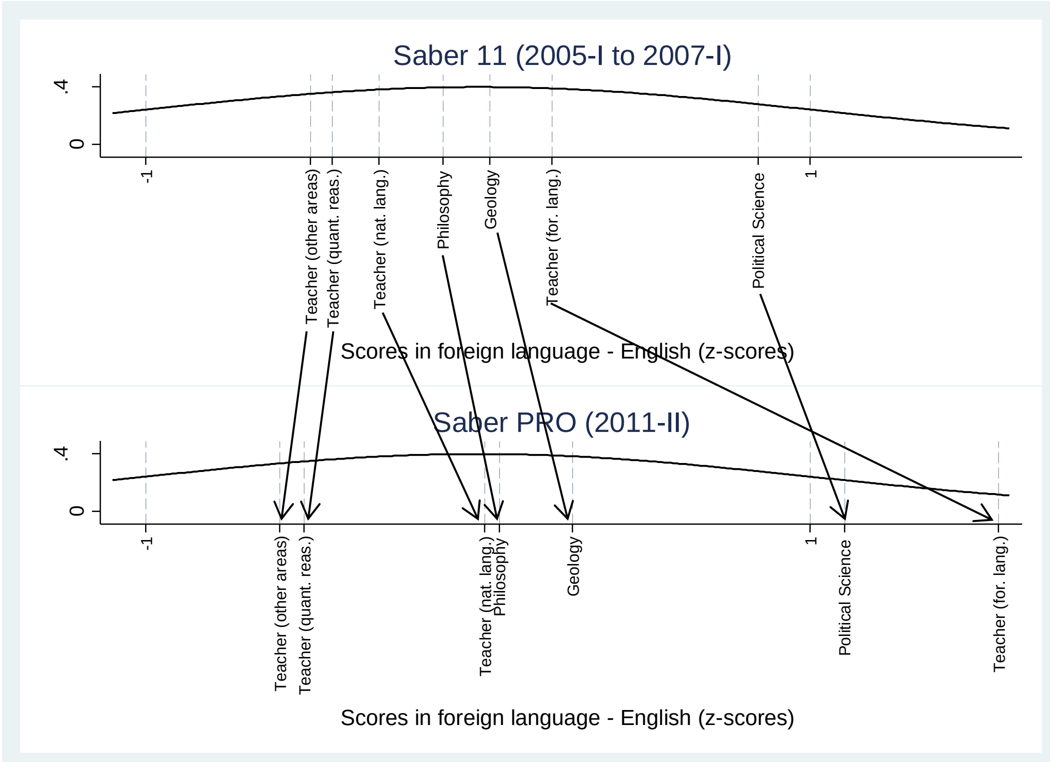


b. Native language (Spanish)



(*Continues on next page*)

c. Foreign language (English)



Source: Authors' calculations based on ICFES data.

To assess relative learning mobility we estimate:

$$y_{i,1} = \beta_1 + \beta_2 t_{i,1} + \beta_3 y_{i,0} + \beta_4 \left(t_{i,1} * y_{i,0}\right) + \varepsilon_i \qquad (1)$$

where $y_{i,1}$ represents the z-score attained in Saber Pro by individual i; $t_{i,1}$ is a dummy variable that takes the value 1 if the individual i is found in saber PRO as an education major and 0 if she/he is in another program; $y_{i,0}$ is the z-score attained in Saber 11 by i; ε_i is an idiosyncratic error term.

Given that some education majors emphasize on the subjects assessed, we proceed to capture this particularity by adding a dummy variable, $h_{i,1}$, that takes the value 1 if the individual i is studying an education major that makes emphasis in either quantitative reasoning, native language or foreign language –in accordance with the dependent variable. Therefore (1) becomes:

$$y_{i,1} = \beta_1 + \beta_2 t_{i,1} + \beta_3 y_{i,0} + \beta_4 \left(t_{i,1} * y_{i,0}\right) + \beta_5 h_{i,1} + \beta_6 \left(h_{i,1} * y_{i,0}\right) + \varepsilon_i \quad (2)$$

However, both (1) and (2) may suffer from selection bias. We address selection on observables by using two approaches: The first one allows us to address the differences in the distribution of characteristics between teachers and other professionals by modelling self-selection on observables into the regression equation: a (two-step) *heckit* model (Heckman, 1979; Maddala, 1983). The second approach: a non-parametric matching, allow us to address the differences in the distribution of observable characteristics in and out of the common support between teachers and other professionals (Ñopo, 2008). We compare teachers and other professionals on the basis of the same observable characteristics –thus, in the common support. We describe both approaches next.

3.1 Two-step model (Maddala, 1983)

To address self-selection, researchers usually resort to Heckman self-selection correction models. Heckman proved that, under incidental truncation, including the inverse of the Mills ratio would provide consistent estimators (Heckman, 1979). Later on, Maddala (1983) proved that Heckman's approach also provides consistent estimators in the presence of self-selection but the absence of incidental truncation. Therefore, the impact of selection bias is neither thrown away nor assumed to be random; we model it into the equation estimating the outcome regression.

In the first step we estimate the selection equation

$$t_{i,1} = \theta X_{i,1} + \epsilon_{i,1}, \qquad (3)$$

where $X_{i,1}$ denotes the vector of observable characteristics. Thus, we are able to obtain

$$Prob\left(t_{i,1} = 1 \middle| X_{i,1}\right) = \varphi\left(\theta X_{i,1}\right)$$

and

$$Prob\left(t_{i,1} = 0 \middle| X_{i,1}\right) = 1 - \Phi\left(\theta X_{i,1}\right),$$

and compute the inverse of the Mills ratio, i.e.

$$\lambda = \frac{\varphi(\theta X_{i,1})}{1-\Phi(\theta X_{i,1})}.$$

Given sample selection (t endogenous) we plug λ into (1) and estimate

$$y_{i,1} = \beta_1 + \beta_2 t_{i,1} + \beta_3 y_{i,0} + \beta_4\left(t_{i,1} \times y_{i,0}\right) + \rho \frac{\varphi(\theta X_{i,1})}{1-\Phi(\theta X_{i,1})} + \varepsilon_i. \qquad (4)$$

Analogously, we plug λ into (2) and estimate

$$y_{i,1} = \beta_1 + \beta_2 t_{i,1} + \beta_3 y_{i,0} + \beta_4\left(t_{i,1} \times y_{i,0}\right) + \beta_5 h_{i,1} + \beta_6\left(h_{i,1} \times y_{i,0}\right) + \rho \frac{\varphi(\theta X_{i,1})}{1-\Phi(\theta X_{i,1})} + \varepsilon_i. \quad (5)$$

3.2 Non-parametric matching (Ñopo, 2008)

There might be combinations of characteristics for which we can find a teacher, but not another professional. These differences in observables might lead to biased estimates given the differences in the supports of the distribution of observable characteristics. Ñopo (2008) propose a way to solve this problem, by comparing individuals in the common support. The matching procedure goes as follows:

- Select one teacher from the sample (without replacement).
- Select all non-teachers that have the same observable characteristics as the teacher previously selected.
- Put the observations of both individuals (the synthetic teacher and non-teacher) in their respective new samples of matched individuals, reweighting the observations.
- Repeat until exhausting the original teachers' sample.

We use gender, year of birth, year and semester in which Saber 11 was taken, parents' education, school type (public or private), type of university (public or private) and scores in Saber 11 (as deciles of the distribution of scores), as variables for the matching. Therefore, we not only compare individuals on the basis of the same observable characteristics but also on the same proxy of academic ability before college education.

By matching on observables we obtain a new distribution of observable characteristics for other professionals that mimics the one for teachers. (See Ñopo, 2008, for further methodological details.) Therefore, we proceed to estimate:

$$y_{i,1}\, w_{matching} = [\beta_1 + \beta_2 t_{i,1} + \beta_3 y_{i,0} + \beta_4 (t_{i,1} \times y_{i,0}) + \varepsilon_{i,1}] w_{matching} \qquad (6)$$

and (subsequently)

$$y_{i,1}\, w_{matching} = [\beta_1 + \beta_2 t_{i,1} + \beta_3 y_{i,0} + \beta_4 (t_{i,1} \times y_{i,0}) + \beta_5 h_{i,1} + \beta_6 (h_{i,1} \times y_{i,0}) + \varepsilon_{i,1}] w_{matching}, \quad (7)$$

where $w_{matching}$ denotes the weights after matching (that is, after the differences in the distribution of observable characteristics have vanished).

4. Results

Table 3 shows the estimates of relative learning mobility for 9 specifications; each triplet of columns corresponds to a subject: quantitative reasoning, native language and foreign language. The first column of each triplet shows the results obtained after estimating (1) through Ordinary Least Squares (OLS); the second column after estimating the two-step estimators described by equation (4),[5] and the third column shows the results obtained from estimating (6).[6]

Being enrolled in an education major is negatively related to scores in quantitative reasoning and native language in Saber Pro. Interestingly, being enrolled in an education major is positively related to performance in foreign language in Saber Pro. As we will uncover next, when estimating Equation (2) instead of Equation (1), most of this result is explained by the skill gains of those who study to become English teachers.

[5] The first-step estimates can be found in Table A1 in the appendix.

[6] Table A2 in the appendix shows the size of the common support after adding each matching variable at a time. After controlling for the full set of observable characteristics we are left with 2/3 of the teachers sample and around 15% of the other professionals sample. We compute the p-values of Kolmogorov-Smirnov tests for the two distributions: teachers and other professionals, to guarantee we are comparing individuals with the same characteristics, including academic performance. After controlling for all observable characteristics the distributions of scores in Saber 11 between teachers and other professionals are statistically the same in each subject.

It comes as no surprise that performance in Saber 11 is a good predictor of performance in Saber Pro across the board. For the performance of teachers in quantitative reasoning, however, the evidence is less favorable. The interaction between scores in Saber 11 and being a teacher shows a robust statistically significant negative value on the regressions. That is, on average, teachers' skills in quantitative reasoning relatively deteriorate compared to those of their peers in other university programs. We do not find similar evidence of a statistically significant change in teachers' relative learning in native language (Spanish) and foreign language (English).

The evidence of selection on observables is highlighted by the high significance of the inverse of the Mills ratio (λ). This also means that under OLS, on average, the conditional z-scores of a teacher are lower than those from another professional with similar observable characteristics drawn at random from the sample. That is, under OLS β_2 is biased downwards.

Table 3. Relative learning mobility

Variable	Dependent variable: Z-scores in Saber PRO								
	Quantitative reasoning			Native language (Spanish)			Foreign language (English)		
	Ordinary Least Squares	Two-step	Matching	Ordinary Least Squares	Two-step	Matching	Ordinary Least Squares	Two-step	Matching
Teacher	-0.4220***	-0.3588***	-0.3874***	-0.1644***	-0.0613***	-0.0902***	0.0452**	0.1570***	0.2104***
	(0.0151)	(0.0156)	(0.0226)	(0.0146)	(0.0150)	(0.0221)	(0.0186)	(0.0188)	(0.0243)
Saber 11 scores in									
Quantitative reasoning	0.4962***	0.4660***	0.4535***						
	(0.0043)	(0.0047)	(0.0164)						
Native language				0.5028***	0.4654***	0.5044***			
				(0.0041)	(0.0042)	(0.0149)			
Foreign language							0.6580***	0.6044***	0.4928***
							(0.0046)	(0.0051)	(0.0156)
Teacher * Saber 11 scores in									
Quantitative reasoning	-0.1457***	-0.1603***	-0.1000***						
	(0.0177)	(0.0180)	(0.0267)						
Native language				0.0319**	0.0057	0.0141			
				(0.0138)	(0.0137)	(0.0221)			
Foreign language							0.0574**	0.0327	0.1584***
							(0.0280)	(0.0279)	(0.0322)
Constant	0.0301***	0.3089***	0.0440***	0.0143***	0.4512***	-0.0113	-0.0020	0.5022***	-0.1825***
	(0.0040)	(0.0180)	(0.0144)	(0.0040)	(0.0169)	(0.0144)	(0.0034)	(0.0190)	(0.0114)
λ		-0.5421***			-0.8486***			-0.9784***	
		(0.0339)			(0.0318)			(0.0356)	
Observations	**50772**	**50772**	**9743**	**50772**	**50772**	**9605**	**50772**	**50772**	**10312**

Robust standard errors in parentheses. * Significant at ten percent; ** significant at five percent; *** significant at one percent.
Note: Programs with emphasis in mathematics are those related to the study of mathematics, physics, biology and the like; programs with emphasis in native language are those related to the study of social sciences, humanities, philology and the like; programs with emphasis in foreign language are those that are focused on the study, mainly, of a foreign language (including English).
Source: Authors' calculations based on ICFES data.

Not all teachers receive the same training while in college. Some of them specialize in teaching math, some others in teaching Spanish, some others in teaching English and so on. It is reasonable to expect different learning gains in different subjects of Saber Pro according to their fields of specialization. Table 4 shows estimations exploring such differences in relative learning mobility. That is, when analyzing relative learning gains in quantitative reasoning we will pay special attention to those education majors emphasizing on math teaching. Similarly, when analyzing relative learning gains in native language (foreign language) we will pay special attention to those education majors emphasizing on Spanish (foreign languages) teaching.

Similar to Table 3, Table 4 shows the estimates of relative learning mobility for 9 specifications: each triplet of columns corresponds to a subject; the first column of each triplet shows the results obtained after estimating (2) through OLS; the second column after estimating the two-step regression model (5), and the third column shows the results obtained from estimating (7).

As in the previous estimations, being a teacher is negatively correlated to performance in Saber Pro, but there are some differences with the results shown in Table 3. For quantitative reasoning and native language this correlation is more negative. For foreign language, this correlation, which was positive, becomes negative. Nonetheless, the results indicate that being an English teacher (as opposed to a teacher in other subjects) has a strongly positive correlation with performance in English in Saber Pro. Along similar lines, Spanish teachers also show a positive performance with respect to the average. This is not the case for teachers in programs with emphasis in quantitative reasoning. The coefficient corresponding to the $h_{i,1}$ dummy is positive and significant, but the total effect $\beta_2+\beta_5$ is still negative.

The results on the predictive power of Saber 11 across the board (β_3) and the smaller learning mobility of teachers with emphasis in quantitative reasoning (β_4) still hold. Such smaller learning mobility of teachers, nonetheless, gets counterbalanced among those teachers whose programs emphasized math. They show higher mobility than the other teachers ($\beta_6 > 0$), although their mobility is still smaller than that of other professionals ($\beta_4 + \beta_6 < 0$). The most intriguing and possibly discouraging result, however, is that learning mobility in native language (foreign language) for teachers in programs whose emphasis is in Spanish (English) is smaller than that of other teachers ($\beta_6 < 0$). That is, the more able an education major student showed to be in native language (foreign language) in Saber 11, the smaller her/his relative progress in native language (foreign language) skill acquisition, as measured by Saber Pro.

We explored this result further by estimating Equation (2) using quantile regressions. Figure 3 shows the coefficients of β_6 by percentiles of the residuals of the regression. In the case of foreign language, the results clearly show that this effect is more pronounced at the higher percentiles of the distribution of residuals of the regression. That is, it is more pronounced among the good performers in Saber Pro. Interestingly, this is not the case among teaching majors that emphasize in neither math nor native language (Spanish). This reduces the scope for an explanation based on a regression to the mean argument. A plausible interpretation for this result, on the other hand, could come from the existence of (negative) peer effects. The few able and talented students who choose a career in English education see their skills relatively reduced as a result of such choice.

Table 4. Relative learning mobility controlling for career emphasis for teachers

Variable	Dependent variable: Z-scores in Saber PRO								
	Quantitative reasoning			Native language (Spanish)			Foreign language (English)		
	Ordinary Least Squares	Two-step	Matching	Ordinary Least Squares	Two-step	Matching	Ordinary Least Squares	Two-step	Matching
Teacher	-0.5154***	-0.4572***	-0.4712***	-0.2273***	-0.1275***	-0.1456***	-0.1683***	-0.0549***	0.0147
	(0.0163)	(0.0166)	(0.0239)	(0.0167)	(0.0170)	(0.0240)	(0.0189)	(0.0189)	(0.0236)
Saber 11 scores in									
Quantitative reasoning	0.4962***	0.4647***	0.4535***						
	(0.0043)	(0.0047)	(0.0164)						
Native language				0.5028***	0.4650***	0.5044***			
				(0.0041)	(0.0042)	(0.0149)			
Foreign language							0.6580***	0.6031***	0.4928***
							(0.0046)	(0.0050)	(0.0156)
Teacher * Saber 11 scores in									
Quantitative reasoning	-0.2003***	-0.2122***	-0.1401***						
	(0.0191)	(0.0192)	(0.0282)						
Native language				0.0409***	0.0186	0.0369			
				(0.0157)	(0.0155)	(0.0239)			
Foreign language							-0.0882***	-0.1100***	0.0274
							(0.0315)	(0.0315)	(0.0323)
Teacher program has emphasis in the assessed subject in Saber PRO	0.3874***	0.4259***	0.3454***	0.2418***	0.2608***	0.2105***	1.6211***	1.6544***	1.5936***
	(0.0359)	(0.0351)	(0.0424)	(0.0316)	(0.0311)	(0.0378)	(0.0506)	(0.0489)	(0.0646)

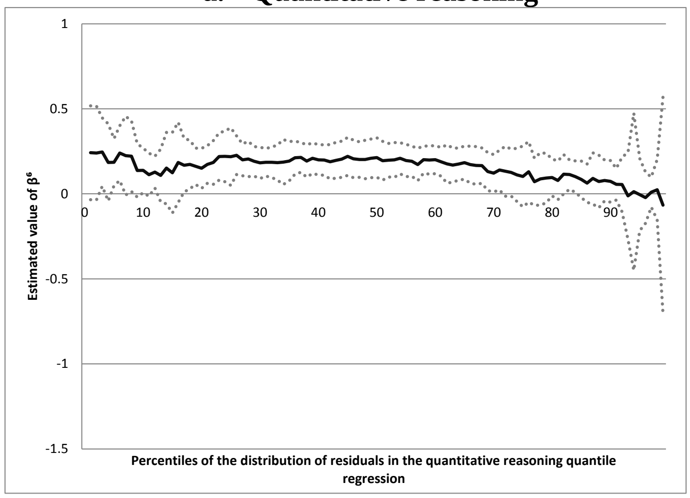
(***Continues on next page***)

Variable	Dependent variable: Z-scores in Saber PRO								
	Quantitative reasoning			Native language (Spanish)			Foreign language (English)		
	Ordinary Least Squares	Two-step	Matching	Ordinary Least Squares	Two-step	Matching	Ordinary Least Squares	Two-step	Matching
Teacher program has emphasis in the assessed subject in Saber PRO * Saber 11 scores in									
Quantitative reasoning	0.1454***	0.1249***	0.0906*						
	(0.0420)	(0.0437)	(0.0507)						
Native language				-0.0681**	-0.0871***	-0.1095***			
				(0.0304)	(0.0304)	(0.0376)			
Foreign language							-0.1496***	-0.1858***	-0.1678***
							(0.0559)	(0.0539)	(0.0611)
Constant	0.0301***	0.3209***	0.0440***	0.0143***	0.4560***	-0.0113	-0.0020	0.5150***	-0.1825***
	(0.0040)	(0.0180)	(0.0144)	(0.0040)	(0.0167)	(0.0144)	(0.0034)	(0.0178)	(0.0114)
λ		-0.5654***			-0.8579***			-1.0032***	
		(0.0340)			(0.0313)			(0.0333)	
Observations	**50772**	**50772**	**9743**	**50772**	**50772**	**9605**	**50772**	**50772**	**10312**

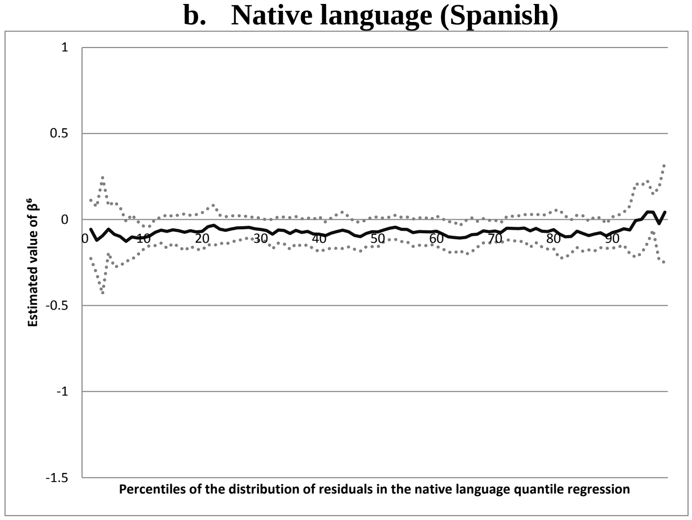
Robust standard errors in parentheses. * Significant at ten percent; ** significant at five percent; *** significant at one percent.
Note: Programs with emphasis in mathematics are those related to the study of mathematics, physics, biology and the like; programs with emphasis in native language are those related to the study of social sciences, humanities, philology and the like; programs with emphasis in foreign language are those that are focused on the study, mainly, of a foreign language (including English).
Source: Authors’ calculations based on ICFES data

Figure 3. Relative learning mobility of teacher majors.
Values of $\widehat{\beta}_6$ from Equation (2) obtained from quantile regressions

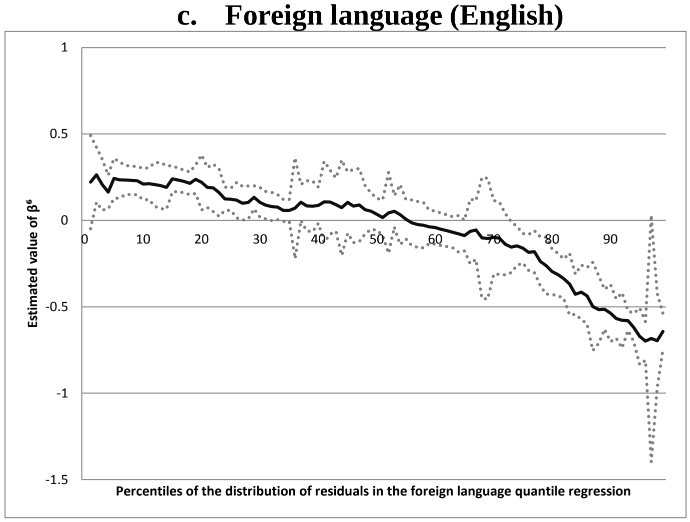
a. Quantitative reasoning



b. Native language (Spanish)



c. Foreign language (English)



Source: Authors' calculations based on ICFES data.
Note: Robust confidence intervals at 95% added.

5. Conclusions

It is unquestionable that teachers are essential for a high quality educational system. Nonetheless, there is evidence of negative selection into teaching as students from disadvantaged socioeconomic environments with low academic performance are more likely to enroll into programs in school education. This study builds onto that providing another piece of evidence. The skills of students in education majors deteriorate in comparison to those who enroll in others. This is specially the case for: (i) those who enroll in education majors with emphasis in math (physics, mathematics, biology, and, in general, the sciences) and (ii) those who show better skills in English before entering college and enroll in education majors with emphasis in English. After nearly 5 years of academic training, learning gaps between teachers and other professionals widen in favor of the latter. The results hold even after accounting for differences in observable characteristics and academic performance before college. This raises an additional red flag regarding the quality of education for teachers.

What can be done? Perhaps the most obvious place to look at is precisely the programs that nurture the future teachers. It could be the case that the teaching bodies of the future teachers, or the curricula they follow, or their pedagogical approaches, need some reforms. There is also room for action on the selection of students into teaching, such as stricter admission standards. This may work positively by two channels: a direct effect on the skills of students in teaching majors and an indirect effect through peers. Nonetheless, better teacher education programs and higher admissions standards, alone, most likely will have only modest effects. Or, more generally, to think that the solution to the problem of an inadequate teaching force lies only within the teaching community would be extremely myopic. It is necessary to push for ambitious policies aimed at making the teaching profession more attractive so that the most talented youngsters opt to teach and develop a good career path.

The recent decades have seen innovation and progress in teacher policies both in the developed and developing world. Many of them have been fortunately accompanied by rigorous impact evaluations. Thus, although perhaps we are still away from answering the question of the design of optimal policies for the teaching profession in different contexts, knowledge is in the making (Vegas and Ganimian, 2013).

References

Ballou, D., & Podgursky, M. (1997). “Teacher Pay and Teacher Quality,” W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.

Barón, J. D., & Bonilla, L. (2011). “La calidad de los maestros en Colombia: Desempeño en el examen de Estado ICFES y la probabilidad de graduarse en el área de educación,” Documentos de Trabajo sobre Economía Regional 152, Banco de la República, Colombia.

Barón, J. D., Bonilla, L., Cardona-Sosa, L., & Ospina, M. (2013). “Quénes eligen la disciplina de la educación en Colombia? Caracterización desde las pruebas Saber 11,” Borradores de Economía 785, Banco de la República, Colombia.

Chetty, R., Friedman, J., & Rockoff, J. (2013a). “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates,” NBER Working Papers 19423, National Bureau of Economic Research, Inc.

Chetty, R., Friedman, J., & Rockoff, J. (2013b). “Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood,” NBER Working Papers 19424, National Bureau of Economic Research, Inc.

Clotfelter, T., Ladd, H. F., & Vigdor, J. L. (2007). “Teacher credentials and student achievement: Longitudinal analysis with student fixed effects,” Economics of Education Review 26(6): 673-682.

Cunha, J., & Miller., T. (2014). “Measuring value-added in higher education: Possibilities and limitations in the use of administrative data,” Economics of Education Review 42: 64–77

Denzler, S., & Wolter. S. (2008). “Selbstselektion bei der Wahl eines Lehramtsstudiums: Zum Zusammenspiel individueller und institutioneller Faktoren,” Beiträge zur Hochschulforschung, 30(4): 112-141.

De Paola, M. (2009). “Does Teacher Quality Affect Student Performance? Evidence From An Italian University,” Bulletin of Economic Research, 61(4): 353–377.

Dolton, P. (1990). “The economics of UK teacher supply,” Economic Journal, 100: 91–104.

Dolton, P. (2006). “Teacher supply,” chapter 19 in E. Hanushek and F. Welch (eds.), Handbook of the Economics of Education vol. 2, Elsevier, The Netherlands.

Dolton, P., & Marcenaro-Gutierrez, O. (2011). “If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance,” Economic Policiy, 26(65): 5-55.

Dolton, P., & van der Klaauw, W. (1999). “The turnover of teachers: A competing risks explanation,” Review of Economics and Statistics, 81(3): 543–550.

Eide, E., Goldhaber, D., & Brewer, D. (2004). “The Teacher Labour Market and Teacher Quality,” Oxford Review of Economic Policy, 20(2): 230–244.

García, S., Maldonado, D., Perry, G., Rodríguez, C., & Saavedra., J. (2014). “Tras la excelencia docente: Cómo mejorar la calidad de la educación para todos los colombianos.” Fundación Compartir. Bogotá, Colombia.

Giesen, H., & Gold, A. (1993). “Leistungsvoraussetzungen und Studienbedingungen bei Studierenden verschiedener Lehrämter,” Psychologie in Erziehung und Unterricht, 40: 111–124

Goldhaber, D., & Brewer, D. (1997). “Why don’t schools and teachers seem to matter?: Assessing the impact of unobservables on educational productivity”, Journal of Human Resources 32(3): 505–523.

Hanushek, E. A., & Pace, R. (1995). “Who Chooses To Teach And Why?” Economics of Education Review, 14(2): 101–117

Hanushek, E. A. (2002). “Publicly provided education,” chapter 30 in A. J. Auerbach and M. Feldstein (eds.), Handbook of Public Economics, vol. 4, Elsevier, The Netherlands.

Hanushek, E. A., & Rivkin, S. G. (2006). “Teacher Quality,” chapter 18 in E. Hanushek and F. Welch (eds.), Handbook of the Economics of Education, vol. 2, Elsevier, The Netherlands.

Klein, S. P., Kuh, G., Chun, M., Hamilton, L., & Shavelson, R. (2005). “An approach to measuring cognitive outcomes across higher education institutions,” Research in Higher Education, 46: 251–276.

Kukla-Acevedo, S. (2009). “Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement,” Economics of Education Review 28(1): 49-57.

Liu, O. (2011). “Value-added assessment in higher education: A comparison of two methods,” Higher Education, 61(4): 445–461.

Maddala, G. S. (1983). “Limited-Dependent and Qualitative Variables in Econometrics,” Cambridge University Press, New York.

Mizala, A., & Ñopo, H. (2011). “Teachers’ Salaries in Latin America. How Much are They (under or over) Paid?,” IZA Discussion Papers 5947, Institute for the Study of Labor (IZA).

Murnane, R., Singer, J., Willett, J., Kemple, J., & Olson, R. (1991). “Who Will Teach? Policies that Matter,” Harvard University Press, Cambridge, MA.

Ñopo, H. (2008). “Matching as a Tool to Decompose Wage Gaps,” Review of Economics and Statistics, 90(2): 290-299.

Odden, A., & Kelley, C. (1997). “Paying Teachers for What They Know and Do: New and Smarter Compensation Strategies to Improve Schools,” Corwin Press, Thousand Oaks, CA.

OECD (2011). “Finland: Slow and steady reform for consistently high results” in Strong Performers and Succesful Reformes in Education, Lessons from PISA for the United States, OECD Publishing.

Podgursky, M., Monroe, R., & Watson, D. (2004). “The academic quality of public school teachers: an analysis of entry and exit behavior,” Economics of Education Review, 23(5): 507–518

Rivkin, S., Hanushek, E., & Kain, J. (2005). “Teachers, Schools, and Academic Achievement,” Econometrica, 73(2): 417-458.

Rockoff, J. E. (2004). "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," The American Economic Review, 94(2): 247–252.

Saavedra, J. (2009). "The learning and early labor market effects of college quality: a regression discontinuity analysis," Mimeo, Harvard University.

Saavedra, A., & Saavedra, J. (2011). "Do colleges cultivate critical thinking, problem solving, writing and interpersonal skills?" Economics of Education Review, 30(6): 1516–1526.

Sahlberg, P. (2009). "Educational Change in Finland" in A. Hargreaves, M. Fullan, A. Lieberman, and D. Hopkins (Eds.), International Handbook of Educational Change, Springer, The Netherlands.

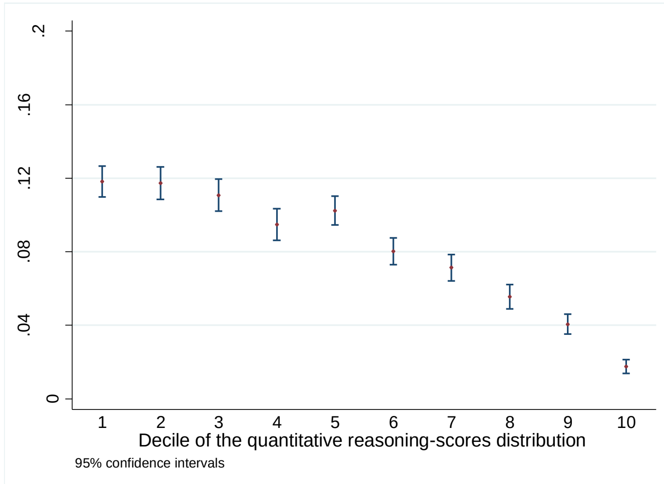
Sahlberg, P. (2011). "Finnish lessons. What can the world learn from educational change in Finland?" Teachers College Press, New York.

Vegas, E., & Ganimian, A. (2013). "Theory and Evidence on Teacher Policies in Developed and Developing Countries," IDB Working Paper Series No. IDB-WP-438.

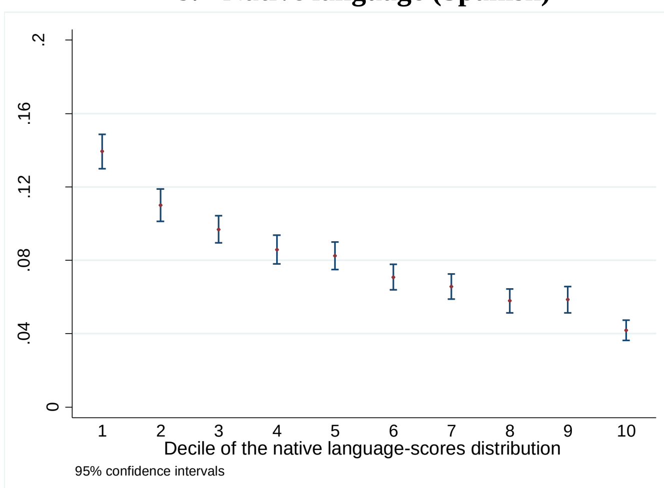
Appendix A

Figure A1. Probability of being a teacher by deciles of the z-score distribution

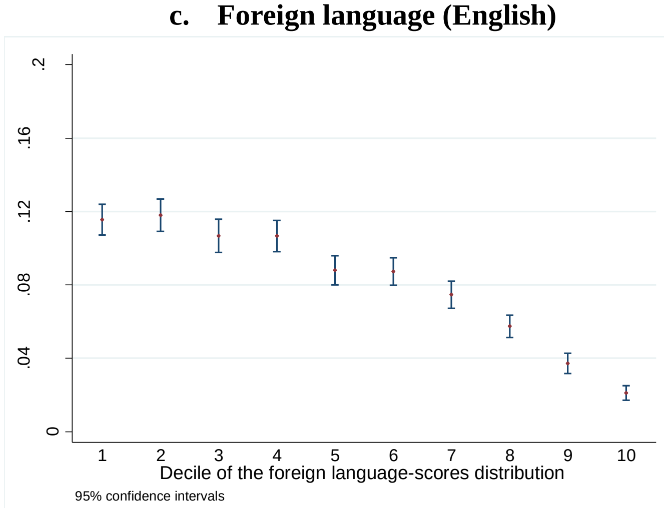
a. Quantitative reasoning



b. Native language (Spanish)



c. Foreign language (English)



Source: Authors' calculations based on ICFES data.

Table A1. Likelihood of choosing an education major (marginal effects after probit)

	Dependent variable: 1 if student studies a program in education, 0 if not		
	Quantitative reasoning	**Native language**	**Foreign language**
Socio-demographic			
Gender (Female)	0.0153***	0.0188***	0.0187***
	(0.0021)	(0.0021)	(0.0021)
Age	0.0026***	0.0027***	0.0027***
	(0.0005)	(0.0005)	(0.0005)
The student moved to another administrative unit†	-0.0335***	-0.0339***	-0.0338***
	(0.0028)	(0.0029)	(0.0028)
Familiy size (more than 5 persons)	0.0180***	0.0181***	0.0178***
	(0.0030)	(0.0030)	(0.0030)
Max education of the parents/guardians			
Secondary complete or tertiary incomplete	-0.0205***	-0.0202***	-0.0203***
	(0.0022)	(0.0023)	(0.0022)
Technical or technician education complete	-0.0303***	-0.0302***	-0.0304***
	(0.0024)	(0.0024)	(0.0024)
Universitary education complete	-0.0615***	-0.0618***	-0.0609***
	(0.0025)	(0.0025)	(0.0025)
School characteristics			
School administration (public)	0.0451***	0.0460***	0.0443***
	(0.0024)	(0.0025)	(0.0025)
School type (mixed gender)	0.0188***	0.0187***	0.0182***
	(0.0025)	(0.0026)	(0.0026)
School calendar (A calendar)††	-0.0119	-0.0113	-0.0128
	(0.0079)	(0.0080)	(0.0081)
School type			
Technical	-0.0026	-0.0028	-0.0037
	(0.0025)	(0.0025)	(0.0025)
Superior Normal school	0.1081***	0.1116***	0.1079***
	(0.0093)	(0.0095)	(0.0093)
Saber 11 scores in			
Quantitative reasoning	-0.0173***		
	(0.0012)		
Native language		-0.0118***	
		(0.0011)	
Foreign language			-0.0133***
			(0.0014)
Pseudo R2	0.13	0.12	0.12
Observations	**50772**	**50772**	**50772**

Standard errors in parentheses. * Significant at ten percent; ** significant at five percent; *** significant at one percent. We also control for dummies that capture when Saber 11 was taken and dummies per administrative unit where the university is located.
† The student moved to another administrative unit for his/her higher education.
†† Colombia has two regular school calendars: The “A calendar”, which goes from February to November; and the “B calendar”, which goes from August to June.
Source: Authors’ calculations based on ICFES data.

Table A2. Size of common support and Kolmogorov-Smirnov tests (p-values after adding observables characteristics)

Variables	Percentage of students		Kolmogorov-Smirnov P-value
	Teachers	Other professionals	
Gender	100	100	0.00
+ year of birth	99.90	99.82	0.00
+ semester and year in which Saber 11 was taken	99.25	98.28	0.00
+ parents' max education	98.58	94.09	0.00
+ the student moved to another administrative unit	97.98	90.99	0.00
+ type of school	96.80	82.84	0.00
+ type of university	92.95	58.64	0.00
& Quantitative reasoning scores deciles	64.89	15.12	0.99
& Native language scores deciles	64.09	14.89	0.49
& Foreing language scores deciles	66.45	16.20	0.15

The Kolmogorov-Smirnov test corresponds to the test of equality of two distributions.
Source: Authors' calculations based on ICFES data.