Economics of Education Review 98 (2024) 102489

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

The impact of incentivizing training on students' outcomes ☆

Tatiana Paredes, Almudena Sevilla *

University College London, United Kingdom

ARTICLE INFO

Keywords:
Professional development
Teachers
Developing countries
Regression discontinuity design

ABSTRACT

This paper studies the effect on students' scores of incentivizing in-service teacher training in a system that conditions teacher promotions to in-service training take-up. In Ecuador, teachers need to pass a compulsory knowledge test with a minimum score and undergo substantial training to qualify for a promotion. We use a regression discontinuity design to identify the causal effect of incentivizing in-service teacher training on students' scores on a standardized national university entrance exam. We find that in-service training significantly improves students' verbal test scores by 0.19 to 0.31 standard deviations (depending on the selected comparison window).

1. Introduction

Teacher quality is one of the most important inputs in the production of student achievement (Bau & Das, 2020; Jackson, 2018; Liu & Loeb, 2019; Rivkin et al., 2005; Rockoff, 2004). For the US, Friedman et al. (2014) find that replacing a teacher whose impact on student value added is in the bottom 5% of the distribution with an average teacher would increase the present value of students' lifetime earnings by $250,000 per classroom. Two tools for improving teacher quality that have been used separately and together are professional development and teacher incentives (Popova et al., 2021). While it is clear that effective teacher education should improve student learning, the effects of teacher incentives are less clear. The reason is that, in most cases, governments reward teachers on the basis of the teacher's production function outcomes rather than effort (which is not verifiable). In these cases, two problems arise: distortion, where measurable outcomes are over-incentivized while others are ignored, and noise, where outcomes are a noisy function of teacher effort (Baker, 2002). This paper provides novel evidence from a program that, unlike traditional professional development programs, incentivizes measures of teacher effort (performance on a teacher assessment and on a PD course) and shows that PD linked to incentives can be an effective way to increase student learning.

We study the effect on students' scores of incentivizing in-service teacher training in a system that conditions teacher promotions to in-service training take-up. In Ecuador, teachers need to pass a compulsory knowledge test with a minimum score and undergo substantial training to qualify for a promotion. We use the discontinuity generated by the knowledge test threshold to identify the causal effect of incentivizing in-service teacher training on students' scores on a standardized national university entrance exam.

Our identification strategy hinges on the similarity of teachers on either side of the threshold in terms of both observed and unobserved characteristics. Teachers in the treatment and control groups have the same opportunities to receive training, which is provided for free by the Ministry of Education. There are no rules to prioritize teachers who passed the nationwide teacher evaluation when these are oversubscribed. The only difference is that teachers in the treated group who pass the threshold are automatically eligible for promotions. Thus, at the margin, teachers in the treated group have a greater incentive to take-up more training and excel in the training criteria to maximize the probability of being promoted.[1]

Our analysis is based on a unique matched teacher–student dataset connecting public school seniors who took the mandatory university entrance exam between 2016 and 2019 with their language teachers. We observe teachers' scores on the 2016 structured knowledge test as

☆ The project PARENTIME has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 770839). We thank the Ministry of Education of Ecuador for access to the data used in this paper. We also thank the attendees of STICERD's Work in Progress Seminar at the London School of Economics and of the PARENTIME workshop at University College London for valuable comments, Nicolás Grau and Javier Brugués for valuable feedback.

* Corresponding author.
E-mail addresses: t.paredes@ucl.ac.uk (T. Paredes), a.sevilla@lse.ac.uk (A. Sevilla).

[1] We lack individual-level data on training attendance. Moreover, even with details on total training hours and a discernible spike at the cut-off, assessing the threshold is ambiguous due to the vast differences in course quality and content available to teachers. Consequently, it is challenging to assert that increased hours in potentially irrelevant or subpar training directly improve student outcomes.

https://doi.org/10.1016/j.econedurev.2023.102489
Received 19 June 2022; Received in revised form 19 October 2023; Accepted 29 October 2023
Available online 27 November 2023

well as students' verbal scores on the 2016–2019 university entrance exam. We also observe teachers' career paths and their scores on the nationwide Curriculum Refresher Course offered to teachers in 2017. Our sample includes students and teachers in predominantly rural and smaller-than-average schools in Ecuador. The latter is because in order to assign the correct treatment status to each pair composed of a teacher and her students, we focus on schools with only one language teacher teaching in the last year of secondary school.

We use an RD design to estimate the effect of a teacher passing the 2016 structured knowledge test on their students' scores on the verbal section of the university entrance exam (2017–2019). When we condition on certain student characteristics, including the pre-treatment verbal scores of the previous cohort, we find a positive and significant impact ranging from 0.19 to 0.31 standard deviations. These results are consistent across several comparison windows (40, 50 and 60 points), although the precision of the estimates is compromised due to small sample sizes in some specifications. The main mechanism we explore is changes in teachers' effort at the threshold. We find that teachers who score above the threshold on the 2016 knowledge test are between 11 and 19 percentage points more likely to pass the 2017 Curriculum Refresher Course, depending on the window considered. We observe a similar pattern of results in terms of the score that teachers obtained in the 2017 Curricular Refresher Course.

We also present additional pieces of evidence that are consistent with the fact that the additional training acquired by teachers in the treated group leads to improvements in student scores. We find that teachers in the treated group who pass the threshold are more likely to pass a compulsory 100-hour curriculum refresher course. The Curricular Refresher course is part of the 330 h of training that all teachers must pass regardless of their performance in the knowledge test, which eliminates concerns about endogeneity in take-up. We also show evidence that teachers who pass the 2016 structured knowledge test are about 30 and 40 percentage points more likely to get a promotion in the following years (2017–2019) and that this effect is largest in the 40-point comparison window and decreases slightly as the window becomes larger. Finally, we analyze the dynamics of the impact on student scores and find that effects grow from 2017 to 2019, which is consistent with the fact that teachers should gradually accumulate the required number of in-service training hours during the years following the teacher evaluation to get a promotion. All in all, these findings suggest that students benefited from the incremental learning undertaken by teachers in the treated group, which translated into enhanced student outcomes in the university entrance examination.

Our paper contributes to a growing literature that seeks to causally estimate the effect of teachers' quality on student outcomes. Much of this literature has focused on improving the quality of teaching by providing test-score-based incentives to teachers (Behrman et al., 2015; Brown & Andrabi, 2021; Duflo et al., 2012; Fryer, 2013; Glewwe et al., 2010; Karachiwalla & Park, 2016; Muralidharan & Sundararaman, 2011). However, test-score-based incentives may risk distorting teachers' incentives, which occur when teachers teach to the test or ignore other outcomes that may be important for students' long-term learning and development. In-service teacher training programs overcome this shortcoming. However, identifying the causal effect of such programs is difficult in real-world settings because of the challenges associated with teachers' and schools' self-selection for training based on unobservable characteristics correlated with student outcomes but unobserved by the researcher. A few small-scale RCTs and quasi-experimental studies aim to circumvent this identification problem, but the evidence so far is mixed.[2] Here we leverage a large administrative data set of students and teachers and the quasi-experimental nature of the program to credibly estimate the causal effect of the training program while addressing external validity concerns present in small RCT studies by providing quasi-experimental estimates from a nationwide teachers' training program with a high participation rate. This helps enhance the generalizability of the findings and ensures that they can be applied to a broader context.

We also add to the literature by informing on the mechanisms at play identified in the theoretical and empirical literature. A recent survey review shows that whereas in-service training for teachers is most effective when participation is linked to promotion or salary increases, a large share of at-scale training programs analyzed in the literature does not link participation to salary or career progression programs. In these cases where participation has no implications for promotion, salary, or status increases, student learning is 0.12 standard deviation lower (Popova et al., 2021).[3] Similarly, Loyalka et al. (2019) acknowledge that the failure of the training program analyzed in their study may be due to the lack of fidelity in implementation because the program lacked incentives for teachers to participate in the program. We add to these studies by providing evidence of the importance of incentivizing teacher training through promotions. In particular, we show that linking participation in the program to salary increases or career progression leads to higher and more sustainable student learning outcomes without distorting teachers' incentives.

For instance, the studies by Garet et al. (2008) and Randel et al. (2011) find no significant effects on student achievement in mathematics or standardized reading scores. Similarly, Jacob and Lefgren (2004) find no statistically significant effects on students' reading and math scores. On the other hand, the studies that have found a positive impact of teacher training on students' test scores include the randomized cluster design by Borman et al. (2007) and a matched comparison design by Angrist and Lavy (2001). Borman et al. (2007) find positive effects of 0.24 and 0.36 standard deviations on word identification and passage comprehension, respectively; and Angrist and Lavy (2001) report positive effects on mathematics and reading test scores of one-half standard deviation, respectively.

The rest of this paper is organized as follows. Section 2 describes the institutional framework with an emphasis on the promotion process for public sector teachers in Ecuador. Section 3 discusses the various administrative data sources used to build the matched teacher–student sample used in this paper. Section 4 discusses our identification strategy and its validity. Section 5 reports the results of this paper, Section 6 discusses the main mechanisms behind the observed effects; Section 7 concludes.

2. Background

2.1. Career progression and PD in Ecuador

In developing countries, many teachers lack the necessary knowledge and skills to improve student achievement (Azam & Kingdon, 2015; Bruns & Luque, 2015; De Talancé, 2017). This is also the case in Ecuador where 51% of 15-year-olds do not reach the minimum level of proficiency in reading (INEVAL, 2018). With the aim of raising teaching quality, regular nationwide teacher evaluations were instituted as the first step for teachers' career progression. Teachers who wish to be promoted need to obtain a minimum score in a mandatory structured knowledge test, undergo substantial in-service training and be four years in their current earning category to qualify.[4] In 2016, the National

[2] Andrabi and Brown (2020), Angrist and Lavy (2001), Behrman et al. (2015), Borman et al. (2007), Duflo et al. (2012), Fryer (2013), Garet et al. (2008), Glewwe et al. (2010), Jacob and Lefgren (2004), Muralidharan and Sundararaman (2011) and Randel et al. (2011).

[3] Muralidharan and Sundararaman (2011) also report that while education and training alone are not significant predictors of value-added, the interaction of teacher education and training with incentives is a positive and significant determinant of value added.

[4] Source: Law of Intercultural Education "Ley Organica de Educación Intercultural (LOEI)" of 2011.

Table 1
Earning categories, Requirements for promotions.
Source: The authors(2021). Articles 302 and 302.1 of the Ecuadorian Education Law and Career Trajectories dataset (2012–2019).

Category	Degree	Minimum experience	Courses	Knowledge test score	Monthly wage
G	Bachelor in Education	0 years	–	–	$ 817
F	Bachelor in Education	4 years	330 h	≥700	$ 901
E	Bachelor in Education	8 years	330 h	≥700	$ 986
D	Masters in Education	12 years	330 h	≥700	$ 1.086
C	Masters in Education	16 years	330 h	≥700	$ 1.212
B	Masters in Education	20 years	330 h	≥800	$ 1.412
A	Masters in Education	24 years	330 h	≥900	$ 1.676

Institute for Evaluation (INEVAL) carried out a structured knowledge test as part of a comprehensive teachers' evaluation process.[5] This test was compulsory for all public sector teachers, who account for 76% of the total teaching population. From the beginning, it was clear to teachers that this test was a prerequisite for promotion in the public sector, which explains the high level of participation. The evaluation process began in May 2016 with 96% of all public school teachers taking the test by October 2016.[6]

The structured knowledge test assesses the teachers' knowledge in the subject with the highest teaching load. The test contains 120 questions with different difficulty levels and is marked out of 1000 points. The overall score takes into account the difficulty level of each item. A score of 700 points reflects a good performance on the test. In comparison, a score between 600 and 699 points reflects only an acceptable performance, which is why the 700 points cut-off was chosen to determine eligibility for promotions.[7] Teachers with a score at or above the cut-off point have the possibility to apply for promotion if they meet the other additional criteria, which include spending at least four years in their current entry category and passing 330 h of PD. Teachers with scores below the cut-off point are excluded from the promotion process until the next teaching evaluation, which means they have to wait at least four years.[8]

Table 1 shows the promotion requirements for the seven earning categories in the public education sector. Teachers join a public school in category G. The highest-earning category is category A. Teachers in category A must have a Master's in Education and over 24 years of experience in the public education system. To get promoted, teachers must be in their category for at least four years and meet the criteria described in Table 1 for each earning category. For example, to move from category G to F, a teacher must have a Bachelor's degree in Education, wait four years in category G, obtain a minimum score of 700 points in the knowledge test, and have at least 330 h of in-service training. Teachers can accumulate training hours in the four years before the promotion request. Any course that is taken before that time is not considered for a promotion request. Despite meeting the other requirements, teachers who do not pass the knowledge test are ineligible for promotion.

The Ministry of Education manages and organizes the supply of courses available to teachers. Most in-service training is provided through the Ministry's online platform, but local universities offer other training that may involve blended learning. Some courses, like the Curriculum Update courses, are mandatory, while others are not. There was a nationwide Curriculum Update course in 2017 which accredited 100 h of in-service training to teachers who approved it. Teachers had to score above 7 points on a 10-point scale to pass the course. Those who do not pass the course are credited with 0 h. We focus on this course since we obtained administrative data, including the teacher ID and her final score and pass or fail status. This information will allow us to test whether teachers who pass the 2016 teacher evaluation are more likely to pass the Curriculum Update course since they have the incentive to put more effort into acquiring the required training to get a promotion.

The 2017 Curriculum Refresher course covers 5 general and 1 specialty courses. General courses last 16 h each, and the specialty course lasts 20 h. Each course must be passed to start the next one. In total, all the courses add up to 100 h of training (See Appendix B for a description of the courses). Teachers are expected to complete the six courses within approximately 60 days with a maximum daily dedication of 1 h and 45 min to avoid interfering with teaching and personal life. Only teachers who pass the 6 courses (100 h) obtain a certificate that is recognized as a valid professional update course for the promotion and recategorization processes of the Ministry of Education. The courses are MOOCs (Massive Open Online Courses). These are non-tutored courses designed to be completed by teachers at their own pace while consulting texts and videos and completing tasks and questionnaires.[9]

2.2. University entrance exam

The Baccalaureate in Ecuador is not mandatory and corresponds to the last three years of high school. It is equivalent to grades 10 to 12 in the US. Most children are 15 to 18 years old during this phase. All students in the last year of high school at public and private institutions take a mandatory university entrance exam at the end of the academic year. The score in the exam corresponds to 30% of the Baccalaureate graduation mark, so it is a compulsory high-stakes exam that determines high school graduation and access to university. In the 2016–2017 academic year, close to 300,000 students took the exam after INEVAL administered the teachers' structured knowledge test in May–October 2016. Of them, 72.3% studied at public high schools, while the remaining 27.7% studied at private or mixed-funded high schools.

INEVAL designs and administers the university entrance exam for all public and private schools in Ecuador. This exam is similar to the SAT exams in the US.[10]

The university entrance exam evaluates students' knowledge in five domains: math, verbal, scientific, social abilities, and abstract reasoning. Each student receives a mark out of 10 points in each of these

[5] The evaluation contains seven components but only the structured knowledge test score is used to determine promotion eligibility. Hence data for the other components is not available: (i) self-evaluation (3%), (ii) peer-evaluation (8%), (iii) headmaster evaluation (5%), (iv) evaluation of classroom practices (15%), (v) students' and parents' evaluation (4%), (vi) structured knowledge test (48%), and (vii) learning management, socio-emotional abilities and leadership (17%) (INEVAL, 2016a).

[6] https://www.expreso.ec/guayaquil/entrevista-magaliramos-evaluacion-maestros-BG3091210.

[7] A group of experts selects the performance category cut-offs. For this, they sort the questions from the easiest to the most difficult and assess each item's difficulty level. When an item corresponds to the next difficulty level, it is considered a marker question. Finally, all the points below the marker are added to get the maximum score that identifies a teacher with a certain performance level.

[8] There has yet to be a new teacher evaluation to date.

[9] https://educacion.gob.ec/wp-content/uploads/downloads/2016/07/Carta-Descripcion-Curso-Actualizacion-Docente.pdf.

[10] See http://admision.senescyt.gob.ec/etapa/toma-del-examen-ser-bachiller/.

Table 2
Administrative datasets. Description and main outcome variables.
Source: 1,2,3 and 4 can be downloaded from the Ministry of Education website and the INEVAL websites (INEVAL, 2014, 2016b; Mineduc, 2016, 2017), except for the teachers' date of birth (used to merge the Career Trajectories dataset and the Teaching Evaluation dataset) and the 2012–2016 Career Trajectories dataset that we obtained after signing confidentiality agreements with the Ministry of Education.

Register name (Period)	Unit of observation	Data description	Outcome measured
University Entrance Exam data set (2014–2019)[1]	Student	Students' university entrance exam verbal scores. Socio-demographic information (student and family) Student and school ID	Students' verbal scores Scores range from 0 to 10 points
Teacher Evaluation data set (2016)[2]	Teacher	Score on the structured knowledge test. Score ranges from 0 to 1000 points. Teacher date of birth and school ID	
Career trajectories data set (2012–2019)[3]	Teacher	Earning categories Wages Place of work (school ID) Number of consecutive years in earning category in 2016. Teacher ID, date of birth and school ID	Job mobility and Promotions
Curriculum Refresher Course data (2017)[4]	Teacher	Final score, pass/fail status Teacher ID	Probability of passing the course

domains. The verbal scores in the dataset that we analyze in this paper are standardized with mean 0 and standard deviation 1. The complete test contains 160 questions and lasts 180 min, and its objective is to generate critical information for policymakers.

3. Data

Table 2 describes the four sources of administrative data that we use in this paper. Information on students' verbal scores in the university entrance exam is from the University Entrance Exam dataset and is available for 2014–2019. Students taking the exam also complete a survey on learning factors that also collects socio-demographic information related to occupation and educational attainment of family members.[11]

We build a matched teacher–student dataset connecting public school seniors who took the mandatory university entrance exam between 2016 and 2019 with their language teachers. We observe teachers' scores on the 2016 structured knowledge test (Teacher Evaluation dataset) as well as students' verbal scores on the 2016–2019 university entrance exam. Our sample of interest contains schools with only one language teacher teaching senior year. We do this to ensure that we assign the correct treatment status to each teacher–student pair. This way, if teachers have more than one group/class, all classes are assigned to the same teacher. This is important because we have no information about classes. Also, pairing students with more than one teacher would bias downwards any effect on learning, as there are more teachers who do not pass the test threshold than those who do pass the threshold.

There are 1694 language teachers teaching in the last year of high school at the national level. Among them, 890 work at schools with only one language teacher in senior year. We match teachers with their students in the University Entrance Exam data set (2014–2019) and restrict the sample to schools where teachers reported a valid score in the teaching knowledge test in 2016. In this step, we lost 238 language teachers. We then dropped the teachers who moved to a different school after 2016. For this, we use the Ministry of Education's Career Trajectories data set, which contains information about each teacher's post, subject(s) and level(s) taught and earning category. Only 54 language teachers moved to a different school after 2016, leaving 598 teachers/schools in the final sample with some variation in the actual number of teachers and students per year. There are 540 teachers/schools matched to 44,293 students in 2016, and 296 teachers/schools matched with 24,581 students within an indicative 50-point window around the structured knowledge test threshold, as shown in Table 3.

We supplemented the main sample with information that describes whether teachers passed or failed the nationwide Curriculum Refresher Course of 2017 and the final score. This information was available for all the teachers in our sample, with 84% of teachers passing the course within the window of 50 points around the threshold.

Table 3 presents the characteristics of teachers and students in 2016. The teachers in our sample have a high level of education. 20% have a master's degree and the rest have a university degree or equivalent. 77.7% of teachers report being in category 1 (category G) so the average earning category among teachers closest to the threshold is 1.59. The average time spent in the reported income category was about three years, while the average age of teachers was 47 years. About 40% of teachers had taken language courses in the two years prior to the nationwide teacher evaluation, and 6% said they had another job in education or in another sector.

Students in our sample report an average score of 7.38 points (out of 10 points) in the verbal section of the university entrance exam in 2016. Fifty-two per cent of students are male, and 33% belong to a racial minority group that includes indigenous, black or mulatto populations. INEVAL calculates a socio-economic status (SES) index using principal components analysis. The most representative latent variables are household services, parents' educational level and household durables (television, mobile phone, computers) and services (internet, telephone service, etc.) (INEVAL, 2017). The index takes values between −2.5 and 2.5, with lower values indicating households with lower SES. In our sample, the average SES index is −0.51.

Table C.10 in Appendix C shows that our sample includes students and teachers in predominantly rural and smaller-than-average schools in Ecuador, but school characteristics in our sample and in the rest of Ecuador's schools are close among various other dimensions.

4. Empirical strategy

The difficulty in estimating the causal effect of in-service training on student outcomes lies in the fact that teachers may select training based on characteristics unobservable to the researcher. For example, participation in in-service training may be determined by or correlated with unobserved teacher ability, which is also expected to affect student learning. Ideally, teachers would be randomly assigned to receive or not to receive training. As this is not possible in this context, we propose to use a regression discontinuity (RD) design that exploits the discontinuity in the cut-off point of the teacher knowledge test

[11] Other questions include access to the internet, the existence of a computer and learning resources at home, and questions related to child work.

Table 3
Characterization of teachers and students.

Variables	All	Window (50p)
Teacher characteristics:		
Male	0.31	0.28
Minority	0.16	0.13
Has a masters degree	0.21	0.21
Age	47.47	46.84
Is single	0.29	0.31
Has other job	0.09	0.06
Training in math (2014–2016)	0.04	0.03
Training in language (2014–2016)	0.40	0.41
Years in earning category	3.09	2.98
Mean earning category	1.86	1.59
Student characteristics:		
Male	0.52	0.52
Minority	0.36	0.33
SES index	−0.53	−0.51
Language score	7.35	7.39
Teachers	540	296.
Students	44,293	24,581

Notes: This table describes the characteristics of teachers and students in the two matched teacher-students samples in 2016. Students' characteristics are averages at the school level.

Table 4
Local polynomial density tests.

	Teachers below the threshold	Teachers above the threshold
Observations	345	195
Effective_Observations	193	104
Bias_corrected_density	0.01	0.00
Standard_error	0.00	0.00
Bandwidth_values	50.00	50.00
Standard_error_test	0.00	
p-value	0.39	

Notes: The table shows the results of the local polynomial density test proposed by Cattaneo et al. (2016). The selected bandwidth is 50 points. There is no evidence of manipulation of the scores on the teacher evaluations in any of the two samples of teachers.

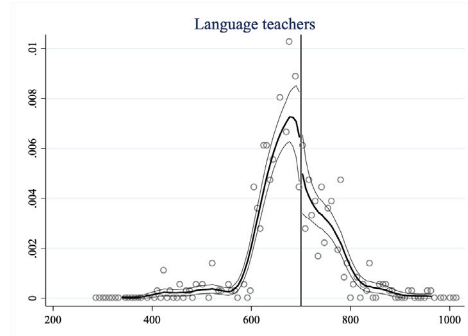


Fig. 1. McCrary density test. **Notes**: Figures show the McCrary density test for the language teachers sample. The selected bandwidth is 50 points. There is no evidence of manipulation of the scores on the teacher evaluation.

that determines eligibility for promotions in the Ecuadorian education system.

The rules that define teachers' eligibility for promotions in the Ecuadorian education system create a natural experiment. Teachers on either side of the cut-off point have very similar observable and unobservable characteristics, except that those who pass the threshold become eligible for promotion and have an incentive to undertake further training, while teachers who do not, do not have the same motivation as they are excluded from the promotion process. Consequently, the probability of passing the training is expected to be higher for teachers who pass the structured knowledge test. Hence, it is possible to use the discontinuity in the knowledge test score as an instrument for the adoption of training, which in turn is expected to improve students' performance in university entrance examinations. We take a reduced form approach and use the cut-off point in the teachers' knowledge test as an instrument to determine the ITT effect of passing the structured knowledge test on students' performance in the university entrance exam and on the likelihood of approving the 2017 Curriculum Refresher Course of 100 h. For this, we use a regression discontinuity design (Hahn et al., 2001; Lee & Lemieux, 2010).

4.1. RD design validity

In this section, we address the main assumptions considered in the literature for the validity of our RD design. The first assumption is that individuals (teachers) are unable to manipulate the assignment variable near the cut-off (structured knowledge test scores). While it is possible that teachers may have studied hard to pass the threshold and become eligible for a promotion, getting a score that falls on either side of the cut-off point is a matter of chance. We argue that manipulation in this case is unlikely because the structured knowledge test is a very complex test where the overall score takes into account the difficulty level of each question. Furthermore, teachers taking the test in 2016 were not familiar with the scoring mechanism ex-ante, considering that this was the first time INEVAL was applying this test as part of a wider teacher evaluation process.

To formally test for manipulation of the scores in the teachers' knowledge test, we use the local polynomial density estimator proposed by Cattaneo et al. (2016) and the McCrary test (McCrary, 2008). If there is no manipulation, we should not observe a gap right before the threshold score of 700 points in the distribution nor a bunching after this cut-off. Table 4 shows the resulting robust-corrected *p*-value of the local polynomial density test (p-value = 0.39). Fig. 1 shows the results from the McCrary test, where we see no bunching at the 700-point threshold. Taken together, these results confirm that there is no statistical evidence of systematic manipulation of the running variable.

The second assumption for the RD design to hold is that the assignment to treatment is approximately randomized around the threshold so that any observed differences in the neighborhood of the cut-off stem solely from differences in the score and not from other observable and unobservable teachers' characteristics. Since we are able to match language teachers with their students, it is possible to compare the average characteristics of the students located on both sides of the threshold and the characteristics of the schools where these students are enrolled. Table C.11 in Appendix C shows that the characteristics of teachers who scored just above and just below the threshold in the knowledge test are balanced in 2016 (prior to the realization of the assignment variable). Balance tests are conducted using the regression discontinuity estimates, with teacher/student baseline characteristics as outcomes. Table C.11 also shows balance across several students' characteristics. Only one out of the close to 40 teacher and student characteristics considered is unbalanced within a window of ±50 points to the threshold. Teachers who pass the knowledge test threshold are slightly more likely to have a second job. However, taken together, the observed differences across all the variables considered are not statistically significant, as reflected by the omnibus F-test (p-value = 0.81).

The third assumption requires that all other factors evolve smoothly with respect to the running variable (Lee & Lemieux, 2010). One possible threat to identification is that students sorted towards teachers who passed the minimum score of 700 points in the knowledge test. Table C.12 in Appendix C shows balance tests on students' socio-economic status, gender and a minority dummy observed in 2017, 2018 and 2019. Despite small statistically significant differences at the threshold in gender and minority indicators, in particular, in 2017

Table 5
Students' verbal scores: Parametric local linear regression estimates.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.341	0.305	0.137	0.247	0.313	0.189
	(0.315)	(0.241)	(0.233)	(0.0929)	(0.110)	(0.0946)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	66 030	82 033	96 123	66 030	82 033	96 123
Teachers	240	296	342	240	296	342

Standard errors in parentheses.
*** p < 0.01, ** p < 0.05, * p < 0.1.
Notes: This table presents the effect of passing the knowledge test on students' verbal scores during the period 2017–2019 for the sample of language teachers and their students for different windows around the cut-off, namely 40, 50 and 60 points to the cut-off.

and 2018, there are no statistically significant differences in students' socio-economic status in any of the years. Furthermore, taken together, the observed differences are not statistically significant for any of the cohorts, as shown by the p-values of the joint F-tests. We also argue that the publication of the teacher evaluation results could not induce student sorting in 2017–2019 for two reasons. First, even though teachers could check their results in the evaluation shortly after the test, there was a lag of at least one year before the results of the teacher evaluation were released to the public, so at least in the very short term, results should not be contaminated by students' sorting. Second, at public schools, students re-enroll automatically for the next academic year. It is possible that some students may choose to change schools at the beginning of the Baccalaureate, but it is less likely that they will move after that. We can also rule out teacher sorting since we drop from the sample teachers who changed schools after 2016.[12]

Finally, in Appendix C, we perform placebo tests to confirm that there are no jumps in student scores or in the likelihood of teachers passing the Curriculum Refresher Course at other cut-off points apart from the 700-point threshold. Given that very few teachers score below 600 points on the teacher knowledge test, for the placebo tests, we choose 620, 640 and 650 points as the threshold of interest. Tables C.13 to C.18 show the results of these placebo tests. We refrained from choosing cut-off points near or at 800 and 900 points because these cut-offs apply to teachers in the highest income categories (as shown in Table 1), which makes them focal points, even though there are very few of them in our sample.

5. RD results

5.1. RD results: University entrance exam scores

Our main objective is to investigate the effect of a teacher passing the structured knowledge test on her student's test scores in the university entrance exam. For this, we estimate the following parametric linear RD regression:

$$Y_{ijc} = \alpha_c + \lambda_t + \beta T_{ij} + \delta D_{ij}^n + \gamma (T_{ij} \times D_{ij}^n) + \epsilon_{ijc} \tag{1}$$

where Y_{ijc} is the standardized verbal score in the university entrance exam for student i with teacher j in parish c in 2017–2019. α_c are parish fixed effects,[13] λ_t are calendar year fixed effects, T_{ij} is an indicator variable equal to 1 if teacher j of student i got a score that is above 700 points and 0 otherwise. D_{ij} is the distance between the teacher's score and the cut-off point, and D_j^n denotes a polynomial of order n for D_{ij}. As before, β captures the effect of passing the knowledge test cut-off on students' scores in the university entrance exam. Standard errors ϵ_{ijc} are clustered at the school level.

The choice of bandwidth in the results tables is given by the optimal bandwidth method proposed by Calonico et al. (2014); Calonico et al. (2015);and Calonico et al. (2020). The optimal bandwidths obtained with rdbwselect are sensitive to the inclusion of covariates, although the values remain within the range of ±40 and ±60 points around the threshold for all the outcomes of interest.[14]

Table 5 reports the estimates from Eq. (1) where D_j^n is a polynomial of order 1. Columns 1 to 3 report the basic specification without controls for three different windows around the cut-off, while Columns 4 to 6 add controls to the specifications reported in Columns 1 to 3. These controls include student's gender, race, socio-economic index and the mean pre-treatment verbal score at the school in 2016. The estimates reported in Columns 1 to 3 are very similar in size to those reported in Columns 4 to 6. However, the latter are estimated with more precision and thus appear as statistically significant at five and one per cent level. According to this, having a teacher who passed the teacher knowledge test has a positive and statistically significant effect on her students' verbal scores observed in 2017–2019 that ranges from 0.19 to 0.31 standard deviations depending on the window considered. Fig. 2 shows the RDD graph depicting mean residuals of a regression of students' test scores on parish-fixed effects, time-fixed effects and other students' characteristics as a function of the structured knowledge test threshold.

Note that the results in column 4 are noticeably different to the ones in column 1, which shows the results of the local linear regression without covariates. The results are especially sensitive to the introduction

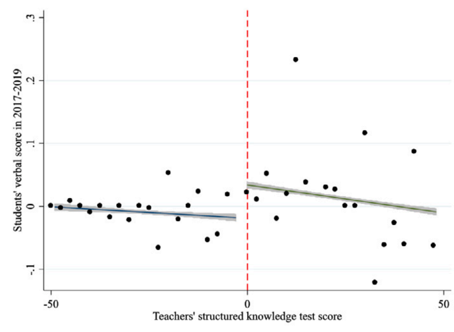


Fig. 2. RD graph: University entrance exam scores. **Notes:** The RDD graph shows the mean residuals of a pooled regression of students' test scores (2017–2019) on parish fixed effects, time fixed effects, student's gender, race, socio-economic index and the mean pre-treatment verbal score of the school in 2016.

[12] Given that there is evidence that teachers who do well in these types of tests have higher mobility (see, for example, Berlinski and Ramos (2020)), we tested if the likelihood of changing schools for teachers is discontinuous at the cut-off after the 2016 evaluation. The results of the analysis are included in the paper in Table C.19 of Appendix C. Teachers who pass the threshold of the knowledge test are slightly less likely to change schools, and the difference is not statistically significant.

[13] We use parish fixed effects instead of canton fixed effects to increase precision considering that we have more observations than in the teacher level regressions where we use canton-level fixed effects. There are more parishes (1499) than cantons (221). Clustering the standard errors at the canton level does not change our results.

[14] For the local linear regression where the student's verbal score is the outcome of interest, the optimal bandwidth ranges from ±52.7 without covariates to ±54.8 with covariates. When the outcome is teacher's promotion, it ranges from ±47.5 without covariates to ±48.8 with covariates; and, when the outcome is passing the curriculum refresher course it ranges from ±50.6 without covariates to ±44.7 with covariates. For this reason, we chose to present 3 bandwidths across all the results tables, namely ±40, ±50 and ±60 points.

Table 6
Probability of passing the Curriculum Refresher Course of 2017: Parametric local linear regression estimates.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.114	0.106	0.186	0.118	0.122	0.193
	(0.0891)	(0.0811)	(0.0764)	(0.0841)	(0.0767)	(0.0734)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	240	296	342	240	296	342
Teachers	240	296	342	240	296	342

Standard errors in parentheses.
*** p < 0.01, ** p < 0.05, * p < 0.1.
Notes: This table presents the effect of passing the knowledge test on the probability of passing the curricular updating course of 2017 among the language teachers in our main sample for different windows around the cut-off, namely 40, 50 and 60 points to the cut-off. Columns 1 to 3 report the results of the RD regression in Eq. (2) without controls, while Columns 4 to 6 add controls to the regressions in Columns 1 to 3. Standard errors are clustered at the parish level.

Table 7
Probability of getting a promotion: Parametric local linear regression estimates (One cross-section).

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.420	0.304	0.274	0.401	0.279	0.256
	(0.204)	(0.182)	(0.151)	(0.216)	(0.189)	(0.156)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	240	296	342	240	296	342
Teachers	240	296	342	240	296	342

Standard errors in parentheses.
*** p < 0.01, ** p < 0.05, * p < 0.1.
Notes: This table presents the effect of passing the knowledge test on promotions during the period 2017–2019 for our main sample of language teachers and their students for different windows around the cut-off, namely 40, 50 and 60 points to the cut-off. Columns 1 to 3 report the results of the RD regression in Eq. (3) without controls. Columns 4 to 6 add controls to the regressions in Columns 1 to 3. These controls include the teacher's gender, race, age, age squared and a dummy for a graduate degree. Standard errors are clustered at the parish level.

of the mean socio-economic index and remain almost unaffected after the introduction of the remaining covariates, namely, student gender, a minority dummy (that indicates whether the student belongs to a racial minority) and the average verbal score in 2016. While these differences are more pronounced with the smaller bandwidth of ±40 points, they are less pronounced when we consider larger bandwidths of ±50 points and ±60 points. Importantly, for this regression, the optimal bandwidth ranges from ±52.7 without covariates to ±54.8 with covariates, so the difference between the point estimates reported in columns 1 and 4 considering a bandwidth of ±40 points should not be a cause of concern. The mean socio-economic index is a composite variable that is positively correlated with an omitted variable like student's health or IQ that is also positively correlated with verbal scores. Hence, by omitting it in columns 1 to 3, it introduces upward bias to the point estimates in columns 1 to 3. As expected, when only the mean socio-economic index is included as a covariate, the point estimates are always smaller than those in regressions without controls.

6. Mechanisms: In-service teacher training and teacher incentives

6.1. In-service teacher training

We expect the probability of passing PD courses to be higher for teachers who pass the structured knowledge test threshold. We use the discontinuity in the teacher knowledge test as an instrument for the adoption/approval of training which in turn is expected to explain the observed positive effects on students' performance in the university entrance exams. For this, we estimate the effect of passing the structured knowledge test on the probability of passing the Curriculum Refresher Course of 2017, a nationwide compulsory 100-hour PD course offered by the Ministry of Education. We use a parametric RD regression of the form:

$$Y_{jc} = \alpha_c + \beta T_j + \delta D_j^n + \gamma (T_j \times D_j^n) + \epsilon_{jc} \tag{2}$$

where Y_{jc} indicates whether teacher j in canton c passed the curriculum update training in 2017. α_c are canton fixed effects, T_j is a binary variable equal to 1 if teacher j got a score above 700 points and 0 otherwise. D_j is the distance between the teacher's score and the cut-off point, and D_j^n denotes a polynomial of order n for D_j. β captures the effect of passing the structured knowledge test on the probability of passing the course. We cluster the standard errors ϵ_{jc} at the parish level.

Columns 1 to 3 of Table 6 show the results of the RD regression in Eq. (2) with D_j^n being a polynomial of order 1 for various windows (40, 50 and 60 points of the cut-off). Columns 4 to 6 add controls to the regressions in Columns 1 to 3. These controls include the teacher's gender, race, age, age squared and a dummy for a graduate degree. Teachers who passed the teacher evaluation are 11 to 19 percentage points more likely to pass the nationwide Curriculum Refresher Course in 2017. The effect is statistically significant at five and one per cent level respectively, for the sample of teachers who scored within the 60 points window in the specifications without controls (Column 3) and with controls (Column 6). A similar pattern of results is observed when we use the score as the outcome of interest.[15]

6.2. Promotions

To estimate the effect of passing the structured knowledge test on the probability of earning a promotion at any point in 2017–2019, we estimate the effect of passing the structured knowledge test on the probability of a promotion using the following parametric RD regression:

$$Y_{jc} = \alpha_c + \beta T_j + \delta D_j^n + \gamma (T_j \times D_j^n) + \epsilon_{jc} \tag{3}$$

where Y_j is a binary variable that indicates whether teacher j got a promotion during the period 2017–2018 or not. α_c are canton fixed effects, T_j is a binary variable equal to 1 if teacher j got a score that is above 700 points and 0 otherwise. D_j is the distance between the teacher's score and the cut-off point, and D_j^n denotes a polynomial of order n for D_j. β captures the effect of passing the structured knowledge test on the likelihood of a promotion. We cluster the standard errors ϵ_j at the parish level.

Table 7 shows the results of the RD regressions in Eq. (3) with D_j^n being a polynomial of order 1. As before, the estimates in Columns 1 to 3 do not control for covariates, while Columns 4 to 6 show the estimates of regressions that control for teacher's gender, race, age, age squared and a dummy for graduate degree. The likelihood of getting a promotion is between 26 and 40 percentage points higher for teachers who passed the knowledge test cut-off than for those who did not cross the cut-off, and the effects seem to decrease in size and significance as the window gets larger.

As a robustness check, we exploit the panel structure of the data at the teacher level and the fact that we have information on the year in which the promotion occurred. We create a binary variable equal to 1 in the year of promotion and in all subsequent years and zero in the years prior to promotion. Note that since the treatment variable is fixed across time, we do not estimate an individual fixed effects model but a pooled parametric linear RD regression and control for time and canton fixed effects as follows:

$$Y_{jct} = \alpha_c + \lambda_t + \beta T_j + \delta D_j^n + \gamma (T_j \times D_j^n) + \epsilon_{jct} \tag{4}$$

[15] Results in Table C.20 in Appendix C.

Table 8
Probability of getting a promotion: Parametric local linear regression estimates.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.304 (0.153)	0.159 (0.136)	0.120 (0.115)	0.279 (0.154)	0.123 (0.136)	0.0940 (0.116)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	720	888	1026	720	888	1026
Teachers	240	296	342	240	296	342

Standard errors in parentheses.
*** p < 0.01, ** p < 0.05, * p < 0.1.
Notes: This table presents the effect of passing the knowledge test on promotions during the period 2017–2019 for our main sample of language teachers and their students for different windows around the cut-off, namely 40, 50 and 60 points to the cut-off. The outcome variable is a binary variable equal to 1 in the year of promotion and in all subsequent years and zero in the years prior to promotion.

where Y_{jct} indicates whether teacher j in canton c received a promotion in year t. α_c are canton fixed effects, λ_t are calendar year fixed effects, T_j is a binary variable equal to 1 if teacher j got a score that is above 700 points and 0 otherwise. D_j is the distance between the teacher's score and the cut-off point, and D_j^n denotes a polynomial of order n for D_j. β captures the effect of passing the structured knowledge test on the likelihood of receiving a promotion. We cluster the standard errors ϵ_{jct} at the teacher/school level.

As previously, Columns 1 to 3 of Table 8 show the results of the RD regression in Eq. (4) for various windows from the cut-off. Columns 4 to 6 add controls to the regressions in Columns 1 to 3 including the teacher's gender, race, age, age squared and a dummy for graduate degree. Estimates in Columns 1 and 4 that focus on teachers whose scores fall within 40 points of the cut-off point show that teachers who pass the test are about 30 percentage points more likely to be promoted in subsequent years (2017–2019). The effects appear to diminish and become less significant as the window of analysis becomes larger, although they remain sizeable and quite similar in the specifications with and without controls.[16]

6.3. Dynamics

In this section, we present additional evidence that the training acquired by teachers who pass the teacher evaluation leads to improvements in student scores. Analyzing the differences in the scores obtained by students taught by teachers who passed the knowledge test threshold relative to students taught by teachers who missed the threshold for each year between 2017 and 2019 separately, we find that the effect grows in 2018 and is maintained in 2019. This long-lasting result contrasts with findings from the literature studying remedial teacher training programs, which have shown that the positive effects on students' learning effects disappear after the first year (Lombardi, 2019). Unlike other incentive and training programs, the program we study in Ecuador has the potential to incentivize continuous knowledge acquisition since it is part of the teacher promotion system. Table 9 shows estimates from Eq. (1) for each cohort (i.e. 2017, 2018 and 2019 cohorts). The corresponding RD graphs can be found in Fig. 3 of the C. Overall, the fact that the positive impacts grow over time aligns with a progressive acquisition of knowledge through in-service training as part of their career progression, whereby teachers were expected to gradually accumulate 330 h of PD during the years following the teacher evaluation to get a promotion.

Furthermore, it appears unlikely that the observed positive effects on verbal scores are driven by a drop in performance among teachers who do not pass the knowledge test. In fact, the existing evidence about teachers assigned to remedial training after a bad evaluation result suggests that since teachers are reevaluated shortly after the remedial training is over, they put more effort into preparing for their teaching evaluations, causing a temporary drop in student performance during preparation time (Lombardi, 2019). In the program that we analyze, teachers who do not pass the test are not tested again, so they do not need to use the time that would otherwise be spent in teaching in preparing for a new test. Consequently, we do not expect a drop in student performance.

7. Conclusions

We study the impact of incentivizing teachers to participate in training resulting from the introduction of a career incentives program on students' test scores. We use the discontinuity generated by a national promotion program to estimate the impact of a teacher passing the structured knowledge test on her student's test scores in the verbal section of the university entrance exam. Conditional on students' characteristics, we find a positive and significant effect on the range of 0.19 to 0.31 standard deviations on student's scores. The main mechanism we explore is the role of increased teacher effort in acquiring in-service training, which in this case is incentivized through the promotion program.

We find that teachers who pass the 2016 structured knowledge test put more effort (relative to teachers who did not pass the threshold) into passing and not only attending the 2017 Curriculum Refresher Course. We expected this to be true since this would credit them with 100 h of in-service training out of the 330 h that they need to credit to get a promotion.

This paper contributes to the literature on the impact of teacher training on students' performance by showing that incentivizing teacher training with promotions leads to higher student learning more sustainably and without distorting teachers' incentives as opposed to traditional test-score-based incentives.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Appendix A. Data appendix

A.1. Students' university entrance exams dataset

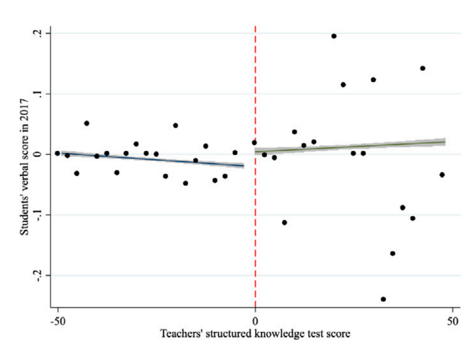
We use an administrative data set that contains the scores of all public school students who took the SB university entrance exam in 2014–2019 and a set of factors associated with school learning. The SB is a requirement for high school graduation and serves also as a university entrance exam. In 2018–2019, 298,317 students took the exam, of which 72.3% studied at public schools. Students taking the exam also fill out a survey about other factors affecting learning. Some questions include access to the internet, the existence of a computer and learning resources at home, and questions related to child work. The survey also collects socio-demographic information about other family members, including the education level of the parents and their occupation.

[16] We lack data on the share of teachers who had the required seniority to go up for promotion in the period 2017–2019 to understand the magnitude of these findings further.

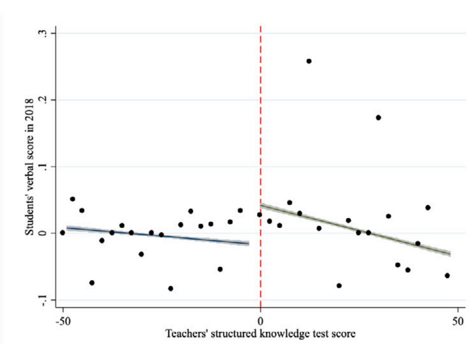
Table 9
Students' verbal scores: Parametric local linear regression estimates for the three cohorts.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: RD results for 2017 cohort						
Treatment	0.369	0.0602	−0.0537	0.268	0.121	0.0461
	(0.366)	(0.276)	(0.260)	(0.183)	(0.139)	(0.115)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	21 439	26 563	31 087	21 439	26 563	31 087
Teachers	240	296	342	240	296	342
Panel B: RD results for 2018 cohort						
Treatment	0.331	0.388	0.135	0.164	0.370	0.182
	(0.360)	(0.276)	(0.268)	(0.117)	(0.123)	(0.111)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	22 173	27 551	32 290	22 173	27 551	32 290
Teachers	240	296	342	240	296	342
Panel C: RD results for 2019 cohort						
Treatment	0.509	0.503	0.355	0.355	0.496	0.359
	(0.298)	(0.242)	(0.228)	(0.123)	(0.156)	(0.123)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	22 418	27 919	32 746	22 418	27 919	32 746
Teachers	240	296	342	240	296	342

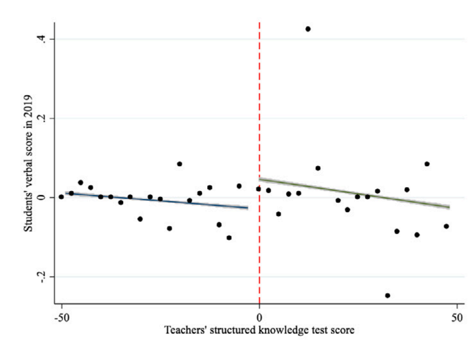
Notes: This table presents the effect of passing the knowledge test on students' verbal scores for each cohort of students for different windows around the cut-off, namely 40, 50 and 60 points to the cut-off. Columns 1 to 3 report the results of the RD regression in equation 2 without controls. Columns 4 to 6 add controls to the regressions in Columns 1 to 3. These controls include the teacher's gender, race, age, age squared and a dummy for a graduate degree.



(A) 2017



(B) 2018



(C) 2019

Fig. 3. RD graphs: University entrance exam scores for each cohort. **Notes**: The RDD graphs show the mean residuals of a regression of students' test scores on various controls (parish fixed effects, time fixed effects, student's gender, race, socio-economic index and the mean pre-treatment verbal score of the school in 2016) for each cohort separately.

A.2. Teacher evaluation dataset

The micro-dataset of the teacher evaluation contains the scores in the knowledge test of all the public school teachers that participated in the 2016 Teacher Evaluation (TE) and several socio-economic characteristics of the teachers. A total of 146,261 public teachers were evaluated in 2016, corresponding to over 96% of all public school teachers in the country.[17] Teachers who took the knowledge test in 2016 also filled out a survey that asks about several socio-economic

[17] https://www.expreso.ec/guayaquil/entrevista-magaliramos-evaluacion-maestros-BG3091210.

characteristics including age, gender, race, education level (degree), courses taken in the past two years, other jobs, among others.

A.3. The career trajectories dataset

The Career trajectories dataset (CT) (2012–2019) comes from two sources. For the period 2012–2016, it was provided by the Ministry of Education after the team signed a confidentiality agreement. This dataset includes teachers' anonymized IDs and names as well as the school code and earning category and subjects and grades taught. For the years 2017–2019, the data was downloaded from the Ministry of Education website.

A.4. Professional development dataset

We complement the main sample with information that describes whether teachers passed or failed the nationwide Curriculum Refresher Course of 2017 and the final score.

A.5. Schools' administrative records

The administrative records of private and public schools are compiled on a yearly basis in a school master file called AMIE. In Ecuador, 2353 schools, equivalent to 62% of all schools that offer Baccalaureate-level education, are public. The data set contains several school-level variables, including school ID, geographic location (city, urban/rural), staff, number of teachers and number of male and female students in different levels, among others.

A.6. Main sample

We merge the Career Trajectory (CT) data (2016) with the Teacher Evaluation (TE) data (2016) using the date of birth of the teachers, gender, and school id. We do this to recover all the subjects and grades taught by each teacher contained in the Career Trajectory data (2016) and use this information to select our sample of interest, which contains schools that have one language teacher teaching in the last year of high school. We are able to merge 95,231 teachers (in 10,978 schools) across all subjects and levels, or 68% of the 140,694 teachers in the TE dataset (2016). Among those teachers, 20,297 teach all subjects in senior year in 2025 high schools.

There are 1694 language teachers teaching in the last year of high school. Among them, 890 work at schools with only one language teacher in senior year. After we match teachers with their students from the University Entrance Exam data set (2014–2019), we obtain an unbalanced panel that includes 652 language teachers who took the teacher knowledge test in 2016 matched to their students. There is some variation in the number of teachers per year. For example, there are 606 teachers in 2016 and 631 in 2019, which is explained by an increase in the number of schools over time. Language teachers are matched to an average of 48,000 senior year students per academic year, totaling over 240,000 students across the 2014–2019 period.

A final step was to restrict the data to schools where teachers kept working in the same school in 2017–2019. Only 54 schools had teachers who moved to a different school after 2016, or 8.3% of the sample, leaving 598 schools in the final sample. To identify the teachers who moved to other schools, we use the Career Trajectory dataset to identify mobility across schools. For this, we build a panel at the teacher level with school ID, subject, and grades taught and earning category from 2012–2019 for senior-level language teachers. To homogenize the data for the two periods (2012–2016 and 2017–2019), we relied on the teacher IDs and school IDs available for the years 2012–2016. First, we homogenized the names of the teachers and schools (dealing with spelling mistakes and common expressions) that introduce inconsistencies across time. We used the names reported in 2016 (or the last year reported) to merge the 2012–2016 dataset with the 2017–2019 administrative data. In this way, we built a panel that follows language teachers who taught senior-year high school students in 2016 from 2012 until 2019. To identify mobility across schools at the teacher level, we look for changes in the reported school (school name) within the same teacher name.

Identifying Promotions with Teachers' Administrative Records

To identify promotions after 2016, we use the panel at the teacher level with school ID, subject and grades taught and earning category from 2012–2019. We create a dichotomous variable equal to 1 if there was a change in the earning category that would indicate a promotion in any year between 2017 and 2019.

Appendix B. Description of the courses covered in the 2017 Curriculum Refresher Course

General Courses[18]

An open and flexible curricular proposal to attend to the diversity of the classrooms This course wants to open a space for reflection on the usefulness of the curriculum documents, their conception, the decisions that are made in their design and the implications that these have for the work of authorities and teachers in educational institutions. It presents the main characteristics of the curricular proposal that entered into force in 2016.

The institutional curriculum The purpose of this course is to guide the entire educational community in the construction and development of the PEI and PCI, to strengthen educational management within educational institutions and the design of a clear path to achieve educational quality.

Planning in the second and third level of curricular concretion "Planning allows organizing and conducting the teaching and learning processes necessary to achieve educational objectives. In addition, it leads to reflect and make timely, relevant decisions, to be clear about what learning needs students have, what should be brought to the classroom and how methodological strategies, projects and processes can be organized so that learning is acquired by all, and in this way cater to the diversity of students" (AFCEGB, 2010). This course wants to shed light on all these aspects of planning.

The collaboration of teaching teams in the development of the institutional curriculum In this course we will work on collaboration mechanisms of the teaching teams of the educational institutions for the curricular development in the second and third level of concretion. The operation of the Academic Boards, the Technical-Pedagogical commissions and graduation boards will be addressed.

The evaluation in the classroom This course's purpose is to deal with the most relevant points of evaluation as a process in the classroom from a constructivist approach through themes, strategies, activities, techniques and learning instruments, which, strategically and systematically, serve the teacher to implement them in a creative, proactive and experiential way, in order to improve their professional performance in decision-making aimed at improving the teaching and learning process.

Specialty courses[19]

Specialty courses delve into the peculiarities of each one of the areas of the curriculum and of the sublevels of General Basic Education at Preparatory, Elementary and Middle levels.

Appendix C. Additional tables and figures

See Fig. 3 and Tables C.10–C.20.

[18] https://educacion.gob.ec/wp-content/uploads/downloads/2016/07/Carta-Descripcion-Curso-Actualizacion-Docente.pdf.

[19] https://educacion.gob.ec/wp-content/uploads/downloads/2016/07/Carta-Descripcion-Curso-Actualizacion-Docente.pdf.

Table C.10
School characteristics: Language teachers sample versus excluded public schools.

Variables	In sample	Rest of schools in Ecuador	Difference
Location Urban/Rural	0.54	0.60	−0.07
Language Spanish/Indigenous	0.93	0.90	0.03
Property ownership	0.92	0.88	0.04
Total teachers	35.80	37.15	−1.35
Total admin. workers	4.29	4.44	−0.16
N. female sophomores	50.72	53.70	−2.98
N. male sophomores	56.77	55.39	1.38
N. female juniors	41.55	46.92	−5.37
N. male juniors	44.66	45.15	−0.49
N. female seniors	38.52	41.66	−3.14
N. male seniors	41.85	38.64	3.21

Notes: This table presents school-level covariate balance tests using local linear regression where the treatment is whether the school is in the sample of language teachers or not. The sample is all public schools that offer Baccalaureate level education in 2018–2019, whose teachers scored within 50 points of the cut-off. Clustered standard errors at the school level are used in the omnibus F-test.

Table C.11
Balance tests: Language teacher - Students sample.

Variables	Above threshold	Below threshold	Difference
Teacher characteristics:			
Male	0.33	0.26	0.07
Minority	0.11	0.16	−0.05
Has a masters degree	0.30	0.29	0.01
Has a university degree	0.69	0.71	−0.03
Age	46.47	46.31	0.16
Is single	0.33	0.31	0.02
Has a desk at home	0.30	0.32	−0.02
Has internet at home	0.72	0.75	−0.03
Has a tablet at home	0.22	0.17	0.05
Has a computer at home	0.91	0.91	0.00
Number of family members	3.65	3.75	−0.10
Has other job	0.11	0.01	.09
Is the only earner at home	0.44	0.55	−0.11
Has more than 35 students per class	0.36	0.25	0.11
Number of classes that supervises	5.92	5.57	0.35
Is pursuing a masters degree	0.04	0.04	0.00
University degree is related to the class she teaches	0.90	0.91	−0.01
Has received training in math	0.01	0.05	−0.04
Has received training in language	0.43	0.32	0.11
Has received other type of training	0.15	0.10	0.05
Assigns more than 10 h per week to prepare class	0.45	0.45	−0.00
Assigns more than 10 h per week to review homework	0.54	0.41	0.13
Reads more than 10 h per week	0.03	0.12	−0.09
Years in earning category in 2016	2.80	2.82	−0.02
Mean earning category in 2016	1.61	1.62	−0.00
Student characteristics:			
Male	0.48	0.53	−0.05
Minority	0.32	0.28	0.04
SES index	−0.40	−0.41	0.01
Language score	7.38	7.54	−0.16
Mother finished high school or more	0.28	0.28	0.00
Father finished high school or more	0.28	0.26	0.02
Consider himself good or excellent at math	0.48	0.49	−0.01
Consider himself good or excellent in language	0.74	0.69	0.04
Reads more than 5 h per week	0.15	0.14	0.02
Has internet at home	0.38	0.38	0.00
Has a computer at home	0.36	0.35	0.01
Has more than 10 books at home	0.56	0.54	0.02
Works	0.29	0.33	−0.04
Works for pay	0.32	0.34	−0.03
Joint F test			
Chi2			31.05
p-value			0.81
Teachers			296
Students			24,581

Notes: This table presents covariate balance tests for the complete matched language-teacher-students sample. Balance tests are conducted using the regression discontinuity estimates, with teacher/student baseline characteristics as outcomes. Students' characteristics are averages at the school level. The chosen bandwidth is ±50 points. All covariates are observed in 2016. Clustered standard errors at the school level are used in the omnibus F-test.

Table C.12
Balance tests for student characteristics in 2017–2019.

Variables	Above threshold	Below threshold	Difference	
Student characteristics in 2017:				
Male	0.49	0.55	−0.06	26,563
Minority	−0.07	−0.04	−0.03	26,563
SES index	0.01	0.02	−0.01	25,382
Joint F test				
Chi2			4.66	
p-value			0.20	
Student characteristics in 2018:				
Male	0.52	0.55	−0.03	27,551
Minority	0.00	0.04	−0.05	27,551
SES index	−0.06	−0.04	−0.02	27,271
Joint F test				
Chi2			2.18	
p-value			0.54	
Student characteristics in 2019:				
Male	0.53	0.51	0.01	27,919
Minority	−0.04	−0.02	−0.02	27,919
SES index	0.00	0.01	−0.01	27,650
Joint F test				
Chi2			0.59	
p-value			0.90	

Notes: This table presents balance tests for various student characteristics observed in 2017, 2018 and 2019. The chosen bandwidth is ±50 points. Clustered standard errors at the school level are used in the omnibus F-test.

Table C.13
Placebo test: Students' verbal scores.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	−0.126	−0.113	−0.0495	−0.134	−0.168	−0.0989
	(0.110)	(0.105)	(0.130)	(0.139)	(0.151)	(0.193)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	41 611	51 659	64 403	41 611	51 659	64 403
Teachers	164	202	249	164	202	249

Standard errors in parentheses.
Notes: This table presents the effect of passing the knowledge test on students' verbal scores during the period 2017–2019 for our main sample of language teachers and their students for different windows around the cut-off, namely 40, 50 and 60 points, to a cut-off point of 620 points. Columns 1 to 3 report the results of the RD regression in equation 2 without controls. Columns 4 to 6 add controls to the regressions in Columns 1 to 3. These controls include the teacher's gender, race, age, age squared and a dummy for a graduate degree.

Table C.14
Placebo test: Students' verbal scores.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.178	0.00898	−0.103	0.0881	0.0340	−0.0671
	(0.194)	(0.159)	(0.167)	(0.204)	(0.134)	(0.147)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	63 595	73 089	83 092	63 595	73 089	83 092
Teachers	245	287	330	245	287	330

Standard errors in parentheses.
Notes: This table presents the effect of passing the knowledge test on students' verbal scores during the period 2017–2019 for our main sample of language teachers and their students for different windows around the cut-off, namely 40, 50 and 60 points to a cut-off point of 640 points. Columns 1 to 3 report the results of the RD regression in equation 2 without controls. Columns 4 to 6 add controls to the regressions in Columns 1 to 3. These controls include the teacher's gender, race, age, age squared and a dummy for a graduate degree.

Table C.15
Placebo test: Students' verbal scores.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0649	0.197	0.200	−0.0561	0.209	0.205
	(0.130)	(0.122)	(0.119)	(0.106)	(0.136)	(0.133)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	70 151	82 438	87 725	70 151	82 438	87 725
Teachers	272	328	344	272	328	344

Standard errors in parentheses.
Notes: This table presents the effect of passing the knowledge test on students' verbal scores during the period 2017–2019 for our main sample of language teachers and their students for different windows around the cut-off, namely 40, 50 and 60 points to a cut-off point of 650 points. Columns 1 to 3 report the results of the RD regression in equation 2 without controls. Columns 4 to 6 add controls to the regressions in Columns 1 to 3. These controls include the teacher's gender, race, age, age squared and a dummy for a graduate degree.

Table C.16
Placebo test: Probability of passing the Curriculum Refresher Course of 2017.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	−0.0556	−0.0395	0.0326	−0.109	−0.0811	−0.00945
	(0.171)	(0.168)	(0.121)	(0.174)	(0.170)	(0.116)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	164	202	249	164	202	249
Teachers	164	202	249	164	202	249

Standard errors in parentheses.
Notes: This table presents the effect of passing the knowledge test on the probability of passing the curricular updating course of 2017 among the language teachers in our main sample for different windows around the cut-off, namely 40, 50 and 60 points to a cut-off point of 620 points. Columns 1 to 3 report the results of the RD regression in Eq. (2) without controls, while Columns 4 to 6 add controls to the regressions in Columns 1 to 3.

Table C.17
Placebo test: Probability of passing the Curriculum Refresher Course of 2017.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0260	0.0578	0.0155	0.100	0.123	0.0794
	(0.111)	(0.108)	(0.102)	(0.110)	(0.107)	(0.105)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	245	287	330	245	287	330
Teachers	245	287	330	245	287	330

Standard errors in parentheses.
Notes: This table presents the effect of passing the knowledge test on the probability of passing the curricular updating course of 2017 among the language teachers in our main sample for different windows around the cut-off, namely 40, 50 and 60 points to a cut-off point of 640 points. Columns 1 to 3 report the results of the RD regression in Eq. (2) without controls, while Columns 4 to 6 add controls to the regressions in Columns 1 to 3.

Table C.18
Placebo test: Probability of passing the Curriculum Refresher Course of 2017.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0142	−0.0119	−0.00265	0.0216	0.0148	0.0239
	(0.110)	(0.0974)	(0.0962)	(0.108)	(0.0949)	(0.0938)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	272	328	344	272	328	344
Teachers	272	328	344	272	328	344

Standard errors in parentheses.
Notes: This table presents the effect of passing the knowledge test on the probability of passing the curricular updating course of 2017 among the language teachers in our main sample for different windows around the cut-off, namely 40, 50 and 60 points to a cut-off point of 650 points. Columns 1 to 3 report the results of the RD regression in Eq. (2) without controls, while Columns 4 to 6 add controls to the regressions in Columns 1 to 3.

Table C.19
Testing discontinuity in teacher mobility to a different school.

	(1)	(2)	(3)	(4)	(5)	(6)
Treat	−0.0707	−0.0700	−0.0445	−0.0728	−0.0684	−0.0448
	(0.0674)	(0.0612)	(0.0531)	(0.0673)	(0.0615)	(0.0523)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	258	320	368	258	320	368
Teachers	258	320	368	258	320	368

Standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Notes: This table presents the effect of passing the knowledge test on the probability that teachers move to a different school for our main sample of language teachers for different windows (40, 50 and 60 points) around the 700 points cut-off. Columns 1 to 3 report the results of the RD regression in equation 2 without controls. Columns 4 to 6 add controls to the regressions in Columns 1 to 3. These controls include the teacher's gender, race, age, age squared and a dummy for a graduate degree.

Table C.20
Score of the Curriculum Refresher Course of 2017.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	1.031	0.930	1.630**	1.072	1.087	1.698**
	(0.831)	(0.752)	(0.682)	(0.784)	(0.711)	(0.662)
Bandwidth	±40	±50	±60	±40	±50	±60
Controls	No	No	No	Yes	Yes	Yes
Observations	240	296	342	240	296	342
Teachers	240	296	342	240	296	342

Standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Notes: This table presents the effect of passing the knowledge test on the score of the curricular updating course of 2017 for our main sample of language teachers for different windows (40, 50, and 60 points) around the 700 points cut-off. Columns 1 to 3 report the results of the RD regression in equation 2 without controls. Columns 4 to 6 add controls to the regressions in Columns 1 to 3. These controls include the teacher's gender, race, age, age squared, and a dummy for a graduate degree.

References

Andrabi, T., & Brown, C. (2020). Subjective versus Objective Incentives and Employee Productivity. *6*(11), 951–952.

Angrist, J. D., & Lavy, V. (2001). Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools. *Journal of Labor Economics*, *19*(2), 343–369.

Azam, M., & Kingdon, G. (2015). Assessing teacher quality in India. *Journal of Development Economics*, *117*(C), 74–83.

Baker, G. (2002). Distortion and risk in optimal incentive contracts. *Journal of Human Resources*, *37*(4), 728–751.

Bau, N., & Das, J. (2020). Teacher value added in a low-income country. *American Economic Journal: Economic Policy*, *12*(1), 62–96.

Behrman, J. R., Parker, S. W., Todd, P. E., & Wolpin, K. I. (2015). Aligning learning incentives of students and teachers : Results from a social experiment in Mexican high schools. *Journal of Political Economy*, *123*(2), 325–364.

Berlinski, S., & Ramos, A. (2020). Teacher coaching in Kenya: Examining instructional support in public and nonformal schools. *Journal of Public Economics*, *186*.

Borman, G. D., Slavin, R. E., Cheung, A. C. K., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of success for all. *American Educational Research Journal*, *44*(3), 701–731.

Brown, C., & Andrabi, T. (2021). Inducing positive sorting through performance pay: Experimental evidence from Pakistani schools.

Bruns, B., & Luque, J. (2015). *Great teachers. How to raise student learning in Latin America and the Caribbean* (pp. 1–13). Number February 2019.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, *23*(2), 192–210, https://academic.oup.com/ectj/article/23/2/192/5625071.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs: robust nonparametric confidence intervals. *Econometrica*, *82*(6), 2295–2326, http://doi.wiley.com/10.3982/ECTA11757.

Calonico, S., Cattaneo, M., & Titiunik, R. (2015). Rdrobust: an r package for robust nonparametric inference in regression-discontinuity designs. *The R Journal*, *7*(1), 38, https://journal.r-project.org/archive/2015/RJ-2015-004/index.html.

Cattaneo, M. D., Jansson, M., & Ma, X. (2016). rddensity : Manipulation testing based on density discontinuity. *The Stata Journal*, (ii), 1–18.

De Talancé, M. (2017). Better teachers, better results? Evidence from rural Pakistan. *The Journal of Development Studies*, *53*(10), 1697–1713.

Duflo, B. E., Hanna, R., & Ryan, S. P. (2012). Incentives work : Getting teachers to come to school. *American Economic Review*, *102*(4), 1241–1278.

Friedman, J. N., Rockoff, J. E., & Chetty, R. (2014). Measuring the impacts of teachers I : Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632.

Fryer, R. G. (2013). Teacher incentives and student achievement : Evidence from New York city public schools. *Journal of Labor Economics*, *31*(2), 373–407.

Garet, M., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H. S., Doolittle, F., Zhu, P., & Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement*: *Technical Report September*, U.S. Department of Education. NCEE 2008-4030.

Glewwe, P., Ilias, N., & Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, *2*(3), 205–227.

Hahn, B. Y. J., Todd, P., & Klaauw, W. V. d. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209.

INEVAL (2014). University entrance score dataset (ser bachiller). http://evaluaciones.evaluacion.gob.ec/BI/bases-de-datos-ser-bachiller/. (accessed 2021-06-25).

INEVAL (2016a). *Resolución N.º 026-INEVAL-2016 harvey*: *Technical report*.

INEVAL (2016b). Teacher evaluation dataset (ser maestro). http://evaluaciones.evaluacion.gob.ec/BI/bases-de-datos-ser-maestro/. (accessed 2021-06-25).

INEVAL (2017). *Índice socioeconómico: Ser estudiante y ser bachiller*: *Technical report*.

INEVAL (2018). *Educación en ecuador. Resultados de PISA para el desarrollo*: *Technical report*, Quito - Ecuador: Instituto Nacional de Evaluación Educativa.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), 2072–2107.

Jacob, B. A., & Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *The Journal of Human Resources*, *39*(1), 50–79.

Karachiwalla, N., & Park, A. (2016). Promotion incentives in the public sector: Evidence from Chinese schools. *Journal of Public Economics*, *146*, 109–128.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355.

Liu, J., & Loeb, S. (2019). Engaging teachers: Measuring the impact of teachers on student attendance in secondary school. *Journal of Human Resources*, 1216–8430R3.

Lombardi, M. (2019). Is the remedy worse than the disease? The impact of teacher remediation on teacher and student performance in Chile. *Economics of Education Review*, *73*(C).

Loyalka, P., Popova, A., Li, G., & Shi, Z. (2019). Does teacher training actually work? Evidence from a large-scale randomized evaluation of a national teacher training program. *American Economic Journal: Applied Economics*, *11*(3), 128–154.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698–714.

Mineduc (2016). Career trajectories data set. https://educacion.gob.ec/wp-content/uploads/downloads/2016/02/b2_Distributivo_de_personal.pdf (accessed 2021-06-25).

Mineduc (2017). Curriculum refresher course data. https://educacion.gob.ec/cursos-de-actualizacion-docente-2020/.

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay : Experimental evidence from India. *Journal of Political Economy*, *119*(1), 39–77.

Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2021). Teacher professional development around the world: The gap between evidence and practice. *TheWorld Bank Research Observer*, 1–51.

Randel, B., Beesley, A. D., Apthorp, H., Clark, T. F., Wang, X., Cicchinelli, L. F., & Williams, J. M. (2011). *Classroom assessment for student learning: Impact on elementary school mathematics in the central region. Final report. NCEE 2011-4005*: *Technical report*, U.S. Department of Education.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement : Evidence from panel data. *The American Economic Review. Papers and Proceedings of the One Hundred Sixteenth Annual Meeting of the American Economic Association*, *94*(2), 247–252.