I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13830

Does Test-Based Teacher Recruitment Work in the Developing World? Experimental Evidence from Ecuador

María Daniela Araujo P.
Guido Heineck
Yyannú Cruz-Aguayo

OCTOBER 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13830

Does Test-Based Teacher Recruitment Work in the Developing World? Experimental Evidence from Ecuador

María Daniela Araujo P.
University of Bamberg and Bamberg Graduate School of Social Science

Guido Heineck
University of Bamberg, IZA and LIfBi

Yyannú Cruz-Aguayo
Inter-American Development Bank

OCTOBER 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Does Test-Based Teacher Recruitment Work in the Developing World? Experimental Evidence from Ecuador*

Since 2007, the Ecuadorian government has required teacher candidates to pass national skill and content knowledge tests before they are allowed to participate in merit-based selection competitions for tenured positions at public schools in an attempt to raise teacher quality. We evaluate the impact of this policy using linked administrative teacher information to data from a unique experimental study where almost 15,000 kindergarten children were randomly assigned to their teachers in the 2012-2013 school year in Ecuador. We find positive and significant effects of test-screened tenured teachers of at least a 0.105 standard deviation for language and a 0.085 standard deviation for math, which persist even after controlling for teacher education, experience, cognitive ability, personality traits and classroom practices.

JEL Classification: I20, I21, I25, I28, J45

Keywords: teacher quality, education policy evaluation, Latin America

Corresponding author:
María Daniela Araujo P.
University of Bamberg
Feldkirchenstr. 21
96052 Bamberg
Germany

E-mail: daniela.araujo@uni-bamberg.de

* We would like to thank the Inter-American Development Bank and Ecuador's Ministry of Education for their crucial collaboration and for providing the data for this study. We also thank Maria Caridad Araujo, Pedro Carneiro, Eric A. Hanushek, David Post and seminar participants at the IAB Workshop on Field Experiments in Policy Evaluation, the 5th IZA Workshop: The Economics of Education, the ifo-WZB Workshop on Causal Analyses of School Reforms, and the Universities of Bamberg, Halle, Jena and Leipzig for helpful comments and discussions. This work was supported by the Bamberg Graduate School of Social Sciences, which is funded by the German Research Foundation (DFG) under the German Excellence Initiative (GSC1024).

1. INTRODUCTION

Identifying high-quality teachers who substantially contribute to student learning has been one of the main challenges faced by policy-makers and researchers in education in recent decades. Much current research has shown that teachers are a key factor in student learning, although what makes a good teacher remains a puzzle. Value-added to student achievement models derived from the education production function literature have not only found substantial teacher effects or individual teacher contribution to student achievement, but also substantial variation in this contribution (Chetty, Friedman, and Rockoff 2014; Hanushek and Rivkin 2006; 2010; 2012; Jackson, Rockoff, and Staiger 2014; Koedel, Mihaly, and Rockoff 2015). Interestingly, it has also been observed that easily quantifiable teacher characteristics such as academic degree, experience beyond the first years, training or test scores explain little of the individual teacher contribution to learning.

Teacher cognitive skill[1] and content knowledge of the subject taught measured by certification tests are among the widely-used observable characteristics as a signal of teacher quality for recruitment purposes in high-income economies. In the US, written tests have been used to certify teachers since the early-20th century, as a means to ensure high and uniform academic standards for teacher pre-service programs and safeguard the public from faulty teacher candidates (D'Agostino and Powers 2009). Nonetheless, the evidence regarding the effectiveness of teacher test scores as predictors of future quality remains mixed. On the one hand, early education production function studies typically found positive effects of teacher skill and certification test scores on student achievement (Wayne and Youngs 2003). On the other hand, more recent studies using large longitudinal data and applying parametric and non-parametric value-added to student achievement models have not been able to consistently find positive or significant effects (Angrist and Guryan 2008; Boyd et al. 2008; Clotfelter, Ladd, and Vigdor 2007; Goldhaber 2007; Goldhaber and Anthony 2007; Goldhaber, Gratz, and Theobald 2017; Harris and Sass 2007; 2011; Kane, Rockoff, and Staiger 2008; Rockoff et al. 2011). On top of the mixed evidence, all such studies suffer from potential estimation bias caused by the non-random matching of teachers to students (Clotfelter, Ladd, and Vigdor 2006; Rothstein 2010; 2009; Koedel and Betts 2011).

In contrast to the US, Latin American countries have not traditionally had certification processes for the teaching profession to ensure that quality standards are met. However, in the last fifteen years several countries have implemented teacher recruitment policies based on

[1] For example, approximated by teachers' verbal and mathematical skills.

skill and subject knowledge tests to improve teacher and school quality. At present, Colombia (since 2002), Ecuador (2007), Mexico (2008) and Peru (2012) all require teacher candidates to pass national mandatory tests before they can opt for long-term careers at public schools (Elacqua et al. 2017; Bruns and Luque 2015). However, much remains to be learned about the effects of these teacher recruitment policies, since the vast majority have not been subject to evaluation.

In this article, we evaluate whether teachers who passed national entry tests and were tenured by Ecuador's new recruitment policy have positive effects on student learning outcomes by linking unique administrative teacher records to the rich experimental data produced by the "Closing Gaps" project. In the project, a representative cohort of Ecuadorian kindergarten children were assigned to their teachers in the 2012-2013 school year, using a rule that is as good as random (Araujo et al. 2016). We confirm that successful random assignment of these kindergarten children to teachers tenured through the new recruitment policy is also given in our data, whereby potential bias caused by matching these teachers to students is no concern. Our results show that kindergarten students benefit from randomly-assigned teachers who passed mandatory entry tests and won merit-based competitions for tenure in Ecuador, having significantly higher end-of-year test scores of at least a 0.105 standard deviation in language and a 0.085 standard deviation in math.

In addition to the experimental nature of the data, the "Closing Gaps" project collected rich information on students, families and teachers, which allows us to check the robustness of our experimental estimations. Moreover, we not only have access to common observable teacher characteristics as such as gender, education and experience, but also measurements of teacher cognitive ability (Wechsler Adult Intelligence Scale, WAIS-III), personality traits (the Big Five personality test) and classroom practices (the Classroom Assessment Scoring System, CLASS). The positive and significant effects of teachers tenured by the new recruitment policy persist even after controlling for the full set of teacher characteristics.

Our estimations are highly robust to several specifications. We run our original analysis for the whole representative sample of public schools that were part of "Closing Gaps" project using school fixed effects and cluster standard errors at the school level. We also test different control group specifications for our model. In addition, we run our analysis for the subsample of schools that had at least one teacher tenured by the new recruitment policy. The size and significance of our estimates are consistent for all specifications.

This paper makes three key contributions to teacher quality research in the context of the education production function literature. First, we provide the first experimental estimations of the effects of teachers screened by skill and subject knowledge tests and tenured by merit-based competitions in a Latin American country. Accordingly, we thus provide an insight into the effectiveness of the new teacher recruitment policies implemented in the region in recent years.

Second, we contribute to the current debate around the teacher characteristics associated with student learning in the developing world. We show that teachers screened by entry tests who competed to earn a permanent job position have positive and significant effects on student learning. This outcome is to some extent aligned with previous findings on the positive learning effects associated with teacher skill and subject knowledge in developing countries (Bau and Das 2020; Bietenbeck, Piopiunik, and Wiederhold 2018; Glewwe et al. 2014; Metzler and Woessmann 2012). However, our results also provide new evidence of the connection between teacher job status and performance. Contrary to recent evidence on the positive effects of fixed-term contract teachers in developing countries (Duflo, Dupas, and Kremer 2015; Muralidharan and Sundararaman 2013), our estimations show that tenured teachers significantly outperform contract teachers in Ecuador.

Third, our study also confirms the potential effectiveness of highly-qualified teachers in closing learning gaps between socioeconomically advantaged and disadvantaged children in the developing world. We find that the effects of *test-screened tenured* teachers on language learning are stronger for children who started the school year with lower baseline test scores or came from socioeconomically disadvantaged households.

The remainder of the paper is structured as follow. In Section 2, we provide background on the Ecuadorian education system, the new teacher recruitment policy and previous evidence connected to our research question. Section 3 reviews the "Closing Gaps" data experimental design and assesses the validity of the experiment. Section 4 presents our model, estimation strategy and results. Section 5 details our heterogeneous effects analysis. Section 6 presents our robustness check, and finally Section 7 concludes.

2. BACKGROUND AND EVIDENCE

2.1. Ecuadorian Education System

Ecuador is a South American middle-income country with compulsory schooling from 5 to 17 years of age. The Ecuadorian education system is organized at three levels: initial education, general basic education and high school (Asamblea Nacional del Ecuador 2011). The initial

education or early education serves children under 5 years of age (equivalent to the ISCED level 0). The general basic education starts at 5 years of age and comprises one year of kindergarten, six years of primary education (ISCED level 1) and three years of lower secondary education (ISCED level 2). Finally, high school corresponds to three years of upper secondary education (ISCED level 3).

The Ecuadorian government implemented major education reforms between 2006 and 2017 to increase enrollment and improve learning outcomes (Schneider, Cevallos Estarellas, and Bruns 2019; Araujo P. and Bramwell 2015). There is evidence of the success achieved in education expansion: while kindergarten and primary school net enrollment remained above 90 percent, secondary school net enrollment increased from 62 percent in 2005 to 85 percent in 2017 (World Bank 2019). Nonetheless, questions about the reforms' impact on the system quality remain open. These reforms particularly targeted the teaching profession, whose prestige and quality had progressively declined since the 1970s due to the lack of academic standards for teacher pre-service programs and the drastic decrease in teaching wages (Elacqua et al. 2017).

2.2. New Teacher Recruitment Policy in Ecuador

In November 2007, the Ecuadorian government issued the Executive Order No. 708[2], which required teacher candidates to pass mandatory skill and content knowledge tests before they were allowed to participate in merit-based selection competitions for tenure at public schools. The new regulation was exclusively applied to teachers seeking tenure[3] from December 2007 onwards. Already-tenured teachers were exempted from the regulation[4]. The policy reform was institutionalized in 2011, when Ecuador's new Intercultural Education Law (*Ley Orgánica de Educación Intercultural, LOEI*) ratified national entrance exams as a mandatory requirement for teacher candidates (Asamblea Nacional del Ecuador 2011).

Prior to December 2007, teacher selection processes were locally organized by Provincial Directorates of Education without national standards other than academic degree requirements. After December 2007, teacher recruitment processes were centrally organized by Ecuador's Ministry of Education. Permanent teacher vacancies had to be filled by teachers who had

[2] The Executive Order No. 708 reformed the Regulation to Ecuador's Education Law of 1990 (*Reglamento de Ley de Carrera Docente y Escalafón del Magisterio Nacional*). Even though Ecuador's Education Law of 1990 already established that teachers should be tested and recruited through merit-based competitions, national entrance examinations for teacher candidates were not implemented until the release of the Executive Order No. 708 in 2007.

[3] In Ecuador, tenured teachers hold permanent job positions in public educational institutions.

[4] Tenured teachers who wanted to be transferred to another school were also required to take the exams and compete for an available position.

passed national entrance exams and won merit-based selection competitions. Local education authorities were only allowed to hire teachers who had not undergone the new recruitment process with fixed-term contracts until the positions were permanently filled by tenured teachers.

Ecuador's Ministry of Education regulated the new competitive teacher recruitment processes through Ministerial Resolutions (*Acuerdos Ministeriales, AM*) and organized them into stages or components (Ministerio de Educación del Ecuador 2007; 2008; 2010; 2011). In a first stage, teacher candidates were required to pass a logical-verbal reasoning test, a pedagogical knowledge test and a subject-specific knowledge test. Teacher candidates who achieved a required minimum score became eligible candidates to compete for a permanent position. Test scores were part of a total competition score for each candidate. A second stage of the recruitment process comprised evaluating the eligible candidates' credentials: academic degrees, teaching experience, additional training courses and academic publications. Credentials were graded and added up to the total competition score. Finally, teachers were also required to present a demonstration class in front of a school board[5], which was also evaluated and added to the total competition score. The demonstration class was part of the first stage in the beginning of the policy implementation, but afterwards it became a third stage of the process itself.

The recruitment process stage weighting was slightly changed between 2007 and 2013. Nonetheless, test scores permanently represented the highest weighting, going from 45 to 55 percent of the total competition score. Table 1 describes the Ministerial Resolutions applied from the beginning of the process to the first semester of 2013[6]. Our analysis focuses on this period because it covers all of the possible recruitment processes for teachers tenured after December 2007 and employed in the 2012-2013 school year.

Along with the new recruitment regulation, the Ecuadorian government opened around 34,000 new permanent teacher positions between 2007 and 2012 (Ministerio de Educación del Ecuador 2012). Strong economic incentives to attract highly competitive teacher candidates into the public educational system were also introduced. The nominal monthly entry wage for a new tenured teacher steadily increased from US$291 in 2006 to US$396 in 2010 and quite

[5] The school board comprised the school principal or deputy, a peer teacher and two parents elected by the school's general assembly. For positions in lower and upper secondary education, a student was also included.

[6] The competitive teacher recruitment process is still in place in Ecuador. In July 2013, it was relaunched as the "I Want to be Teacher" (*Quiero Ser Maestro*) competition. Since then, entry tests have been independently designed by the National Institute of Education Evaluation (*Instituto National de Evaluación Educativa, INEVAL*).

strongly to US$775 in 2011[7], when Ecuador's new Intercultural Education Law homogenized the teacher payment scale in line with the public service payment scale (Ministerio de Educación del Ecuador 2012; Schneider, Cevallos Estarellas, and Bruns 2019). The incentives were very attractive to contract teachers[8] working at public schools, teachers working at private schools, recently-graduated teachers, and university graduates who did not hold teaching degrees but were specialists in subjects taught at schools[9].

The first years of the new Ecuadorian competitive teacher recruitment process were highly competitive. Between 2007 and 2012, 320,000 teacher candidates registered for eligibility tests, 21,200 eligible candidates passed entry tests and 18,820 successful candidates were granted a permanent teaching position (Ministerio de Educación del Ecuador 2012).

2.3. Previous Policy Evaluation Studies in Ecuador

There is scarce and mixed evidence on the effects of mandatory certification tests and competitive teacher recruitment in Ecuador. Cruz-Aguayo, Ibarrarán and Schady (2017) use data on a representative sample of children in first primary school grades to analyze whether children taught by teachers with higher test scores in Ecuador's new competitive recruitment had higher achievements in language and math in the 2011-2012 school year. They report no indication that teachers with higher (or lower) test scores were assigned to children with different observable characteristics, which allows them to estimate level and value-added to achievement models with OLS regressions. Their results do not suggest that test scores on the teacher entry competition were associated with child achievement in language or math. Subsequently, Cruz-Aguayo et al. (2017) conclude that the instrument used to decide which teachers receive tenure in Ecuador does not predict how effective a teacher is at raising math and language achievement. Nonetheless, a serious limitation with this study is that it only compares student outcomes among successful teacher candidates who passed entry tests, who therefore belong to less than 10 percent of all tested teachers. Since test score data from teachers who did not pass entry exams is not available in the study, its findings do not seem to fully support its conclusion.

Araujo P. (2019) assesses the effectiveness of Ecuador's teacher recruitment process as a quality screening device using the same school sample, but taking into account information on

[7] Teaching wage increases were not the product of high inflation rates. Ecuador's average inflation rate between 2006 and 2011 was 4.37 percent (INEC, 2020).
[8] Contract teachers are fixed-term employed teachers. The regular contract period is one year with the possibility of renewal.
[9] Ecuador's new Intercultural Education Law of 2011 officially opened the teaching career to university graduates who did not hold teaching degrees.

teachers who were not recruited through the selective entry competitions and were working at the same schools and grades in the 2011-2012 academic year. Her analysis shows that students assigned to teachers tenured by the new recruitment policy had on average parents with fewer years of education and were more likely to live in poor households, which suggests a matching of more vulnerable students to these teachers. Using propensity score matching to estimate a value-added to student achievement model, her results suggest that teachers who passed national entrance examinations and won tenured positions were more effective in raising language achievement among students living in poverty. The average treatment effect of a test-screened teacher who won an entry competition is estimated to be at least a 0.08 standard deviation gain in language for a student living in a poor household. By contrast, no effect was found for math achievement.

Nonetheless, the aforementioned studies and much of the teacher quality literature suffer from potential bias caused by the non-random assignment of students to teachers. We address this issue by using the data gathered by the "Closing Gaps" project, where a representative cohort of Ecuadorian kindergarten children were assigned to their classes and teachers with a rule as good as random starting in the 2012-2013 school year. Araujo et al. (2016) were the first to use the "Closing Gaps" data to examine the impact of teacher quality on learning outcomes in kindergarten. Their study finds teacher effects of a 0.09 standard deviation in language and math learning. The results also suggest that children assigned to teachers with higher classroom practice scores had significantly higher achievement. A one standard deviation higher teacher score is associated with higher student end-of-year test scores of between a 0.06 and 0.08 standard deviation. By contrast, children assigned to inexperienced teachers had test scores that were 0.17 standard deviations lower. None of the other teacher characteristics analyzed – including cognitive skills and personality traits – were associated with student learning.

We combine the "Closing Gaps" data with administrative teacher recruitment information from Ecuador's Ministry of Education. This allows us to estimate the causal effect on student learning of teachers screened by national entrance examinations and tenured through the new competitive recruitment process.

3. EXPERIMENTAL DESIGN AND DATA

3.1. "Closing Gaps" Project Data

The Inter-American Development Bank in cooperation with Ecuador's Ministry of Education started in 2011 the "Closing Gaps" project, a longitudinal experimental study to evaluate different dimensions of teacher quality in public schools, starting with the first grade of general basic education or kindergarten.

The "Closing Gaps" Project randomly chose a sample of 204 public schools from the coastal region of Ecuador[10] for its implementation. Even though the study sample was drawn from the coastal region, children and households in the study were generally similar to national samples (Araujo et al. 2016). The sample was limited to schools that had at least two kindergarten classes. Starting in the 2012-2013 school year, 14,930 children enrolled in kindergarten in the participating schools were randomly assigned to their classrooms and teachers. In terms of assignment, children enrolled for kindergarten in a given school were ordered by their last and first names, and then assigned to kindergarten classrooms going down the list in alternating order. Compliance with the assignment rule was very high: only 1.7 percent of children were found in classrooms other than those to which they had been assigned[11] (Araujo et al. 2016).

At the beginning of the 2012-2013 school year, the "Closing Gaps" project collected baseline data on characteristics of children's age, gender and attendance of preschool. Children were also tested with the *Test de Vocabularion en Imagenes Peabody (TVIP)* [12], a measurement of children's past learning. The project also ran a household survey that included questions on parents' education and living conditions. We use information on household assets and access to basic services from this survey to calculate a household living standard indicator[13], with a scale from 0 to 6.

---

[10] Ecuador has four natural regions: Coastal (Costa), Andean (Sierra), Amazon (Amazonía) and Insular (Islas Galápagos). Due to the particular weather conditions of each region, the school year starts in different months. The 2012-2013 school year started in April 2012 and ended in February 2013 in the Coastal and Insular regions. The same school year in the Andean and Amazon Regions started in September 2012 and ended in June 2013.

[11] No-compliers in our analysis are assigned to the classrooms to which they were originally randomly assigned. This means that we actually estimate intention to treat. Nonetheless, our intention to treat estimators should be very close to the average treatment effects, since compliance is almost 100 percent.

[12] It is the Spanish version of the Peabody Picture Vocabulary Test (PPVT) (Dunn et al. 1986). The "Closing Gaps" project provided TVIP test scores standardized on a sample of Mexican and Puerto Rican children, whose mean was set at 100 and the standard deviation at 15 at each age. In the standardizing procedure, some of the observations were imputed to the lowest or highest possible value if the raw score obtained by the child did not have a correspondence to the standardized score in the population of reference.

[13] The household living standard indicator is based on the Global Multidimensional Poverty Index (MPI) developed by the UNDP Human Development Report Office, in collaboration with the Oxford Poverty & Human Development Initiative (OPHI) for the Human Development Reports (HDRs) (Alkire et al. 2016). It aggregates

To measure student learning at the end of the school year, four tests of language and early literacy and four tests of math were applied to children individually[14]. The language and early literacy tests covered child vocabulary[15], oral comprehension and sound, letter and word recognition[16]. The math tests[17] covered number recognition, sequencing, applied math problems and identifying basic geometric figures. Test aggregates for language and math were normalized to have zero mean and unit standard deviation.

Table 2 reports summary statistics for children and families' characteristics. Children are approximately 5 years old at the beginning of the school year. Girls account for almost half of the sample and around 56 percent of children attended early childhood education. The average TVIP score is approximately 83, one standard deviation lower than the Mexican and Puerto Rican reference population used to norm the test. Parents had completed almost nine years of education on average.

In addition, Table 2 show that 14,930 children were enrolled in kindergarten in the participating schools. However, 740 children did not start the academic year in these schools, whereby 14,190 actually started and have a TVIP score. We have full baseline information for around 98 percent of the children who enrolled and actually started the school year in the participating schools, and household information on about 94 percent of them. However, our student sample size decreases to 12,632 children when additional information on teacher characteristics is taken into account in further analyses.

At the beginning of the school year, the project also collected data on conventional teacher characteristics such as gender, experience, education and tenure status. Additional rich non-conventional data on teachers' cognitive ability and personality was collected by the end of the school year. The Spanish version of the Wechsler Adult Intelligence Scale (WAIS-III) was applied to measure teacher cognitive skills. The NEO PI-R psychometric instrument (Costa Jr. and McCrae 2008) was used to assess teacher personality by measuring the so-called Big Five personality traits of neuroticism, extraversion, openness, agreeableness and

---

the following households' characteristics: access to improved sanitation and safe drinking water, type of floor, roof and exterior walls material, and asset ownership.

[14] Tests of children's inhibitory control, working memory, capacity to pay attention, and cognitive flexibility were also applied. These processes are jointly known as executive function. Our study did not find effects on children's executive function. Results are available upon request.

[15] The TVIP was applied again to evaluate child vocabulary.

[16] Oral comprehension and sound, letter and word recognition tests were taken from the Spanish-speaking version of the Woodcock-Johnson battery of tests of child development and achievement (Muñoz-Sandoval et al. 2005) and an adapted version of the Early Grade Reading Assessment (RTI International 2009b).

[17] All of the math tests were taken from the Spanish-speaking version of the Woodcock-Johnson battery (Muñoz-Sandoval et al. 2005) and an adapted version of the Early Grade Math Assessment (RTI International 2009a).

conscientiousness[18]. Under the NEO PI-R profile scale, each trait can be scored as very low (20-35), low (35-45), average (45-55), high (55-65) or very high (65-80).

Finally, to retrieve information on teacher behaviors and classroom practices, the CLASS was applied in the middle of the school year. CLASS measures teacher behaviors in three domains: emotional support, classroom organization and instructional support (Hamre, La Paro, and Pianta 2007). Each domain's score ranges from low (1-2) to medium (3-5) or high (6-7). As part of the standard application of CLASS, all teachers were filmed teaching for a full-day[19] during the 2012-2013 school year. The videos were coded by experts according to the CLASS protocol and scores for each domain were calculated. All kindergarten teachers in the school sample were also filmed in the previous school year, although there are fewer observations than for the 2012-2013 school year due to changes in the allocation of teacher staff within schools and teachers leaving the participating schools.

Table 3 reports summary statistics for teacher characteristics. Virtually all teachers are females[20]. Around 64 percent are tenured. On average, they have 15 years of experience and almost all of them have a university degree. The mean cognitive skill score is around 87, which is in the low average range of the WAIS-III international scale[21] (Wechsler and Psychological Corporation 1997). With respect to personality traits, the NEO PI-R profile scale shows a high mean score for conscientiousness, a low mean score for neuroticism, and average mean scores for openness, extraversion and agreeableness. CLASS average scores are in the medium range for socioemotional support and class management, and in the low range for instructional support in both school years. The CLASS scores show that even though teachers maintain positive relations with their students and moderately well-organized classrooms, they engage

[18] The APA Dictionary of Psychology (VandenBos 2007) defines the traits as follows: i. Neuroticism: chronic level of emotional instability and proneness to psychological distress; ii. Extraversion: an orientation of one's interest and energies toward the outer world of people and things rather than the inner world of subjective experience, characterized by positive affect and sociability; iii. Openness to Experience: the tendency to be open to new aesthetic, cultural or intellectual experiences; iv. Agreeableness: the tendency to act in a cooperative, unselfish manner; v. Conscientiousness: the tendency to be organized, responsible and hardworking. Psychology and economics' research have well stablished that standardized tests of cognitive ability and personality traits predict a variety of job performance outcomes across professions (Almlund et al. 2011; Barrick and Mount 1991; Kuncel, Ones, and Sackett 2010; Heineck and Anger 2010).

[19] Teachers did not have previous knowledge about the day on which they would be filmed.

[20] Traditionally, teachers in the first school grades in Ecuador are women. Nonetheless, as a robustness check, we ran all of our analyses only for the female teacher sample. The results were consistent and are available upon request.

[21] The average teacher cognitive skill score is not a source of concern for this study. Vast neuropsychological literature has shown that there are cultural differences in cognitive test performance. The performance of people from different cultures in the same cognitive test may vary according to the importance of the specific cognitive ability in one's own culture (Fasfous et al. 2013; Bakos et al. 2010; Rosselli and Ardila 2003).

in very few interactions that support learning. We start our teacher sample with 450 observations, but we have full information on 430 teachers in the 2012-2013 school year.

The fact that measurements of teacher cognitive skills and personality were collected by the end of the school year could raise concern about reverse causality, in case one assumes that these characteristics could be affected by the classroom environment. However, it should be noted that we do not try to find the causal effect of teacher cognitive skills and personality on student learning, but that we add these covariates at the end of our estimations to observe whether our experimental results are affected. In addition, despite evidence on a change in personality over the life cycle, most of the evidence does not point to single environmental events as the source of the change but rather a combination of factors importantly influenced by genetics (Almlund et al. 2011).

The same reverse causality concern could apply to the information on teacher practices, given that CLASS was applied in the middle of the 2012-2013 school year. We use the average CLASS scores of 2011-2012 and 2012-2013 to deal with this problem, as suggested by Araujo et al. (2016). Furthermore, we add our CLASS covariate at the end of our estimations as a robustness check.

We normalized the cognitive skills, the Big Five personality traits and the total CLASS scores to have zero mean and a unit standard deviation for our regression analyses.

3.2. Data on the New Teacher Recruitment Policy in Ecuador

We obtained additional unique information from administrative records of Ecuador's Ministry of Education on the tenure status and recruitment processes of all teachers participating in the "Closing Gaps" project at the beginning of the 2012-2013 school year. In our sample, about 13 percent of teachers had passed national entry tests and won a merit-based competitions organized by the Ministry of Education between 2007 and 2013, which granted them tenure.

Table 4 reports summary statistics of teacher characteristics for the full analyzed sample, as well as whether the teacher passed national entry tests and won a competition for tenure (henceforth *test-screened tenured* teacher). Some characteristics significantly differ between the group of *test-screened tenured* teachers and their colleagues. *Test-screened tenured* teachers have on average about two years' less experience, and all of them have university degrees and tenured positions. In terms of personality, on average they have significantly higher cognitive skill scores and are more open, extroverted and agreeable than their peers. Moreover, *test-screened tenured* teachers have slightly more years of education on average,

are more conscientious and less neurotic, and have marginally higher CLASS scores, although these differences are not statistically significant in our sample.

3.3. Validity of the Experimental Design

The successful randomization of students into classrooms and to teachers within schools is the fundamental condition behind the validity of the causal inferences that we aim to establish. To evaluate whether the random assignment to *test-screened tenured* teachers was successfully implemented, we test for balance in predetermined variables across classrooms. For this purpose, we regress teacher *test-screened tenured* status on student and family predetermined variables. We condition our estimations on school fixed effects because the random assignment of students to teachers was conducted within schools.

Table 5 provides the results of our randomization test. As shown, none of the student or family characteristics predict the likelihood that a child is assigned to a *test-screened tenured* teacher, with the exception of parents' years of schooling. Nonetheless, this variable has a regression coefficient close to zero. Moreover, an F-test for the joint significance of all of the predetermined demographic variables is statistically insignificant (p=0.32). We conclude that the random assignment to *test-screened tenured* teachers was successful, and that there is no threat to our identification strategy.

We also look at the sample attrition, which poses another threat to the experimental design. In our sample, we have children who were enrolled in kindergarten in the participating schools but never showed up (5 percent of the original sample), children who dropped out during the school year (2.8 percent), and finally children who were enrolled later (5 percent). We evaluate whether being a no-show, attritor or late enrollment is correlated with assignment to our treatment by running individual linear probability regressions of these conditions on the *test-screened tenured* teacher status, controlling for school fixed effects. Our results are presented in Table 6. Columns 1, 3 and 5, respectively, show that correlations between no-shows, attritors or late enrollments and *test-screened tenured* teacher status are statistically insignificant and virtually equal to zero. These results do not change even when we add additional teacher characteristics in Columns (2), (4) and (6). We conclude that there is no evidence that the decisions for being a no-show, attritor or late enrollment are affected by being randomly assigned to a *test-screened tenured* teacher.

4. ESTIMATION STRATEGY AND RESULTS

4.1. Estimates of Test-Screened Tenured Teacher Effects

We evaluate the effect of *test-screened tenured* teachers on learning outcomes by estimating a value-added to student achievement model – empirically developed by Hanushek (1971) and formalized by Todd and Wolpin (2003) – with the following OLS regression:

$$Y_{ics} = \rho_0 + \alpha_s + \rho_1 test\_tenured_{cs} + \rho_2 X_{ics} + \rho_3 \bar{X}_{ics} + \rho_4 C_{cs} + \rho_5 T_{cs} + \rho_6 P_{cs} + \rho_7 CLASS_{cs} + u_{ics}, \quad (1)$$

where Y_{ics} represents the end-of-year test score in language or math of child *i* in classroom *c* in school *s*. $Test\_tenured_{cs}$ is a dummy variable indicating whether the student was assigned to a *test-screened tenured* teacher and α_s is a school fixed effect component, which must be included because the random assignment of students to teachers was conducted within schools. Given that students were randomly assigned to *test-screened tenured* teachers, we do not need to include any additional control in the regression. However, we also estimate our model with additional covariates to examine the robustness of our results and increase the precision of the estimates (Duflo, Glennerster, and Kremer 2008). We gradually include as controls in the regression a vector of observable student and parent characteristics (X_{ics}), a vector of classroom averages of student and parent characteristics ($\bar{X}_{ics}$), an indicator of class size (C_{cs}), a vector of teacher observable characteristics (T_{cs}), a vector of teacher cognitive ability and personality (P_{cs}) and an indicator of teacher class practices ($CLASS_{cs}$).[22].

Standard errors are clustered at the school level in all regressions. We take this approach because although treatment occurs more precisely at the classroom level, clustering by school is a more conservative estimate of standard errors that takes into account cross-classroom correlations in errors within schools[23] (Chetty et al. 2011). In addition, our school sample comprises purely treated schools where all kindergarten classrooms are taught by a *test-screened tenured* teacher, purely control schools where no kindergarten classroom is taught by

[22] We use the average CLASS scores of the 2011-2012 and 2012-2013 school years in our estimations. Our findings do not change if we run the entire analysis with the lagged CLASS score of 2011-2012 instead of the average CLASS score. Results for math are even stronger in significance level when the lagged CLASS score is used for the estimations. However, the sample is restricted to teachers who taught kindergarten in the school sample for the entire 2011-2012 and 2012-2013 school years. All results are available upon request.

[23] We also estimated all the regressions with standard errors cluster at the classroom level. The results are in line with those provided here, but with higher significance level for the variables of interest. They are available upon request.

a *test-screened tenured* teacher, and schools with a combination of treated and control kindergarten classrooms.

As previously mentioned, the final student sample size of 12,632 children in our estimations is limited to the number of observations obtained after including all student, parent, classroom and teacher controls[24].

We report regression results for language learning in Table 7. Column (1) presents the effect of *test-screened tenured* teachers on language learning estimated without additional controls and taking into account school fixed effects. We find a significant and positive causal effect of *test-screened tenured* teachers on language learning at the 5 percent significance level. Children assigned to *test-screened tenured* teachers have a 0.105 standard deviation higher end-of-year test score in language. We present regression results that incorporate controls for student characteristics in Column (2), family characteristics in Column (3), and classroom characteristics in Column (4). Here, the results do not substantially change in terms of significance or size. In Column (5), we include controls for additional teacher observable characteristics of gender, experience and education. We find that the effect of *test-screened tenured* teachers is significant and its size increases to a 0.115 standard deviation higher end-of-year language test score. Subsequently, we add controls for teacher cognitive ability and personality in Column (6). The effect does not change in significance or size. Finally, in Column (7) we introduce our measure of teacher classroom practices (CLASS score) as a control. Interestingly, we still find a significant and positive effect of *test-screened tenured* teachers, which increases to a 0.125 standard deviation higher end-of-year language test score.

We report our estimates for math learning in Table 8, following the same structure of Table 7. As shown in Column (1), we observe a positive causal effect of *test-screened tenured* teachers of a 0.085 standard deviation higher end-of-year test score in math when no additional controls are included, albeit only at the 10 percent significance level. Once additional teacher observable characteristics are taken into account as controls in Column (5), we find some evidence of a significant effect at the 5 percent level. It finally rises to a 0.099 standard deviation higher end-of-year math test score when we include a full set of teacher characteristics in Column (7).

Aside from the effect of *test-screened tenured* teachers, the only teacher characteristics that seem to correlate with language and math end-of-year test scores are teacher experience and

[24] We also estimated all regressions without limiting the sample size to the final number of observations obtained after a full set of controls are included. All results reconfirm our findings here and are available upon request.

classroom practices (CLASS), which is in line with the conclusions of Araujo et al. (2016). In addition, we find a significant and positive association between teacher cognitive skills and language learning.

It is important to note that the size of the estimated effects are substantial. Our basic estimations of the effect of a *test-screened tenured* teacher range between 10.5 and 12.5 percent of a standard deviation of end-of-year test scores for language, and from 8.5 to 9.9 percent for math. By contrast, existing estimations of the effects of certified or test-screened teachers in the US typically range between 1 and 7 percent of a standard deviation for reading and math test scores (Clotfelter, Ladd, and Vigdor 2007; Goldhaber 2007; Goldhaber and Anthony 2007; Goldhaber, Gratz, and Theobald 2017; Harris and Sass 2007).

4.2. Estimates of Test-Screened Tenured, Other-Tenured and Test-Screened Contract Teachers

As shown in Table 4, about 13 percent of the teachers in our sample are *test-screened tenured* teachers who passed national entry exams and won a merit-based competition. Previously, we compared this group with all of their peer teachers who had not undergone the new competitive recruitment process. However, in our comparison group there are teachers tenured before 2007 by local authorities (45 percent of the full sample), contract teachers who had passed national entry exams but had not yet won a competition for tenure (12 percent of the full sample), and contract teachers who had not passed national entry exams (30 percent of the full sample). Tenured teachers and contract teachers faced very different incentives in terms of wages, job security and prospective careers. For instance, while the nominal monthly entry wage for a new tenured teacher rose to US$775 in 2011, the wage of a fixed-term contract teacher stagnated around US$300. In addition to lower wages, contract teachers faced high job uncertainty since fixed-term contracts were renewed annually up to two years. In this context, it is important to explore possible differences within the original comparison group.

Accordingly, we estimate the effect of *test-screened tenured* teachers, other-tenured teachers and test-screened contract teachers[25] compared with contract teachers on learning outcomes using a specification analogous to Equation (1):

$$Y_{ics} = \rho_0 + \alpha_s + \rho_1 Test\_tenured_{cs} + \rho_2 Other\_tenured_{cs} + \rho_3 Test\_contract_{cs} + \rho_4 X_{ics} + \rho_5 \bar{X}_{ics} + \rho_6 C_{cs} + \rho_7 T_{cs} + \rho_8 P_{cs} + \rho_9 CLASS_{cs} + u_{ics}, \quad (2)$$

[25] As a robustness check, we run all our estimations excluding test-screened contract teachers from our analysis. The results are confirmatory and available upon request.

Table 9 presents regression results for language learning. Column (1) shows a positive and significant effect of *test-screened tenured* teachers on child learning in language when no controls are taking into account other than school fixed effects. A kindergarten student taught by a *test-screened tenured* teacher has a 0.169 standard deviation higher end-of-year test score in language compared with a student assigned to a contract teacher. Children assigned to other-tenured teachers also have significantly higher end-of-year language test scores than those assigned to contract teachers. Nonetheless, the size of this effect is about half of the *test-screened tenured* teacher effect. The impact of *test-screened tenured* and other-tenured teachers on student language learning persists when controlling for additional child, family and classroom covariates, as observed in Columns (2) to (4). The other-tenured teacher effect decreases and becomes statistically insignificant when additional teacher characteristics are taken into account in Columns (5) to (7). By contrast, the effect of a *test-screened tenured* teacher remains positive, statistically significant and only slightly decreases to a 0.148 standard deviation. Curiously, the effect of test-screened contract teachers is not statistically different from the effect of contract teachers.

Regression results for math learning are reported in Table 10. We find a strong, positive and significant effect of *test-screened tenured* teachers on child learning in math. Compared with a contract teacher, the effect of a *test-screened tenured* teacher is a 0.155 standard deviation higher end-of-year math test score when no controls are included aside from school fixed effects, as shown in Column (1). Other-tenured teachers also show a positive significant effect of a 0.105 standard deviation higher end-of-year math test score. These effects hold even when additional student, family and classroom characteristics are taken into account in Columns (2) to (4). When a full set of teacher covariates is included in Columns (5) to (7), the effect of a *test-screened tenured* teacher is still positive and statistically significant, although its size decreases to a 0.129 standard deviation. By contrast, the effect of other-tenured teachers not only decreases in size to a 0.054 standard deviation, but also becomes statistically insignificant. Once again, the effect of test-screened contract teachers is not statistically different from that of contract teachers.

Our estimation findings confirm the positive and significant effect of teachers who have passed national entry exams and won selection competitions for tenure in language and math learning. Nonetheless, they also show that test-screened teachers who have not won a selection competition do not have the same effect. These results suggest that the entire teacher selective process drives the positive effects on student achievement, and not only the testing component. They also indicate that test score differences among teachers who passed entry examinations

might be relevant. Finally, the results point to a positive association between permanent job status and performance, in contrast to Araujo et al.'s (2016) findings[26].

4.3. Estimates of Test-Screened Tenured Teachers, Accounting for Ministerial Resolution Competitions

Among the *test-screened tenured* teachers in our sample, around 25 percent won competitions regulated by the original Ministerial Resolution of December 2007 (AM No. 438-07), only about 4 percent won competitions organized under the Ministerial Resolution of January 2010 (AM No. 018-10)[27], and 71 percent won competitions regulated by the Ministerial Resolution of November 2011 (AM No. 379-11). The impact of each competition might be different. On the one hand, the competition component weighting slightly changed over time, as shown in Table 1. The overall test weighting increased from 45 to 55 percent, while the demonstration class weighting decreased from 20 to 10 percent. On the other hand, the test quality might have differed over time and among competitions. Even though the same type of teacher skill and subject knowledge tests were applied between 2007 and 2013, there is no evidence that they were psychometrically comparable.

Consequently, we estimate the effect of *test-screened tenured* teachers on learning outcomes accounting for the Ministerial Resolution that regulated each selection competition. Accordingly, we use an extended specification of Equation (1):

$$Y_{ics} = \rho_0 + \alpha_s + \rho_1 Test\_tenured\_AM438_{cs} + \rho_2 Test\_tenured\_AM018_{cs} + \rho_3 Test\_tenured\_AM379 + \rho_4 X_{ics} + \rho_5 \bar{X}_{ics} + \rho_6 C_{cs} + \rho_7 T_{cs} + \rho_8 P_{cs} + \rho_9 CLASS_{cs} + u_{ics}, \quad (3)$$

The regression results for language learning are presented in Table 11. All of the model specifications show that the effect of *test-screened tenured* teachers who won competitions organized under the 2007 Regulation is stronger in size and significance than those of teachers who won competitions organized under other regulations. The size of the effect ranges from a

[26] Araujo et al. (2016) did not find a significant association between tenure status and end-of-year student outcomes. Our results differ because we discriminate between *test-screen tenured* teachers and other-tenure teachers. The results of Araujo et al. might be driven by other-tenured teachers whose effects fade out after additional controls are taken into account. In addition, our information on tenure status at the beginning of the school year was confirmed by the Ministry's administrative data, whereas Araujo et al.'s came only from the "Closing Gaps" teacher survey.

[27] The sample size of teachers who won competitions organized under the Ministerial Resolution of January 2010 is too small to obtain robust estimations. However, these teachers are not excluded from our estimations to preserve the integrity of the analysis and sample size.

0.160 standard deviation higher end-of-year language test score in Column (1) with no controls other than school fixed effects to a 0.185 standard deviation in Column (7) where the full set of child, family, classroom and teacher covariates are taken into account. By contrast, the effect of *test-screened tenured* teachers who won competitions organized under the 2011 Regulation ranges from a positive but statistically insignificant 0.090 standard deviation higher end-of-year test score in Column (1) to a significant 0.102 standard deviation in Column (7).

Table 12 presents estimation results for math learning. Once again, we find a stronger positive effect of *test-screened tenured* teachers who won competitions under the 2007 Regulation. Column (1) shows that the children randomly assigned to *test-screened tenured* teachers who won competitions organized under the 2007 Regulation achieve a 0.172 standard deviation higher end-of-year math test score, which is highly significant. When additional child, family, classroom and teacher covariates are included, this effect increases to a 0.183 standard deviation in Column (7). We also find positive effects of *test-screened tenured* teachers who won competitions organized under the 2011 Regulation, although they are not statistically significant.

Finally, estimation results of the effects of *test-screened tenured* teachers who won competitions organized under the 2010 Regulation are not statistically different from zero for language and math, although they should be treated with considerable caution due to the small sample size of these teachers.

To sum up, our findings suggest that the teacher selection competitions regulated by the original Ministerial Resolution of December 2007 were more effective in recruiting teachers who have an impact on student learning.

5. HETEROGENEOUS EFFECTS

There is an ongoing debate on the extent to which highly-qualified teachers can close learning gaps between socioeconomically advantaged and disadvantaged children (Borman and Kimball 2005; Boyd et al. 2008; Hanushek et al. 2020; James and Wyckoff 2020; Phillips 2010). This is particularly important for Ecuador and other Latin American countries due to the large and persistent differences found in the cognitive development of children of high and low socioeconomic status throughout the school (Schady et al. 2015).

Consequently, we look at heterogeneous effects of *test-screened tenured* teachers among children who started kindergarten with different language skill development levels. First, we use the TVIP baseline score to sort the student sample into quintiles from the lowest to the

highest score. Subsequently, we implement our regression model for each TVIP quintile, as formalized in Equation (1)[28]. The results for language learning are presented in Table 13, and they suggest that the effect of *test-screened tenured* teachers on language is stronger in size and significance for children at the lowest TVIP quintile. While this effect is statistically not different from zero for children at the highest TVIP quintile as displayed in Columns (9) and (10), it ranges between a 0.240 and 0.214 standard deviation higher end-of-year test score for children at the lowest TVIP quintile as shown in Columns (1) and (2). Table 14 presents regression results for math learning. In contrast to the previous finding, the effect of *test-screened tenured* teachers on math is stronger for children at the middle-upper TVIP baseline quintiles, as shown in Columns (5) to (8).

We also examine the effects of *test-screened tenured* teachers among children from different socioeconomic backgrounds. We use our living standard indicator (LSI) standardized score to sort the student sample into quartiles starting with the lowest score. Subsequently, we estimate our regression model for each LSI quartile.

Table 15 presents estimation results for language. Our results suggest that the effects of *test-screened tenured* teachers on language are stronger in size and significance for children in the lowest and highest LSI quartile. The effect for children in the lowest LSI quartile ranges from a 0.142 to a 0.200 standard deviation higher end-of-year test score, as presented in Columns (1) to (2). By contrast, we find no heterogeneous effect on math learning as shown in Table 16.

Overall, our analysis of heterogeneous effects suggest that the impact of *test-screened tenured* teachers on language learning is stronger for vulnerable children who started the school year with lower TVIP scores or came from socioeconomically disadvantaged households. This is not the case for math learning. These results are in good agreement with the findings of Araujo P. (2019).

6. ROBUSTNESS CHECK

In this section, we show that our results are robust to using the subsample of schools that have at least one treated kindergarten classroom: in other words, a classroom taught by a *test-screened tenured* teacher.

[28] We also estimated quantile regressions for several quantile values in reading and math. The results are in line with those provided here and are available upon request.

Our original school sample comprises 204 schools and all of their kindergarten classrooms. Among them, 161 are purely control schools that have no kindergarten classroom taught by a *test-screened tenured* teacher, while another six schools are purely treated schools where all kindergarten classrooms are taught by *test-screened tenured* teachers. Finally, 37 schools in our sample have a combination of treated and control kindergarten classrooms. This is the subsample that we are interested in for our robustness check, although first we exclude a unique school whose control group is a classroom taught by a test-screened contract teacher, which guarantees that all of our control classrooms are taught by teachers who have not passed any stage of the new teacher recruitment process. Our final subsample comprises 36 schools and 84 teachers.

We repeat our original analysis with the subsample of 36 schools[29]. First, we evaluate the randomization of students into classrooms by regressing teacher *test-screened tenured* status on student and family predetermined covariates, conditioned on school fixed effects. Our results are presented in Table 17 and show that none of the student or family characteristics predict the likelihood that a child is assigned to a *test-screened tenured* teacher at the 5 percent significance level. Once again, parents' years of schooling is marginally correlated with assignment to a *test-screened tenured* teacher, albeit only at the 10 percent significance level and with a very small coefficient. Moreover, the F-test for the joint significance of all of the predetermined demographic variables is statistically insignificant (p=0.287). We conclude that the random assignment to *test-screened tenured* teachers was also successful for the subsample of 36 schools.

Subsequently, we evaluate whether being a no-show, attritor or late enrollment is correlated with assignment to the treatment in our subsample of schools. The results are presented in Table 18, again showing that there is no evidence that being randomly assigned to a *test-screened tenured* teacher has an effect on the decision to be a no-show, attritor or late enrollment.

Finally, we estimate our regression model as described in Equation (1) for language and math end-of-year test scores and present them in Table 19. The student sample size substantially decreases from 12,632 to 2,393 kindergarten children. Nonetheless, we still find that students randomly assigned to *test-screened tenured* teachers achieve at least a 0.102 standard deviation significantly higher end-of-year language test score, as shown in Columns

[29] We re-run all our econometric analyses for the subsample of 36 schools, but we present here the main estimations. All results were confirmatory and are available upon request.

(1) to (4). In math, the effect of *test-screened tenured* teachers is at least a 0.089 standard deviation higher end-of-year test score as shown in Columns (5) to (8), although its significance is only found at the 10 percent level. In order to have a conservative estimation of our standard errors, all regression are clustered at the school level. However, when standard errors are clustered at the classroom level where treatment occurred, the significance level of the *test-screened tenured* teacher effect increases to 1 percent for language and math in all of our estimations[30].

These results are practically identical to those found in our original analysis and have further strengthened our confidence in the significant and unbiased effect of Ecuadorian *test-screened tenured* teachers on kindergarten learning.

7. CONCLUSIONS

In this paper, we have assessed the effectiveness of Ecuador's new teacher recruitment policy, which from 2007 onwards required teacher candidates to pass mandatory skill and content knowledge tests before they were allowed to participate in merit-based selection competitions for tenure at public schools. For our identification strategy, we combined administrative teacher recruitment information from Ecuador's Ministry of Education with data provided by the "Closing Gaps" project, which randomly assigned a representative sample of kindergarten children to their classrooms and teachers in the 2012-2013 school year.

Our work has led us to conclude that teachers who passed national entry tests and won a competition for tenure in Ecuador have positive and significant effects on language learning in kindergarten, which persist even after controlling for teacher cognitive skill, personality and classroom practice. Children randomly assigned to a *test-screened tenured* teacher achieved between a 0.105 and a 0.125 standard deviation higher end-of-year language test score in the 2012-2013 school year. Moreover, *test-screened tenured* teachers have a robust effect of at least a 0.137 standard deviation compared with contract teachers. *Test-screened tenured* teachers also outperform teachers tenured before 2007 who were not required to pass national entry tests.

The evidence obtained in our study also suggests positive effects of *test-screened tenured* teachers on math learning in kindergarten. Children randomly assigned to these teachers achieved between a 0.085 and a 0.099 standard deviation higher end-of-year math test score in

[30] The results of our estimations for the subsample of 36 schools with standard errors clustered at the classroom level were confirmatory and are available upon request.

the 2012-2013 school year. When compared with a contract teacher, the effect of a *test-screened tenured* teacher is at least a 0.133 standard deviation, which is robust and highly significant. Likewise, the effects of *test-screened tenured* teachers are larger in size and significance than those of teachers tenured by previous processes.

Surprisingly, we do not find similar effects for teachers who have passed entry tests but have not won a completion for tenure, and therefore work with fixed-term contracts. These results suggest that differences in skill and subject knowledge test scores among *test-screened* teachers might be a relevant indicator of teacher quality. Moreover, these results point to a positive association between job status and performance in Ecuador.

We also explored potential differences among teacher selection processes organized under three different regulations between 2007 and 2012. Our results suggest that competitions regulated by the original Ministerial Resolution of December 2007 recruited more effective teachers. There are two potential explanations behind this result. On the one hand, this is the regulation that assigns the highest weighting to the teacher demonstration class component. Our results, the findings of Araujo et al. (2016) as well as increasing international evidence (Bacher-Hicks et al. 2019; Kane et al. 2011; 2013) suggest that teacher classroom practices are good indicators of teacher effectiveness. Accordingly, it is possible that the results are driven by the weighting assigned to the classroom practice evaluation. On the other hand, it is possible that the results are driven by quality differences in skill and content knowledge tests among the processes organized under different regulations.

Remarkably, our study also confirms the potential effectiveness of *test-screened tenured* teachers in closing learning gaps between socioeconomically advantaged and disadvantaged children in Ecuador. We find that the effects of *test-screened tenured* teachers on language learning are stronger for children who started the school year with the lowest baseline TVIP scores or those who came from socioeconomically disadvantaged households. This is not the case for math learning.

In addition, as part of our robustness check, we conducted our entire analysis for the subsample of schools that have at least one classroom randomly assigned to a *test-screened tenured* teacher. Even though our student sample decreased to one-fifth of the original size, our estimation results are practically the same in size and significance as those of the full sample. These results emphasize the validity of our original estimations and further strengthen our confidence in our findings of the unbiased significant and positive causal effect of Ecuadorian *test-screened tenured* teachers on kindergarten learning.

We conclude by drawing some policy implications from our analysis. Our results show that between 2007 and 2012 the Ecuadorian teacher reform succeeded in recruiting more effective teachers. *Test-screened tenured* teachers were significantly more successful than their peers in raising kindergarten student learning in the 2012-2013 school year. It is likely that the mechanism behind the observed results is a combination of screening teacher candidates who demonstrate higher cognitive skills, greater content knowledge of the subject taught and better class practices, along with the provision of an economically attractive permanent job position. Under this scenario, it is desirable for policy-makers in other Latin American countries to introduce objective and highly competitive teacher recruitment processes to screen and select the best available candidates to improve teacher quality.

REFERENCES

Alkire, Sabina, Adriana Conconi, Gisela Robles, José Manuel Roche, María Emma Santos, Suman Seth, and Ana Vaz. 2016. "The Global Multidimensional Poverty Index (MPI): 5-Year Methodological Note." 37. *OPHI Briefing*. OPHI Briefing. Oxford. https://www.ophi.org.uk/wp-content/uploads/MPI_Methodology_2010-2015_Jan2016.pdf.

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. 2011. "Personality Psychology and Economics." In *Handbook of The Economics of Education*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, Volume 4, 1–181. Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-444-53444-6.00001-8.

Angrist, Joshua D., and Jonathan Guryan. 2008. "Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements." *Economics of Education Review*. https://doi.org/10.1016/j.econedurev.2007.03.002.

Araujo, M. Caridad, Pedro Carneiro, Yyannú Cruz-Aguayo, and Norbert Schady. 2016. "Teacher Quality and Learning Outcomes in Kindergarten." *The Quarterly Journal of Economics* 131 (3): 1415–53. https://doi.org/10.1093/qje/qjw016.

Araujo P., M. Daniela. 2019. "Measuring the Effect of Competitive Teacher Recruitment on Student Achievement: Evidence from Ecuador." 150. BERG Working Paper Series. Bamberg. https://www.uni-bamberg.de/fileadmin/uni/fakultaeten/sowi_faecher/vwl/BERG/BERG_150.pdf.

Araujo P., M. Daniela, and Daniela Bramwell. 2015. "Cambios En La Política Educativa En Ecuador Desde El Año 2000." ED/EFA/MRT/2015/PI/35. *Education for All Global Monitorin Report UNESCO*. Paper Commissioned for the EFA Global Monitoring Report 2015, Education for All 2000-2015. Paris. https://unesdoc.unesco.org/ark:/48223/pf0000232430.

Asamblea Nacional del Ecuador. 2011. *LEY ORGÁNICA DE EDUCACIÓN INTERCULTURAL*. *Registro Oficial Nº 417*. Ecuador.

Bacher-Hicks, Andrew, Mark J. Chin, Thomas J. Kane, and Douglas O. Staiger. 2019. "An Experimental Evaluation of Three Teacher Quality Measures: Value-Added, Classroom Observations, and Student Surveys." *Economics of Education Review* 73 (December): 101919. https://doi.org/10.1016/j.econedurev.2019.101919.

Bakos, Daniela Schneider, Natalie Denburg, Rochele Paz Fonseca, and Maria Alice de Mattos Pimenta Parente. 2010. "A Cultural Study on Decision Making: Performance Differences on the Iowa Gambling Task between Selected Groups of Brazilians and Americans." *Psychology & Neuroscience* 3 (1): 101–7. https://doi.org/10.3922/j.psns.2010.1.013.

Barrick, Murray R., and Michael K. Mount. 1991. "The Big Five Personality Dimensions and Job Performance: A Meta-Analysis." *Personnel Psychology* 44 (1): 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x.

Bau, Natalie, and Jishnu Das. 2020. "Teacher Value Added in a Low-Income Country." *American Economic Journal: Economic Policy* 12 (1): 62–96. https://doi.org/10.1257/pol.20170243.

Bietenbeck, Jan, Marc Piopiunik, and Simon Wiederhold. 2018. "Africa's Skill Tragedy: Does Teachers' Lack of Knowledge Lead to Low Student Performance?" *Journal of Human Resources* 53 (3): 553–78. https://doi.org/10.3368/jhr.53.3.0616-8002R1.

Borman, Geoffrey D., and Steven M. Kimball. 2005. "Teacher Quality and Educational Equality: Do Teachers with Higher Standards-Based Evaluation Ratings Close Student Achievement Gaps?" *The Elementary School Journal* 106 (1): 3–20. https://doi.org/10.1086/496904.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, Jonah E. Rockoff, and James Wyckoff. 2008. "The Narrowing Gap in New York City Teacher Qualifications and Its Implications for Student Achievement in High-Poverty Schools." *Journal of Policy Analysis and Management* 27 (4): 793–818. https://doi.org/10.1002/pam.20377.

Bruns, Barbara, and Javier Luque. 2015. "Great Teachers: How to Raise Student Learning in Latin American and the Caribbean." Washington D.C.: The World Bank. https://doi.org/10.1596/978-1-4648-0151-8.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *The Quarterly Journal of Economics* 126 (4): 1593–1660. https://doi.org/10.1093/qje/qjr041.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–79. https://doi.org/10.1257/aer.104.9.2633.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources* 41 (April 2005): 778–820. https://doi.org/10.2307/40057291.

———. 2007. "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects." *Economics of Education Review* 26 (6): 673–82.

https://doi.org/10.1016/j.econedurev.2007.10.002.

Costa Jr., Paul T, and Robert R McCrae. 2008. “The Revised NEO Personality Inventory (NEO-PI-R).” In *The SAGE Handbook of Personality Theory and Assessment, Vol 2: Personality Measurement and Testing.*, 179–98. Thousand Oaks, CA, US: Sage Publications, Inc. https://doi.org/10.4135/9781849200479.n9.

Cruz-Aguayo, Yyannú, Pablo Ibarrarán, and Norbert Schady. 2017. “Do Tests Applied to Teachers Predict Their Effectiveness?” *Economics Letters* 159 (October): 108–11. https://doi.org/10.1016/J.ECONLET.2017.06.035.

D’Agostino, Jerome V., and Sonya J. Powers. 2009. “Predicting Teacher Performance With Test Scores and Grade Point Average: A Meta-Analysis.” *American Educational Research Journal* 46 (1): 146–82. https://doi.org/10.3102/0002831208323280.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. “School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools.” *Journal of Public Economics*. https://doi.org/10.1016/j.jpubeco.2014.11.008.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2008. “Using Randomization in Development Economics Research: A Toolkit.” In *T. Schultz and John Strauss, Eds., Handbook of Development Economics*. Vol. 4. Amsterdam and New York: North Holland.

Dunn, Lloyd, Eligio Padilla, Delia Lugo, and Leota Dunn. 1986. *Test de Vocabulario En Imagenes Peabody*. Circle Pines, MN: American Guidance Service.

Elacqua, Gregory, Diana Hincapié, Emiliana Vegas, and Mariana. Alfonso. 2017. “Profesión: Profesor En América Latina. ¿Por Qué Se Perdió El Prestigio Docente y Cómo Recuperarlo?” Washington, D.C.: Inter-American Development Bank. https://doi.org/10.18235/0000901.

Fasfous, Ahmed F, Natalia Hidalgo-Ruzzante, Raquel Vilar-López, Andrés Catena-Martínez, and Miguel Pérez-García. 2013. “Cultural Differences in Neuropsychological Abilities Required to Perform Intelligence Tasks.” *Archives of Clinical Neuropsychology* 28 (8): 784–90. https://doi.org/10.1093/arclin/act074.

Glewwe, Paul. W., Eric A. Hanushek, Sarah D. Humpage, and Renato Ravina. 2014. “School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010.” In *Education Policy in Developing Countries*, edited by Paul Glewwe, 13–64. Chicago and London: The University of Chicago Press. https://doi.org/10.3386/w17554.

Goldhaber, Dan. 2007. “Everyone’s Doing It, but What Does Teacher Testing Tell Us about Teacher Effectiveness?” *The Journal of Human Resources* XLII (4): 765–94. https://doi.org/10.2307/40057329.

Goldhaber, Dan, and Emily Anthony. 2007. “Can Teacher Quality Be Effectively Assessed? National Board Certification as a Signal of Effective Teaching.” *The Review of Economics and Statistics* 89 (1): 134–50. https://doi.org/10.1162/rest.89.1.134.

Goldhaber, Dan, Trevor Gratz, and Roddy Theobald. 2017. “What’s in a Teacher Test?

Assessing the Relationship between Teacher Licensure Test Scores and Student STEM Achievement and Course-Taking." *Economics of Education Review* 61 (December): 112–29. https://doi.org/10.1016/j.econedurev.2017.09.002.

Hamre, Bridget K., Karen M. La Paro, and Robert C. Pianta. 2007. *Classroom Assessment Scoring System (CLASS) Manual*. Baltimore: Paul H. Brookes Publishing Co., Inc.

Hanushek, Eric A. 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." *The American Economic Review* 61: 280–88. https://doi.org/10.2307/1817003.

Hanushek, Eric A., Paul Peterson, Laura Talpey, and Ludger Woessmann. 2020. "Long-Run Trends in the U.S. SES-Achievement Gap." 26764. *National Bureau of Economic Research*. https://doi.org/10.3386/w26764.

Hanushek, Eric A., and Steven G. Rivkin. 2006. "Teacher Quality." In *Handbook of the Economics of Education*, edited by E. Hanushek and F. Welch, 1052–75. Elsevier. https://doi.org/10.1016/S1574-0692(06)02018-6.

———. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." In *American Economic Review*. https://doi.org/10.1257/aer.100.2.267.

———. 2012. "The Distribution of Teacher Quality and Implications for Policy." *Annual Review of Economics* 4 (1): 131–57. https://doi.org/10.1146/annurev-economics-080511-111001.

Harris, Douglas N., and Tim R. Sass. 2007. "The Effects of NBPTS- Certified Teachers on Student." *Journal of Policy Analysis and Management*. WileyAssociation for Public Policy Analysis and Management. https://doi.org/10.2307/29738986.

———. 2011. "Teacher Training, Teacher Quality and Student Achievement." *Journal of Public Economics* 95 (7–8): 798–812. https://doi.org/10.1016/J.JPUBECO.2010.11.009.

Heineck, Guido, and Silke Anger. 2010. "The Returns to Cognitive Abilities and Personality Traits in Germany." *Labour Economics* 17 (3): 535–46. https://doi.org/10.1016/j.labeco.2009.06.001.

Instituto Ecuatoriano de Estadísticas y Censos (INEC). 2020. "Serie Histórica IPC." Ecuador En Cifras. 2020. https://www.ecuadorencifras.gob.ec/historicos-ipc/.

Jackson, C. Kirabo, Jonah E. Rockoff, and Douglas O. Staiger. 2014. "Teacher Effects and Teacher-Related Policies." *Annual Review of Economics* 6: 801–25. https://doi.org/10.1146/annurev-economics-080213-040845.

James, Jessalynn, and James Wyckoff. 2020. "Teacher Labor Markets: An Overview." In *The Economics of Education*, edited by Steve Bradley and Colin B T Green, 355–70. Elsevier. https://doi.org/10.1016/b978-0-12-815391-8.00026-4.

Kane, Thomas J., Daniel F Mccaffrey, Trey Miller, and Douglas O. Staiger. 2013. "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment." *MET Project Reseach Paper*. Bill & Melinda Gates Foundation. www.metproject.org.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. “What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City.” *Economics of Education Review* 27: 615–31. https://doi.org/10.1016/j.econedurev.2007.05.005.

Kane, Thomas J., Eric S. Taylor, John H. Tayler, and Amy L. Wooten. 2011. “Identifying Effective Classroom Practices Using Student Achievement Data.” *The Journal of Human Resources* 46 (3): 587–613. http://www.jstor.org/stable/41304833.

Koedel, Cory, and Julian R. Betts. 2011. “Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique.” *Education Finance and Policy* 6 (1): 18–42. https://doi.org/10.1162/EDFP_a_00027.

Koedel, Cory, Kata Mihaly, and Jonah E. Rockoff. 2015. “Value-Added Modeling: A Review.” *Economics of Education Review* 47 (August): 180–95. https://doi.org/10.1016/J.ECONEDUREV.2015.01.006.

Kuncel, Nathan R., Deniz S. Ones, and Paul R. Sackett. 2010. “Individual Differences as Predictors of Work, Educational, and Broad Life Outcomes.” *Personality and Individual Differences*. https://doi.org/10.1016/j.paid.2010.03.042.

Metzler, Johannes, and Ludger Woessmann. 2012. “The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from within-Teacher within-Student Variation.” *Journal of Development Economics* 99 (2): 486–96. https://doi.org/10.1016/J.JDEVECO.2012.06.002.

Ministerio de Educación del Ecuador. 2007. *Acuerdo Ministerial 438-07: NORMAS DE LOS CONCURSOS DE MÉRITOS Y OPOSICIÓN PARA EL INGRESO AL MAGISTERIO NACIONAL*. *Acuerdos Ministeriales*. Quito, Ecuador: Ministerio de Educación del Ecuador (MinEduc).

———. 2008. *Acuerdo Ministerial 363: NORMAS PARA LLENAR VACANTES A TRAVÉS DEL SISTEMA DE RUEDAS DE CAMBIOS*. *Acuerdos Ministeriales*. Ecuador: Ministerio de Educación del Ecuador (MinEduc).

———. 2010. *Acuerdo Ministerial 018-10: NORMATIVA PARA LOS CONCURSOS DE MÉRITOS Y OPOSICIÓN PARA LLENAR VACANTES DEL MAGISTERIO NACIONAL*. *Acuerdos Ministeriales*. Ecuador: Ministerio de Educación del Ecuador (MinEduc).

———. 2011. *Acuerdo Ministerial 379-11: NORMATIVA DE CONCURSOS DE MÉRITOS Y OPOSICIÓN PARA LLENAR VACANTES DE DOCENTES EN EL SECTOR PÚBLICO*. *Acuerdos Ministeriales*. Ecuador: Ministerio de Educación del Ecuador (MinEduc).

———. 2012. “Ecuador 2007-2012 a Transformation in Education. Presentation at the Inter-American Development Bank.” Washington D.C.: Ministerio de Educación del Ecuador (MinEduc).

Muñoz-Sandoval, Ana F., Richard W. Woodcock, Kevin S. McGrew, and N. Mather. 2005. *Batería III Woodcock-Muñoz*. Itasca, IL: Riverside Publishing.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2013. “Contract Teachers:

Experimental Evidence from India. NBER Working Paper No. 19440." *National Bureau of Economic Research*. https://doi.org/10.3386/w19440.

Phillips, Kristie J R. 2010. "What Does 'Highly Qualified' Mean for Student Achievement? Evaluating the Relationships between Teacher Quality Indicators and At-Risk Students' Mathematics and Reading Achievement Gains in First Grade." *Elementary School Journal* 110 (4): 464–93. https://doi.org/10.1086/651192.

Rockoff, Jonah E., Brian A. Jacob, Thomas J. Kane, and Douglas O. Staiger. 2011. "Can You Recognize an Effective Teacher When You Recruit One?" *Education Finance and Policy* 6 (1): 43–74. https://doi.org/10.1162/EDFP_a_00022.

Rosselli, Mónica, and Alfredo Ardila. 2003. "The Impact of Culture and Education on Non-Verbal Neuropsychological Measurements: A Critical Review." *Brain and Cognition* 52 (3): 326–33. https://doi.org/10.1016/S0278-2626(03)00170-2.

Rothstein, Jesse. 2009. "Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables." *Education Finance and Policy* 4 (4): 537–71. https://doi.org/10.1162/edfp.2009.4.4.537.

———. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *The Quarterly Journal of Economics* 125 (1): 175–214. http://www.jstor.org/stable/40506280.

RTI International. 2009a. "Early Grade Mathematics Assessment (EGMA): A Conceptual Framework Based on Mathematics Skills Development in Children." Research Triangle Park, NC. https://www.globalpartnership.org/content/early-grade-mathematics-assessment-egma-conceptual-framework-based-mathematics-skills.

———. 2009b. "Early Grade Reading Assessment Toolkit." Research Triangle Park, NC. https://globalreadingnetwork.net/eddata/early-grade-reading-assessment-toolkit-2009.

Schady, Norbert, Jere R. Behrman, M. Caridad Araujo, Rodrigo Azuero, Raquel Bernal, David Bravo, Florencia Lopez-Boo, et al. 2015. "Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries." *Journal of Human Resources* 50 (2): 446–63. https://doi.org/10.3368/jhr.50.2.446.

Schneider, Ben Ross, Pablo Cevallos Estarellas, and Barbara Bruns. 2019. "The Politics of Transforming Education in Ecuador: Confrontation and Continuity, 2006–17." *Comparative Education Review*, March, 000–000. https://doi.org/10.1086/702609.

Todd, Petra E., and Kenneth I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *The Economic Journal* 113 (485): F3–33. https://doi.org/10.1111/1468-0297.00097.

VandenBos, Gary R. (Ed.). 2007. *APA Dictionary of Psychology*. Washington, DC, US: American Psychological Association.

Wayne, Andrew, and Peter Youngs. 2003. "Teacher Characteristics and Student Achievement Gains: A Review." *Review of Educational Research* 73 (1): 89–122. http://www.jstor.org/stable/3516044.

Wechsler, David, and Psychological Corporation. 1997. *WAIS-III: Administration and Scoring Manual: Wechsler Adult Intelligence Scale*. Third. San Antonio, TX: Psychological Corporation.

World Bank. 2019. “Education Statistics (EdStats).” 2019. http://datatopics.worldbank.org/education/country/ecuador.

TABLES

Table 1: Regulations and Components of Ecuador's Competitive Teacher Recruitment

	MINISTERIAL REGULATION					
	AM No. 438-07 December 2007		**AM No. 018-10 January 2010**		**AM No. 379-11 November 2011**	
Competition Components	Weight	Use of Score	Weight	Use of Score	Weight	Use of Score
Tests	**45%**		**45%**		**55%**	
– Logical-Verbal Reasoning	15%	Eligibility & Ranking	15%	Eligibility & Ranking	15%	Eligibility & Ranking
– Pedagogical Knowledge	15%	Eligibility & Ranking	15%	Eligibility & Ranking	15%	Eligibility & Ranking
– Subject-Specific Knowledge	15%	Eligibility & Ranking	15%	Eligibility & Ranking	25%	Eligibility & Ranking
Demonstration Class	**20%**	Eligibility & Ranking	**15%**	Ranking	**10%**	Ranking
Teacher Credentials	**35%**		**40%**		**35%**	
– Academic Degree	20%	Ranking	20%	Ranking	20%	Ranking
– Training and Publications	10%	Ranking	10%	Ranking	5%	Ranking
– Teaching Experience	5%	Ranking	10%	Ranking	10%	Ranking
Minimum Eligibility Threshold	– 60% of Eligibility Instruments		– 60% of Eligibility Instruments		– 60% of Logical-Verbal and Pedagogical Knowledge Tests – 70% of Subject-Specific Knowledge Tests	
Issued	Dec-07		Jan-10		Nov-11	
Abolished	Jan-10		Nov-11		May-13	

Source: Araujo P. (2019)

Table 2: Summary Statistics for Children and Families' Characteristics

	Mean	Sd.	Obs.
Children:			
Proportion female	0.49	0.50	14930
Proportion who attended preschool	0.56	0.50	14925
Age (months)	60.34	5.11	14841
TVIP	83.24	16.89	14187
Family:			
Parent's years of schooling	8.69	3.42	13275
Living standard indicator	3.33	1.38	13744

Note: This table reports means and standard deviations of the characteristics of kindergarten children and their families. TVIP stands for *Test de Vocabulario en Imágenes Peabod*y, the Spanish version of the Peabody Picture Vocabulary Test (PPVT). The test was standardized to have a mean of 100 and the standard deviation of 15 at each age, based on a reference sample of Mexican and Puerto Rican children. Family living standard indicator aggregates the following households' characteristics: access to improved sanitation and safe drinking water, type of floor, roof and exterior walls material, and assets ownership.

Table 3: Summary Statistics for Teacher Characteristics

	Mean	Sd.	Obs.
Proportion female	0.99	0.10	450
Proportion tenure	0.64	0.48	450
Years of experience	14.91	8.88	450
Years of education	17.14	1.93	450
Teacher has university degree	0.99	0.11	450
Cognitive skills	86.46	9.53	430
Neuroticism	43.85	6.72	430
Extraversion	45.65	6.83	430
Openness	50.82	6.75	430
Agreeableness	48.22	7.56	430
Conscientiousness	57.55	8.15	430
CLASS total score 2011-2012	3.63	0.37	341
Socio emotional support	4.30	0.40	341
Classroom management	4.99	0.63	341
Instructional support	1.36	0.27	341
CLASS total score 2012-2013	3.41	0.28	450
Socio emotional support	4.07	0.33	450
Classroom management	4.79	0.47	450
Instructional support	1.15	0.18	450

Note: This table reports means and standard deviations of teacher characteristics. Cognitive skills were measured with the Spanish version of the Wechsler Adult Intelligence Scale (WAIS-III). The test is internationally normed so that 100 is the median score for the adult population. The Big Five personality trait scores (openness, conscientiousness, extraversion, agreeableness and neuroticism) were obtained with the NEO PI-R psychometric instrument. Each personality trait can be scored as very low (20-35), low (35-45), average (45-55), high (55-65) or very high (65-80). CLASS stands for Classroom Assessment Scoring System. CLASS domains can be scored as low (scores 1-2), medium (3-5) or high (6-7).

Table 4: Teacher Sample Characteristics

	Full sample	Test-screened tenured teachers		
		YES	NO	Difference
Proportion female	0.99	0.96	0.99	-0.03
	(0.005)	(0.026)	(0.004)	(0.026)
Proportion tenured	0.65	1.00	0.60	0.40***
	(0.023)	(.)	(0.025)	(0.025)
Years of experience	14.74	12.84	15.01	-2.17**
	(0.415)	(0.963)	(0.453)	(1.064)
Years of education	17.15	17.46	17.11	0.36
	(0.093)	(0.259)	(0.099)	(0.277)
University degree	0.99	1.00	0.99	0.01**
	(0.005)	(.)	(0.006)	(0.006)
Cognitive skills	86.46	89.93	85.96	3.96***
	(0.459)	(1.324)	(0.485)	(1.410)
Neuroticism	43.85	43.22	43.94	-0.73
	(0.324)	(0.925)	(0.346)	(0.988)
Extraversion	45.65	48.55	45.24	3.31***
	(0.329)	(0.878)	(0.350)	(0.946)
Openness	50.82	52.97	50.51	2.46**
	(0.325)	(0.887)	(0.347)	(0.953)
Agreeableness	48.22	50.63	47.87	2.76***
	(0.365)	(0.955)	(0.391)	(1.032)
Conscientiousness	57.55	58.64	57.39	1.25
	(0.393)	(0.951)	(0.428)	(1.043)
CLASS average 2011-2012	3.48	3.50	3.48	0.02
	(0.014)	(0.037)	(0.015)	(0.040)
Observations	430	54	376	.

Note: This table reports means and standard deviations of teacher characteristics for the full analyzed sample, as well as whether the teacher passed national entry tests and won a competition for tenure (test-screened tenured teacher). Cognitive skills were measured with the Spanish version of the Wechsler Adult Intelligence Scale (WAIS-III). The test is internationally normed so that 100 is the median score for the adult population. The Big Five personality trait scores (openness, conscientiousness, extraversion, agreeableness and neuroticism) were obtained with the NEO PI-R psychometric instrument. Each personality trait can be scored as very low (20-35), low (35-45), average (45-55), high (55-65) or very high (65-80). CLASS stands for Classroom Assessment Scoring System. CLASS domains can be scored as low (scores 1-2), medium (3-5) or high (6-7). Standard errors in parentheses. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 5: Randomization Test

	Test-screened tenured teachers
Children:	
Age (months)	-0.000
	(0.000)
Gender	0.001
	(0.003)
TVIP	0.000
	(0.000)
Proportion who attended preschool	-0.005
	(0.005)
Family:	
Parents' years of schooling	0.001**
	(0.001)
Living standard indicator	-0.000
	(0.003)
Observations	12632
R^2	0.580
F	1.09
p	0.372

Note: OLS model estimated with cluster standard errors (in parentheses) at the school level and school fixed effects. Total number of observations restricted to student sample with full information on student, parent, classroom and teacher characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 6: No-Show, Attrition and Late Enrollment Tests

	No-shows		Attritors		Late enrollments	
Teacher	(1)	(2)	(3)	(4)	(5)	(6)
Test-screened tenured	-0.005	-0.007	-0.004	-0.003	-0.004	0.000
	(0.007)	(0.007)	(0.007)	(0.007)	(0.013)	(0.013)
Female		-0.020		0.006		-0.019
		(0.019)		(0.019)		(0.023)
Years of experience		0.000		-0.000		0.001**
		(0.000)		(0.000)		(0.000)
Years of education		0.001		-0.001		-0.001
		(0.002)		(0.001)		(0.001)
Cognitive skills		-0.002		-0.000		-0.002
		(0.003)		(0.003)		(0.003)
Neuroticism		-0.002		-0.001		0.001
		(0.003)		(0.002)		(0.003)
Extraversion		0.002		0.002		-0.005
		(0.003)		(0.002)		(0.003)
Openness		0.002		-0.001		0.003
		(0.003)		(0.002)		(0.004)
Agreeableness		0.005		-0.000		0.001
		(0.003)		(0.002)		(0.004)
Conscientiousness		0.002		0.002		0.001
		(0.003)		(0.002)		(0.003)
CLASS mean 2011-12		-0.001		0.007***		-0.000
		(0.003)		(0.003)		(0.004)
Observations	14909	14235	14170	13543	14495	13855
R^2	0.045	0.044	0.033	0.034	0.045	0.048
F	0.55	1.32	0.42	1.46	0.11	1.11
p	0.459	0.215	0.519	0.150	0.743	0.356

Note: OLS linear probability models estimated with cluster standard errors (in parentheses) at the school level and school fixed effects. No-shows is a dummy variable that takes the value of 1 if an enrolled student did not show up in the beginning of the school year. Attritors is a dummy variable that takes the value of 1 if the student dropped out of the school. Late enrollment is a dummy variable that takes the value of 1 if the student enrolled after the school year started. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 7: Estimates of Effects of Test-Screened Tenured Teachers on Language

Teacher	Language						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Test-screened tenured	0.105**	0.102***	0.096**	0.089**	0.115***	0.115***	0.125***
	(0.043)	(0.038)	(0.038)	(0.039)	(0.037)	(0.038)	(0.035)
Female teacher					0.245**	0.211*	0.162
					(0.113)	(0.108)	(0.101)
Years of experience					0.005**	0.004**	0.003*
					(0.002)	(0.002)	(0.002)
Years of education					0.001	-0.001	-0.002
					(0.007)	(0.007)	(0.007)
Cognitive skills						0.038***	0.037**
						(0.014)	(0.014)
Neuroticism						-0.003	0.001
						(0.013)	(0.012)
Extraversion						-0.001	-0.012
						(0.014)	(0.014)
Openness						0.006	0.006
						(0.015)	(0.015)
Agreeableness						-0.014	-0.009
						(0.018)	(0.017)
Conscientiousness						-0.005	-0.004
						(0.015)	(0.015)
CLASS average 2011-12							0.045***
							(0.015)
School fixed effects	YES	YES	YES	YES	YES	YES	YES
Student controls	NO	YES	YES	YES	YES	YES	YES
Parent controls	NO	NO	YES	YES	YES	YES	YES
Classroom controls	NO	NO	NO	YES	YES	YES	YES
Observations	12632	12632	12632	12632	12632	12632	12632
R^2	0.149	0.433	0.440	0.440	0.441	0.442	0.442

Notes: Each column reports coefficients from OLS regressions estimated with cluster standard errors (in parentheses) at the school level. Columns (2)-(7) control for the following student characteristics: TVIP score, age, gender, attendance to preschool; parent characteristics: years of education and living standard conditions; classroom characteristics: class size, classroom averages of student and parent characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 8: Estimates of Effects of Test-Screened Tenured Teachers on Math

Teacher	Math						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Test-screened tenured	0.085*	0.086*	0.080*	0.068	0.093**	0.086*	0.099**
	(0.048)	(0.045)	(0.045)	(0.046)	(0.046)	(0.045)	(0.044)
Female teacher					0.097	0.092	0.028
					(0.236)	(0.241)	(0.233)
Years of experience					0.005***	0.006***	0.004**
					(0.002)	(0.002)	(0.002)
Years of education					-0.004	-0.004	-0.005
					(0.007)	(0.008)	(0.007)
Cognitive skills						0.014	0.011
						(0.015)	(0.016)
Neuroticism						0.006	0.012
						(0.015)	(0.014)
Extraversion						0.022*	0.008
						(0.013)	(0.014)
Openness						0.013	0.013
						(0.018)	(0.018)
Agreeableness						-0.010	-0.003
						(0.018)	(0.017)
Conscientiousness						-0.018	-0.017
						(0.018)	(0.018)
CLASS average 2011-12							0.059***
							(0.019)
School fixed effects	YES	YES	YES	YES	YES	YES	YES
Student controls	NO	YES	YES	YES	YES	YES	YES
Parent controls	NO	NO	YES	YES	YES	YES	YES
Classroom controls	NO	NO	NO	YES	YES	YES	YES
Observations	12632	12632	12632	12632	12632	12632	12632
R^2	0.123	0.303	0.309	0.309	0.310	0.310	0.311

Notes: Each column reports coefficients from OLS regressions estimated with cluster standard errors (in parentheses) at the school level. Columns (2)-(7) control for the following student characteristics: TVIP score, age, gender, attendance to preschool; parent characteristics: years of education and living standard conditions; classroom characteristics: class size, classroom averages of student and parent characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 9: Estimates of Effects of Test-Screened Tenured, Other-Tenured, Test-Screened Contract vs. Contract Teachers on Language

Teacher	Language						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Test-screened tenured	0.169***	0.166***	0.159***	0.155***	0.157***	0.143***	0.148***
	(0.050)	(0.044)	(0.045)	(0.045)	(0.044)	(0.045)	(0.043)
Other-tenured	0.097**	0.093***	0.091**	0.098***	0.078**	0.058*	0.047
	(0.040)	(0.035)	(0.035)	(0.035)	(0.035)	(0.034)	(0.035)
Test-screened contract	-0.012	0.040	0.044	0.032	0.030	-0.013	-0.005
	(0.061)	(0.054)	(0.054)	(0.056)	(0.055)	(0.060)	(0.061)
Female teacher					0.248**	0.218**	0.171*
					(0.103)	(0.102)	(0.097)
Years of experience					0.002	0.002	0.002
					(0.002)	(0.002)	(0.002)
Years of education					-0.000	-0.001	-0.002
					(0.007)	(0.007)	(0.007)
Cognitive skills						0.038***	0.036**
						(0.014)	(0.014)
Neuroticism						-0.005	-0.001
						(0.013)	(0.013)
Extraversion						0.001	-0.010
						(0.015)	(0.015)
Openness						0.003	0.004
						(0.015)	(0.015)
Agreeableness						-0.013	-0.008
						(0.018)	(0.017)
Conscientiousness						-0.004	-0.003
						(0.015)	(0.015)
CLASS average 2011-12							0.042***
							(0.016)
School fixed effects	YES	YES	YES	YES	YES	YES	YES
Student controls	NO	YES	YES	YES	YES	YES	YES
Parent controls	NO	NO	YES	YES	YES	YES	YES
Classroom controls	NO	NO	NO	YES	YES	YES	YES
Observations	12632	12632	12632	12632	12632	12632	12632
R^2	0.150	0.433	0.441	0.441	0.441	0.442	0.442

Notes: Each column reports coefficients from OLS regressions estimated with cluster standard errors (in parentheses) at the school level. Columns (2)-(7) control for the following student characteristics: TVIP score, age, gender, attendance to preschool; parent characteristics: years of education and living standard conditions; classroom characteristics: class size, classroom averages of student and parent characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 10: Estimates of Effects of Test-Screened Tenured, Other-Tenured, Test-Screened Contract vs. Contract Teachers on Math

Teacher	Math						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Test-screened tenured	0.155***	0.154***	0.147***	0.136**	0.137**	0.123**	0.129**
	(0.057)	(0.052)	(0.052)	(0.052)	(0.053)	(0.052)	(0.051)
Other-tenured	0.105**	0.100***	0.098***	0.100***	0.080**	0.068*	0.054
	(0.040)	(0.036)	(0.036)	(0.036)	(0.038)	(0.038)	(0.038)
Test-screened contract	0.008	0.047	0.049	0.043	0.042	0.020	0.030
	(0.061)	(0.053)	(0.053)	(0.055)	(0.054)	(0.058)	(0.059)
Female teacher					0.101	0.100	0.037
					(0.222)	(0.231)	(0.225)
Years of experience					0.003	0.003*	0.002
					(0.002)	(0.002)	(0.002)
Years of education					-0.005	-0.004	-0.006
					(0.007)	(0.008)	(0.007)
Cognitive skills						0.010	0.007
						(0.015)	(0.016)
Neuroticism						0.004	0.010
						(0.015)	(0.014)
Extraversion						0.022	0.008
						(0.014)	(0.014)
Openness						0.010	0.011
						(0.018)	(0.017)
Agreeableness						-0.008	-0.002
						(0.017)	(0.016)
Conscientiousness						-0.017	-0.016
						(0.018)	(0.018)
CLASS average 2011-12							0.057***
							(0.019)
School fixed effects	YES	YES	YES	YES	YES	YES	YES
Student controls	NO	YES	YES	YES	YES	YES	YES
Parent controls	NO	NO	YES	YES	YES	YES	YES
Classroom controls	NO	NO	NO	YES	YES	YES	YES
Observations	12632	12632	12632	12632	12632	12632	12632
R^2	0.124	0.303	0.309	0.310	0.310	0.311	0.312

Notes: Each column reports coefficients from OLS regressions estimated with cluster standard errors (in parentheses) at the school level. Columns (2)-(7) control for the following student characteristics: TVIP score, age, gender, attendance to preschool; parent characteristics: years of education and living standard conditions; classroom characteristics: class size, classroom averages of student and parent characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 11: Estimates of Effects of Test-screened Tenured Teachers on Language, Accounting for Ministerial Resolution Competitions

Teacher	Language						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Test-screened tenured AM 438-07	0.160***	0.193***	0.193***	0.185***	0.193***	0.184***	0.185***
	(0.061)	(0.069)	(0.069)	(0.069)	(0.067)	(0.068)	(0.064)
Test-screened tenured AM 018-10	-0.105	-0.030	-0.055	-0.019	-0.058	-0.040	-0.001
	(0.234)	(0.177)	(0.167)	(0.180)	(0.249)	(0.255)	(0.237)
Test-screened tenured AM 379-11	0.090	0.068	0.060	0.051	0.088*	0.090*	0.102**
	(0.056)	(0.044)	(0.043)	(0.046)	(0.045)	(0.046)	(0.043)
Female teacher					0.263*	0.228*	0.176
					(0.135)	(0.128)	(0.119)
Years of experience					0.005**	0.004**	0.003*
					(0.002)	(0.002)	(0.002)
Years of education					0.001	-0.001	-0.002
					(0.007)	(0.007)	(0.007)
Cognitive skills						0.037***	0.036**
						(0.014)	(0.014)
Neuroticism						-0.003	0.002
						(0.013)	(0.013)
Extraversion						-0.001	-0.012
						(0.014)	(0.014)
Openness						0.008	0.008
						(0.016)	(0.015)
Agreeableness						-0.012	-0.007
						(0.018)	(0.017)
Conscientiousness						-0.005	-0.005
						(0.015)	(0.015)
CLASS average 2011-12							0.044***
							(0.015)
School fixed effects	YES	YES	YES	YES	YES	YES	YES
Student controls	NO	YES	YES	YES	YES	YES	YES
Parent controls	NO	NO	YES	YES	YES	YES	YES
Classroom controls	NO	NO	NO	YES	YES	YES	YES
Observations	12632	12632	12632	12632	12632	12632	12632
R^2	0.149	0.433	0.440	0.440	0.441	0.442	0.442

Notes: Each column reports coefficients from OLS regressions estimated with cluster standard errors (in parentheses) at the school level. Columns (2)-(7) control for the following student characteristics: TVIP score, age, gender, attendance to preschool; parent characteristics: years of education and living standard conditions; classroom characteristics: class size, classroom averages of student and parent characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 12: Estimates of Effects of Test-Screened Tenured Teachers on Math, Accounting for Ministerial Resolution Competitions

Teacher	Math						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Test-screened tenured AM 438-07	0.172***	0.198***	0.197***	0.183***	0.192***	0.182**	0.183**
	(0.059)	(0.060)	(0.059)	(0.062)	(0.074)	(0.074)	(0.074)
Test-screened tenured AM 018-10	-0.155	-0.094	-0.118	-0.093	-0.093	-0.114	-0.062
	(0.262)	(0.219)	(0.209)	(0.210)	(0.253)	(0.258)	(0.236)
Test-screened tenured AM 379-11	0.058	0.045	0.037	0.025	0.057	0.051	0.067
	(0.063)	(0.058)	(0.058)	(0.059)	(0.060)	(0.059)	(0.057)
Female teacher					0.115	0.113	0.045
					(0.249)	(0.255)	(0.245)
Years of experience					0.005**	0.005***	0.004**
					(0.002)	(0.002)	(0.002)
Years of education					-0.004	-0.004	-0.005
					(0.007)	(0.007)	(0.007)
Cognitive skills						0.012	0.010
						(0.015)	(0.016)
Neuroticism						0.007	0.013
						(0.015)	(0.014)
Extraversion						0.022	0.008
						(0.013)	(0.014)
Openness						0.015	0.015
						(0.018)	(0.018)
Agreeableness						-0.008	-0.001
						(0.018)	(0.017)
Conscientiousness						-0.019	-0.018
						(0.018)	(0.018)
CLASS average 2011-12							0.058***
							(0.019)
School fixed effects	YES	YES	YES	YES	YES	YES	YES
Student controls	NO	YES	YES	YES	YES	YES	YES
Parent controls	NO	NO	YES	YES	YES	YES	YES
Classroom controls	NO	NO	NO	YES	YES	YES	YES
Observations	12632	12632	12632	12632	12632	12632	12632
R^2	0.123	0.303	0.309	0.309	0.310	0.310	0.312

Notes: Each column reports coefficients from OLS regressions estimated with cluster standard errors (in parentheses) at the school level. Columns (2)-(7) control for the following student characteristics: TVIP score, age, gender, attendance to preschool; parent characteristics: years of education and living standard conditions; classroom characteristics: class size, classroom averages of student and parent characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 13: Estimates of Effects of Test-Screened Tenured Teachers on Language by TVIP Quintiles (Heterogeneity)

	Language									
Teacher	TVIP Q1		TVIP Q2		TVIP Q3		TVIP Q4		TVIP Q5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Test-screened tenured	0.240***	0.214***	0.076	0.121*	0.100	0.168**	0.109	0.184**	-0.048	-0.034
	(0.078)	(0.070)	(0.083)	(0.070)	(0.067)	(0.070)	(0.068)	(0.077)	(0.086)	(0.083)
Female		0.068		0.277**		0.144		0.417**		-0.055
		(0.100)		(0.127)		(0.226)		(0.207)		(0.087)
Years of experience		0.005		-0.001		0.004		0.004		0.003
		(0.004)		(0.003)		(0.003)		(0.003)		(0.003)
Years of education		0.001		0.010		0.006		-0.012		-0.004
		(0.015)		(0.013)		(0.012)		(0.011)		(0.011)
Cognitive skills		0.030		0.041*		0.009		0.026		0.065***
		(0.032)		(0.022)		(0.024)		(0.026)		(0.023)
Neuroticism		0.001		-0.004		0.009		0.032		-0.015
		(0.027)		(0.021)		(0.024)		(0.022)		(0.026)
Extraversion		0.014		-0.015		-0.021		-0.051**		0.027
		(0.028)		(0.024)		(0.029)		(0.021)		(0.027)
Openness		-0.020		0.002		0.046*		0.008		0.001
		(0.034)		(0.023)		(0.025)		(0.029)		(0.028)
Agreeableness		-0.043		-0.010		-0.026		0.034		-0.013
		(0.030)		(0.028)		(0.027)		(0.031)		(0.032)
Conscientiousness		0.027		-0.049**		0.012		0.002		0.007
		(0.031)		(0.025)		(0.026)		(0.024)		(0.030)
CLASS mean 2011-12		0.025		0.033		0.026		0.101***		0.067**
		(0.029)		(0.029)		(0.028)		(0.028)		(0.026)
School fixed effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Student controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Parent controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Classroom controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	2544	2544	2793	2793	2318	2318	2550	2550	2427	2427
R^2	0.189	0.263	0.186	0.299	0.220	0.295	0.175	0.269	0.190	0.313

Notes: Each column reports coefficients from OLS regressions estimated with cluster standard errors (in parentheses) at the school level. Columns (2), (4), (6), (8) and (10) control for the following student characteristics: TVIP score, age, gender, attendance to preschool; parent characteristics: years of education and living standard conditions; classroom characteristics: class size, classroom averages of student and parent characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 14: Estimates of Effects of Test-Screened Tenured Teachers on Math by TVIP Quintiles (Heterogeneity)

Teacher	Math									
	TVIP Q1		TVIP Q2		TVIP Q3		TVIP Q4		TVIP Q5	
	1	2	3	4	5	6	7	8	9	10
Test-screened tenured	0.065	0.046	0.112	0.127*	0.179**	0.272***	0.164**	0.218***	-0.052	-0.041
	(0.056)	(0.053)	(0.082)	(0.074)	(0.086)	(0.095)	(0.080)	(0.081)	(0.103)	(0.102)
Proportion female		0.020		0.272**		0.068		-0.015		-0.156
		(0.279)		(0.129)		(0.473)		(0.334)		(0.286)
Years of experience		0.007**		0.001		0.010**		0.001		0.002
		(0.003)		(0.004)		(0.004)		(0.003)		(0.004)
Years of education		-0.009		-0.001		0.007		0.001		-0.011
		(0.011)		(0.014)		(0.014)		(0.010)		(0.016)
IQ		0.012		0.011		0.012		0.007		0.010
		(0.026)		(0.023)		(0.032)		(0.027)		(0.033)
Neuroticism		-0.017		0.027		0.056*		0.040		-0.007
		(0.025)		(0.031)		(0.029)		(0.026)		(0.031)
Extraversion		0.014		0.020		0.037		-0.053*		0.033
		(0.025)		(0.025)		(0.027)		(0.027)		(0.035)
Openness		-0.030		0.013		0.041		0.035		0.023
		(0.034)		(0.029)		(0.036)		(0.028)		(0.038)
Agreeableness		-0.027		-0.002		-0.031		0.031		0.007
		(0.029)		(0.029)		(0.035)		(0.033)		(0.038)
Conscientiousness		-0.004		-0.035		-0.014		-0.035		-0.006
		(0.028)		(0.025)		(0.034)		(0.030)		(0.038)
CLASS mean 2011-12		0.021		0.041		0.078**		0.128***		0.059
		(0.030)		(0.033)		(0.039)		(0.034)		(0.037)
School fixed effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Student controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Parent controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Classroom controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	2544	2544	2793	2793	2318	2318	2550	2550	2427	2427
R^2	0.189	0.227	0.158	0.223	0.183	0.255	0.179	0.257	0.172	0.268

Notes: Each column reports coefficients from OLS regressions estimated with cluster standard errors (in parentheses) at the school level. Columns (2), (4), (6), (8) and (10) control for the following student characteristics: TVIP score, age, gender, attendance to preschool; parent characteristics: years of education and living standard conditions; classroom characteristics: class size, classroom averages of student and parent characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 15: Estimates of Effects of Test-Screened Tenured Teachers on Language by Household Living Standard Indicator Quintiles (Heterogeneity)

Teacher	Language							
	LSI Q1		LSI Q2		LSI Q3		LSI Q4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test-screened tenured	0.142*	0.200***	0.038	0.060	0.049	0.086	0.155*	0.152**
	(0.077)	(0.066)	(0.097)	(0.079)	(0.073)	(0.064)	(0.084)	(0.075)
Female teacher		0.202		0.196*		0.063		0.413**
		(0.234)		(0.104)		(0.177)		(0.187)
Years of experience		0.009***		0.000		0.004		0.000
		(0.003)		(0.003)		(0.003)		(0.003)
Years of education		0.007		0.010		-0.010		-0.017
		(0.013)		(0.011)		(0.009)		(0.016)
Cognitive skills		0.034		0.070***		0.030		-0.000
		(0.024)		(0.023)		(0.022)		(0.027)
Neuroticism		0.001		-0.013		0.005		0.032
		(0.017)		(0.021)		(0.023)		(0.026)
Extraversion		-0.010		0.013		-0.012		-0.021
		(0.021)		(0.021)		(0.025)		(0.029)
Openness		-0.017		-0.017		0.025		0.044*
		(0.025)		(0.025)		(0.028)		(0.026)
Agreeableness		-0.038		-0.013		-0.014		0.039
		(0.023)		(0.028)		(0.030)		(0.035)
Conscientiousness		0.031		-0.002		-0.020		-0.036
		(0.026)		(0.024)		(0.023)		(0.029)
CLASS mean 2011-12		0.066***		0.041**		0.011		0.062**
		(0.023)		(0.021)		(0.027)		(0.030)
School fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Student controls	NO	YES	NO	YES	NO	YES	NO	YES
Parent controls	NO	YES	NO	YES	NO	YES	NO	YES
Classroom controls	NO	YES	NO	YES	NO	YES	NO	YES
Observations	3537	3537	3268	3268	3403	3403	2424	2424
R^2	0.152	0.450	0.162	0.469	0.193	0.453	0.191	0.461

Notes: Each column reports coefficients from OLS regressions estimated with cluster standard errors (in parentheses) at the school level. Columns (2), (4), (6) and (8) control for the following student characteristics: TVIP score, age, gender, attendance to preschool; parent characteristics: years of education and living standard conditions; classroom characteristics: class size, classroom averages of student and parent characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 16: Estimates of Effects of Test-Screened Tenured Teachers on Math by Household Living Standard Indicator Quintiles (Heterogeneity)

Teacher	Math							
	LSI Q1		LSI Q2		LSI Q3		LSI Q4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test-screened tenured	0.056	0.081	0.065	0.077	0.060	0.072	0.117	0.135
	(0.086)	(0.077)	(0.074)	(0.071)	(0.065)	(0.056)	(0.127)	(0.097)
Female teacher		-0.157		0.210		0.162		0.057
		(0.259)		(0.317)		(0.310)		(0.277)
Years of experience		0.007**		0.001		0.002		0.003
		(0.003)		(0.004)		(0.004)		(0.004)
Years of education		-0.005		0.010		-0.004		-0.020
		(0.011)		(0.012)		(0.010)		(0.018)
Cognitive skills		0.006		0.048*		-0.013		0.005
		(0.021)		(0.026)		(0.026)		(0.036)
Neuroticism		-0.008		0.015		0.046**		-0.009
		(0.023)		(0.025)		(0.023)		(0.035)
Extraversion		-0.017		0.015		0.028		0.025
		(0.023)		(0.021)		(0.022)		(0.034)
Openness		0.009		0.008		0.029		0.010
		(0.024)		(0.027)		(0.032)		(0.032)
Agreeableness		-0.005		-0.012		0.009		0.018
		(0.022)		(0.033)		(0.030)		(0.042)
Conscientiousness		-0.005		-0.008		-0.038		-0.034
		(0.028)		(0.027)		(0.024)		(0.038)
CLASS mean 2011-12		0.066***		0.059**		0.038		0.075*
		(0.025)		(0.023)		(0.033)		(0.044)
School fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Student controls	NO	YES	NO	YES	NO	YES	NO	YES
Parent controls	NO	YES	NO	YES	NO	YES	NO	YES
Classroom controls	NO	YES	NO	YES	NO	YES	NO	YES
Observations	3537	3537	3268	3268	3403	3403	2424	2424
R^2	0.163	0.349	0.144	0.346	0.143	0.317	0.169	0.345

Notes: Each column reports coefficients from OLS regressions estimated with cluster standard errors (in parentheses) at the school level. Columns (2), (4), (6) and (8) control for the following student characteristics: TVIP score, age, gender, attendance to preschool; parent characteristics: years of education and living standard conditions; classroom characteristics: class size, classroom averages of student and parent characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 17: Randomization Test, School Subsample (Robustness Check)

	Test-screened tenured teachers
Children:	
Age (months)	-0.002
	(0.002)
Proportion female	0.004
	(0.018)
TVIP	0.000
	(0.001)
Proportion who attended preschool	-0.031
	(0.026)
Family:	
Parents' years of schooling	0.006*
	(0.003)
Living standard indicator	-0.003
	(0.015)
Observations	2393
R^2	0.057
F	1.15
p	0.353

Note: Subsample of 36 schools with treated and control classrooms. OLS model estimated with cluster standard errors (in parentheses) at the school level and school fixed effects. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.

Table 18: No-Show, Attrition and Late Enrollment Tests (Robustness Check)

Teacher	No-shows		Attritors		Late enrollments	
	(1)	(2)	(3)	(4)	(5)	(6)
Test-screened tenured	-0.006	-0.005	-0.005	-0.005	-0.006	0.001
	(0.007)	(0.008)	(0.007)	(0.007)	(0.013)	(0.016)
Female		-0.077***		0.009		-0.105***
		(0.019)		(0.015)		(0.026)
Years of experience		-0.000		0.000		0.001
		(0.001)		(0.001)		(0.001)
Years of education		-0.002		0.002		-0.002
		(0.003)		(0.002)		(0.003)
Cognitive skills		0.007		0.005		-0.006
		(0.005)		(0.006)		(0.005)
Neuroticism		0.001		0.001		-0.003
		(0.004)		(0.003)		(0.005)
Extraversion		-0.000		0.006		-0.011
		(0.005)		(0.005)		(0.007)
Openness		0.006		-0.002		-0.007
		(0.005)		(0.005)		(0.009)
Agreeableness		-0.015***		-0.000		-0.006
		(0.004)		(0.005)		(0.013)
Conscientiousness		0.004		0.002		0.016*
		(0.006)		(0.005)		(0.009)
CLASS mean 2011-12		0.002		0.002		-0.006
		(0.004)		(0.005)		(0.009)
Observations	2688	2654	2589	2555	2681	2645
R^2	0.024	0.027	0.044	0.046	0.037	0.043
F	0.65	47.29	0.45	2.64	0.19	112.27
p	0.424	0.000	0.506	0.016	0.667	0.000

Table 19: Estimates of Effects of Test-Tenured Teachers, School Subsample (Robustness Check)

Teacher	Language				Math			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test-screened tenured	0.102**	0.137***	0.093***	0.102***	0.089*	0.107**	0.094**	0.105*
	(0.045)	(0.042)	(0.029)	(0.034)	(0.050)	(0.052)	(0.044)	(0.053)
Female		0.161***	0.357***	0.319***		0.104	0.362***	0.316**
		(0.055)	(0.086)	(0.102)		(0.089)	(0.127)	(0.156)
Years of experience		0.005**	0.008***	0.007***		-0.002	0.004	0.003
		(0.002)	(0.003)	(0.003)		(0.004)	(0.004)	(0.004)
Years of education		-0.005	0.002	0.002		-0.023*	-0.003	-0.003
		(0.011)	(0.011)	(0.011)		(0.013)	(0.013)	(0.012)
Cognitive skills			-0.010	-0.005			-0.075**	-0.069*
			(0.025)	(0.029)			(0.035)	(0.037)
Neuroticism			-0.012	-0.011			0.042	0.044
			(0.022)	(0.021)			(0.033)	(0.031)
Extraversion			0.072***	0.061**			0.074**	0.061
			(0.017)	(0.027)			(0.030)	(0.038)
Openness			0.037**	0.030			0.021	0.012
			(0.017)	(0.020)			(0.035)	(0.040)
Agreeableness			0.060**	0.054*			0.068*	0.061
			(0.028)	(0.029)			(0.036)	(0.039)
Conscientiousness			-0.072**	-0.062*			-0.087*	-0.075
			(0.027)	(0.034)			(0.044)	(0.051)
CLASS mean 2011-12				0.025				0.030
				(0.033)				(0.043)
School fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Student controls	NO	YES	YES	YES	NO	YES	YES	YES
Parent controls	NO	YES	YES	YES	NO	YES	YES	YES
Classroom controls	NO	YES	YES	YES	NO	YES	YES	YES
Observations	2393	2393	2393	2393	2393	2393	2393	2393
R^2	0.107	0.426	0.430	0.431	0.045	0.262	0.267	0.267

Notes: Subsample of 36 schools with treated and control classrooms. Each column reports coefficients from OLS regressions estimated with cluster standard errors (in parentheses) at the school level. Columns (2)-(4) and (6)-(8) control for the following student characteristics: TVIP score, age, gender, attendance to preschool; parent characteristics: years of education and living standard conditions; classroom characteristics: class size, classroom averages of student and parent characteristics. * Significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level.