



# Do tests applied to teachers predict their effectiveness?<sup>☆</sup>

Yyannú Cruz-Aguayo, Pablo Ibararán, Norbert Schady\*

Inter-American Development Bank, 1300 New York Avenue N.W., Washington, DC, 20577, United States



## HIGHLIGHTS

- We analyze a teacher evaluation system in Ecuador.
- The evaluation includes a written test, and a demonstration class.
- It also rewards teachers for experience, degrees, and in-service training.
- Children taught by teachers with higher scores on the evaluation do not learn more.
- Our results underline the difficulty policy-makers have identifying effective teachers.

## ARTICLE INFO

### Article history:

Received 8 May 2017

Received in revised form 20 June 2017

Accepted 23 June 2017

Available online 1 July 2017

### JEL classification:

I25

I28

### Keywords:

Teachers

Evaluation

Test scores

## ABSTRACT

We analyze a teacher evaluation system in Ecuador. The evaluation includes a written test, a demonstration class, and points for experience, degrees, and in-service training. Children taught by teachers with higher scores on the evaluation do not learn more.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

It is generally believed that teachers are the most important input into the production of learning within schools. Having a better teacher has long-term consequences for a variety of outcomes in adulthood, including college attendance, savings, and wages (Chetty et al., 2011, 2014a). Research from the United States has credibly established that teachers – even teachers in the same school, teaching observationally equivalent students – vary a great deal in their effectiveness (Chetty et al., 2014b; Hanushek and Rivkin, 2012). However, knowing that some teachers produce more learning than others provides no guidance on the attributes of effective teachers.

In this note, we analyze whether an evaluation that is used to determine which teachers receive tenure predicts student learning

<sup>☆</sup> We thank officials in the Ministry of Education in Ecuador for making data available to us, Fiorella Benedetti and Jorge Luis Castañeda for outstanding research assistance, and Pedro Carneiro, Costas Meghir and an anonymous referee for their comments.

\* Corresponding author.

E-mail address: [Norberts@iadb.org](mailto:Norberts@iadb.org) (N. Schady).

in Ecuador, a middle-income country in South America. In earlier work (Araujo et al., 2016), we showed that kindergarten teachers varied in the impact they have on learning: A one standard deviation increase in classroom quality increased test scores in math and language by 0.11 standard deviations.

In Ecuador, about one in three teachers in the public sector work on a contract basis. Tenured teachers are paid substantially more than contract teachers, and have more generous health insurance and pension benefits. Being converted from a contract to a tenured teacher therefore has important benefits for the individual, and implies increases in costs for the Ministry of Education.

Every year the Ministry carries out a planning exercise that takes account of the current distribution of tenured teachers, student population and the projected growth in demand, and any projected school openings, closings or merges. Based on this exercise, additional tenured slots are assigned to some schools. These slots are school-specific, but any individual who has a teaching degree from an accredited institution can apply. Among eligible applicants to a new slot, the applicant with the highest total score on an evaluation known as the *Concurso de Méritos y Oposiciones* (henceforth, *Concurso*) is awarded the tenured position.

The *Concurso* has three components, each of which receives equal weight in an aggregate score. The first component is given by a teacher's score on a test. This test has three parts: logical-verbal reasoning, pedagogical knowledge, and subject-specific knowledge. The second component of the aggregate score is a teacher's score on a demonstration class. The third component of the aggregate *Concurso* score is given by a point system (*Méritos*) which gives higher scores to teachers with more experience, degrees (above and beyond the basic teaching certificate), and in-service training.

We use data on a sample of children in 2nd, 3rd, and 4th grades to analyze whether children taught by teachers with higher scores on the *Concurso* have higher achievement in language and math. Our earlier work (Araujo et al., 2016) was based on a sample of kindergarten children randomly assigned to teachers within schools. The data used for the analysis in this paper do not come from an experiment. We therefore have to make stronger identifying assumptions. In practice, we use a school fixed effects strategy that compares the test scores of observationally equivalent children taught by teachers with higher or lower *Concurso* scores in the same school.

## 2. Data and identification

Our sample consists of 4,479 children taught by 480 teachers in 240 schools. Schools were selected so each would have at least two contract teachers in 2nd through 4th grades who had taken the *Concurso* evaluation. Within each classroom, 9–10 children were randomly selected for testing. Twenty-eight percent of the children in the sample are in 2nd grade, 38 percent in 3rd grade, and 35 percent in 4th grade. To ease with interpretation, we standardize teacher scores on the *Concurso* and its components to have mean zero and unit standard deviation.

Children were tested twice, midway through the school year, and shortly before the end. We used adapted versions of the Early Grade Reading Assessment (EGRA; RTI International 2009; Gove and Cvelich (2011)), and Early Grade Math Assessment (EGMA; Reubens (2009)). Both tests have been applied in more than 20 developing countries. Application takes about 30 min per child. We standardize the four tests (EGRA and EGMA, mid- and end-year) so they have mean zero and unit standard deviation in each grade.

For 3,346 children in the sample (74.7 percent), we also have data from a household survey. This survey includes information on the years of completed schooling of mothers and fathers; whether a child attended preschool; and information on a large number of household characteristics (such as the material of the floors, wall, and ceilings in the house; whether the house is connected to the electricity, water, and sewerage system), and ownership of assets (questions for 23 assets). We aggregate the information on household characteristics and assets into a wealth composite (following, among many others, Araujo et al. (2016), Paxson and Schady (2007) and Paxson and Schady (2010), who also work with data from Ecuador).

The regressions we estimate take the following form:

$$Y_{itgs} = \theta_g + \alpha_s + \beta_1 T_{gs} + \beta_2 X_{itgs} + \varepsilon_{itgs}, \quad Z = 1, 2, \dots, N \quad (1)$$

where the subscripts *i*, *t*, *g*, and *s* refer to individuals, teachers, grades, and schools, respectively;  $\theta_g$  is a set of grade fixed effects;  $\alpha_s$  is a set of school fixed effects;  $T_{gs}$  is the score of a teacher on the *Concurso* or its components;  $X_{itgs}$  always includes controls for child age in months, gender, and classroom size and, in some specifications, the education of mothers and fathers, household wealth, and an indicator variable for whether a child attended preschool;  $\varepsilon_{itgs}$  is the regression error term; and the regressions are run for *z* different tests (the mid- and end-year scores on

EGRA or EGMA, and the change in scores between mid- and end-year). The coefficient of interest is  $\beta_1$ , an estimate of the extent to which children taught by teachers with higher scores on the *Concurso* have better scores in math or language. Standard errors are corrected for clustering at the school level.

The main estimation concern is purposeful sorting of children to teachers. As a check on our identification strategy, we run regressions of the variables in  $T_{gs}$  on the variables in  $X_{itgs}$ . Since the variables in  $X_{itgs}$  are predetermined, any indication that teachers with higher (or lower) test scores on the *Concurso* were assigned to children with different characteristics would raise questions about our identification strategy. Table 1 finds no evidence that this is the case, regardless whether the variables in  $X_{itgs}$  are entered one at a time or jointly. We conclude that our identification strategy is reasonable.

## 3. Results

To put things in context, we first regress the change in math or language test scores between mid- and end-year on the full set of child and household controls. We take the residuals from these regressions, calculate classroom means, and the difference in means between classrooms in the same school. Within-school differences are large: The median difference is 0.18 standard deviations for language, and 0.22 standard deviations for math. Although some of these within-school, cross-classroom differences are probably driven by sampling error or idiosyncratic classroom shocks, it seems likely that a substantial proportion reflect true differences in the quality of teachers.<sup>1</sup>

Next, we show that the characteristics in  $X_{itgs}$  tend to be strongly correlated with test scores, as expected. Table 2 shows that girls score about 0.14 standard deviations higher on language, and 0.23 standard deviations lower in math.<sup>2</sup> Children from households of higher socioeconomic status have higher test scores in both reading and math.

Our main findings are in Table 3. We report the results from specifications with the additional controls (even-numbered columns) and without (odd-numbered columns), and six different dependent variables: the mid- and end-year test scores in language and math, and the change in test scores between mid- and end-year.<sup>3</sup>

Table 3 shows no evidence that the test score or any of its components, the score on the demonstration class, points on the *Méritos* scale, or the aggregate score on the *Concurso* predict child achievement in language or math. There are 84 coefficients in the table, and about one-half (40) are positive and one-half (44) are negative. No coefficient is significant at conventional levels. The estimates in the table are quite precise. For example, in the specifications with the additional controls, we can rule out positive associations between the total *Concurso* score and child test scores larger than 0.02 standard deviations for language, and 0.03 standard deviations for math. Consistent with the results in Table 1, the coefficients with and without the additional controls are very similar.

<sup>1</sup> In Araujo et al. (2016) and Carneiro et al. (2017) we use an Empirical Bayes estimator to shrink the estimated classroom effects. On average, the corrected estimates are about one-third smaller than those that do not correct for sampling error. We do not do this in the current paper because we are not calculating classroom (or teacher) effects.

<sup>2</sup> Carneiro et al. (2017) carefully analyze the gender gap in math achievement in kindergarten through 2nd grade in Ecuador. Consistent with what we report here, they find a gender gap of 0.17 standard deviations, favoring boys, by 2nd grade.

<sup>3</sup> Results are very similar if instead we use the largest possible sample of children and do not include the additional controls. These results are available from the authors upon request.

Download English Version:

<https://daneshyari.com/en/article/5057576>

Download Persian Version:

<https://daneshyari.com/article/5057576>

[Daneshyari.com](https://daneshyari.com)