FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

PAULA REIS KASMIRSKI

ESSAYS ON ECONOMICS OF EDUCATION

São Paulo
2019

PAULA REIS KASMIRSKI

ESSAYS ON ECONOMICS OF EDUCATION

Tese apresentada à Escola de Economia de São Paulo da Fundação Getulio Vargas como requisito para a obtenção do título de Doutora em Economia

Campo de conhecimento: Microeconomia

Orientador: Vladimir Pinheiro Ponczek

São Paulo

2019

Kasmirski, Paula Reis.
Essays on economics of education / Paula Reis Kasmirski. - 2019.
115 f.

Orientador: Vladimir Pinheiro Ponczek.
Tese (doutorado CDEE) – Fundação Getulio Vargas, Escola de Economia de São Paulo.

1. Economia da educação. 2. Educação - Aspectos econômicos - Brasil. 3. Professores - Avaliação. 4. Alfabetização - Brasil. 5. Pressão dos pares. I. Ponczek, Vladimir Pinheiro. II. Tese (doutorado) – Escola de Economia de São Paulo. III. Fundação Getulio Vargas. IV. Título.

CDU 37.014.54(81)

Ficha Catalográfica elaborada por: Isabele Oliveira dos Santos Garcia CRB SP-010191/O
Biblioteca Karl A. Boedecker da Fundação Getulio Vargas - SP

PAULA REIS KASMIRSKI

ESSAYS ON ECONOMICS OF EDUCATION

Tese apresentada à Escola de Economia de São Paulo da Fundação Getulio Vargas como requisito para a obtenção do título de Doutora em Economia

Campo de conhecimento: Microeconomia

Data da aprovação: 21 de maio de 2019

Banca examinadora:

Prof. Dr. Vladimir Pinheiro Ponczek
(Escola de Economia de São Paulo - FGV)

Profa. Dra. Priscilla de Albuquerque Tavares
(Escola de Economia de São Paulo - FGV)

Profa. Dra. Fernanda Gonçalves De La Fuente Estevan
(Escola de Economia de São Paulo - FGV)

Prof. Dr. Sergio Pinheiro Firpo
(Insper)

Prof. Dr. Fernando Botelho
(FEA - USP)

Este trabalho é dedicado à mim mesma.

Agradecimentos

Gostaria de agradecer, em primeiro lugar, ao Vladimir Ponczek, meu orientador. Obrigada pela ajuda e paciência. Aprendi a ser uma pesquisadora melhor com você.

Agradeço à banca examinadora pelas sugestões e críticas.

Agradeço novamente à Priscilla Albuquerque Tavares, grande amiga, "madrinha" de casamento e ótima profissional. Devo à ela a oportunidade de ter participado do Programa de Formação de Professores-Tutores da Escola de Economia de São Paulo (EESP), experiência que enriqueceu muito a minha formação e me fez redescobrir a docência. Espero que futuros doutorandos continuem tendo essa oportunidade.

Essa jornada infinita de quatro anos e meio não teria sido a mesma se eu não tivesse conhecido meu marido, João, e não tivesse a ajuda das minhas melhores amigas Vanda, Simone e Martha. O apoio de vocês foi fundamental pra mim.

Agradeço à Isabela Furtado pela ajuda com o Latex, o Stata e com as inseguranças do dia-a-dia, ao Vinicius Lima por me iniciar no Mata e ao Possebom pela sugestões valiosas que deu para o segundo artigo.

Agradeço à Secretaria da Educação do Estadao do Ceará, Maurício Holanda Maia e George Gomes Ferreira pelos dados fornecidos e por tirarem as muitas dúvidas que tive sobre as bases e sobre o Paic.

Obrigada à Secretária Municipal de Educação de São Paulo e Thiago Costa pelos dados da Prova São Paulo e pela ajuda para compreender a base.

Agradeço à EESP pelo financiamento de parte das bolsas que desfrutei ao longo do curso.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

"ABC, ABC, toda criança tem que ler e escrever."
(Edson Arantes do Nascimento)

Resumo

Essa tese é composta por três ensaios sobre economia da educação. O primeiro capítulo investiga a contribuição dos efeitos de pares para a qualidade do professor. O segundo capítulo avalia o Programa Alfabetização na Idade Certa (Paic) implementado no Ceará e busca detectar seus efeitos no desempenho acadêmico dos alunos. O terceiro capítulo investiga um dos componentes mais importantes do Paic - o treinamento dos professores alfabetizadores - e verifica se ele gerou spillovers na aprendizagem de alunos que não participaram do programa, mas tiveram contato direto com docentes treinados.

Palavras-chaves: efeito de pares, alfabetização, spillover.

Abstract

This thesis consists of three essays about economics of education. The first chapter investigates the contribution of peer effects to the teacher quality. The second chapter evaluates the Literacy Program at the Right Age (LPRA) implemented in Ceara and aims to detect its impacts on student achievement. The third chapter investigates one of the most crucial elements of LPRA - the in-service teacher training - and verifies if it yielded spillover effects on the performance of pupils not treated by the program, but that had direct contact with trained teachers.

Key-words: peer effects, literacy, spillover.

List of Figures

List of Tables

Contents

1 Teacher peer effects

Abstract

This paper aims to investigate the contribution of peer effects to teacher quality, measured by teacher value-added, in an educational system in Brazil's largest city - the Sao Paulo city system. We analyzed if more effective, experienced, and educated teachers would contribute to the effectiveness of their colleagues. To deal with the issues associated with peer effects estimation, we use two approaches that solve the reflection problem and the perfect collinearity between mean outcome and mean characteristics of peers: instrumental variables, as proposed by Bramoulle, Djebbari and Fortin (2009), and out of sample estimates of teacher quality suggested by Jackson and Bruegmann (2009). We treat the self-selection into reference groups using network or schools fixed effects. Using data from Prova Sao Paulo and School Census, our samples are composed of teachers of elementary students. Based on our first empirical strategy, we provide evidence that more effective peers reduce (in an economically relevant amount for Reading but not for Mathematics) teacher contribution to the student learning, and that other peers' features do not matter in statistical and economic terms. We investigated if the negative endogenous effects were explained by a jointed production and shared resources story, and concluded that this was not its source. Our second approach yielded different results that point to positive endogenous peers' effects, which are economically sizable for Reading.

Keywords: peer effects, teacher quality, teacher value-added.

1.1 Introduction

Many production processes have an output that is a function of the combined effort of many workers, and it is hard to identify and reward the exact contribution made by each employee (MAS; MORETTI, 2009). In this environment, peer effects matter, because the amount of effort put on a task by a worker can depend on her peers' productivity - more productive partners can induce less (free-riding) or more effort (if there is some peer sanction or resentment) (MAS; MORETTI, 2009).

A vital production process that uses numerous human resources along a relatively long period and that imposes challenges to recognize how much an employee contributed to the final product is mandatory education. A clear challenge is how to measure output. A common measure in the literature is standardized achievement test scores, usually of reading and mathematics subjects, but, as pointed out by Hanushek (1979), other measures can be appropriate as well, as attendance or dropout rates.

The adoption of test scores as the schooling output is supported by the fact that this measure is meaningful - student test scores are highly related to school attainment and economic outcomes - and there is a methodology widely accepted in the literature to identify the contribution of a teacher to the student achievement called teacher value-added (KOEDEL; MIHALY; ROCKOFF, 2015). This approach assumes that a good teacher is the one able to consistently increase student learning (HANUSHEK; RIVKIN, 2006).

Given the mentioned features of the schooling outcome, we can conclude that peer effects may be important to explain teacher effectiveness, especially when commonly observed teacher characteristics - education, certification, test scores, experience (at least in a linear way) - explains little of student achievement variation (HANUSHEK; RIVKIN, 2010; AARONSON; BARROW; SANDER, 2007; WINTERS; DIXON; GREENE, 2012). This is something that school administrators would like to know because it would make them able to identify good teachers before they earn the protection of tenure (WINTERS; DIXON; GREENE, 2012).

Adapting the concepts provided by Epple e Romano (2011) and Sacerdote (2011), if having teacher A as a colleague affects teacher B's quality by means of an externality, *ceteris paribus*, this is considered as peer effect. As pointed out by Jackson e Bruegmann (2009), there are at least three sources of spillovers between teachers - jointed production and shared resources, motivation and effort, and peer learning.

The first source of peer effects, as discussed by Jackson e Bruegmann (2009), operates through shared duties outside the classroom, as school planning, and resources, like school facilities. These authors argued that its effect on teacher quality is contemporaneous and it has an ambiguous sign, since good peers can increase or decrease duties, and therefore time, and resources available to their colleagues. The second source changes teaching effort and it also generates contemporaneous and unclear sign effect (JACKSON; BRUEGMANN, 2009). In an environment without accountability, Jackson e Bruegmann (2009) believe that the arrival of a good coworker can create social pressure, since teachers can be embarrassed over the unfavorable direct performance comparison, or it can simply motivate them to be better. Mas e Moretti (2009) define social pressure as the situation in which employees have preferences over how they are perceived by their partners; there is a loss of utility if the worker is observed behaving in an uncooperative way by their peers. In the presence of hard accountability, having a new more effective peer can induce free-riding or social pressure, reducing and increasing teacher quality, respectively (JACKSON; BRUEGMANN, 2009).

Direct or indirect peer learning is the only source with lasting and unambiguous positive empirical prediction (JACKSON; BRUEGMANN, 2009). This is supported by the literature finding that teacher effectiveness generally increases with experience, es-

pecially in the first few years, which means that there is substantial on-the-job learning (JACKSON; BRUEGMANN, 2009; HANUSHEK; RIVKIN, 2010; HARRIS; SASS, 2011). Teachers with less experience and greater labor force attachment are more likely to invest in learning and are more sensitive to peer learning (JACKSON; BRUEGMANN, 2009). Altogether, it is an empirical question if peer effects are positive or negative on average.

Accord with Sacerdote (2011), to estimate peer effects credibly, it is necessary to deal with some issues. When the outcome of an individual varies, her peers' mean outcome changes in response and vice versa, what clearly violates the exogeneity assumption and it is called the reflection problem. The second issue refers to self-selection into reference groups and it can be resumed in a simple way - similar people tend to be together and they behave similarly just because they are alike. In a linear-in-means model, another issue arises - perfect collinearity between mean outcome and mean characteristics of peers, implying that may be not possible to separate endogenous (influence of peer outcomes) from exogenous (influence of peer characteristics) social effects (SACERDOTE, 2011; BRAMOULLE; DJEBBARI; FORTIN, 2009).

Jackson e Bruegmann (2009) pursue evidence of peer learning between teachers and their empirical strategy consists in estimate a value-added model with the inclusion of teacher peer attributes as covariates. To deal with the reflection problem and perfect collinearity of social effects, they used value-added estimates from pre-sample data as a measure of quality, and they use school fixed effects to take account of self-selection into reference groups. Their results show that students' test scores gains increase, in both math and reading, when their teachers' peers have higher mean estimated value-added and provide evidence that the mechanism underlying the endogenous effect is peer learning. We can say that Jackson e Bruegmann (2009) estimated an indirect impact of peers' quality on teacher effectiveness since they looked for the contribution of peers' value-added to the teacher's student learning, instead of teacher's value-added. This spillover can include the indirect effects of direct contact between a teacher's students and the students of her colleagues.

This paper aims to investigate the direct contribution of peer effects to teacher value-added in an educational system in Brazil's largest city - the Sao Paulo city system. In other words, we investigate if a teacher colleagues' quality impacts her own quality, net of other spillover sources. Our empirical strategy consists of estimating an extended version of a linear-in-means model using Bramoulle, Djebbari e Fortin (2009)'s approach. These authors proposed a more general interaction pattern between peers, a directed social network, and characterized the networks where peer effects are identified. In short, what makes spillovers identified is reference group heterogeneity, which in our case is possible because Sao Paulo city teachers are allowed to work in more than one school in a given school year. This means that some teachers, say A, have different groups of peers that do

not interact between them; so changes in colleagues characteristics from one school serve as instruments to changes in A's quality to the colleagues in the other school. This solves the reflection problem and the perfect collinearity between exogenous and endogenous effects. The endogenous reference group formation is solved with the inclusion of network fixed effects. Also we estimate an adapted version of the Jackson e Bruegmann (2009) strategy described above.

Based on our first empirical strategy, we provide evidence that more effective peers reduce (in an economically, but not statistically, relevant amount for Reading but not for Mathematics) teacher contribution to the student learning, and that other peers' features do not matter in statistical and economic terms. The only reasonable source for the negative endogenous effect in our context - jointed production and shared resources - does not seem to explain it. Our second approach brought different results. The peers' mean quality effects are positive in most models and, for Reading, they are economically sizable. Some other peers' features presented notable economic sizes in one of our specifications, such as mothers' education and experience.

This paper is organized as follows. The next section presents the data and the samples selected. Section 3 presents the identification strategies of value-added and peer effects. Section 4 shows the main results and last section concludes.

1.2 Data and samples

We use data on elementary students from *Prova São Paulo* (PSP), provided by Sao Paulo City Education Department. PSP is a standardized test applied, between 2007 and 2012, to all second, fourth, sixth and eighth-grade students and to a sample of third, fifth and seventh-grade pupils of the municipal educational system. We focused on elementary school students, who have only one teacher for the entire year, to avoid spillovers driven by pupils having direct contact with their teacher's colleagues. Besides the tests in Reading and Mathematics, questionnaires were applied in students, their parents, and teachers.

We have two samples: the first one, called 2010-2009, is composed of teachers with at least 5 elementary students with test scores available in 2010 and 2009, and with information about gender, age, Intensive Project for Elementary Students (IPE)[1] participation, shift, preschool, turnover, and repetition in 2010. We called this set of students' characteristics, excepting test scores, basic information. The second, named 2010-2007, is composed by the teachers of the 2010-2009 sample that, in addition, had at least 5 students with proficiency scores available in 2008 and 2007, and with basic information in 2008 or 2007[2]. The sample 2010-2009 is used to estimate peer effects

[1] The Intensive Project for Elementary Students aim is to help those not fully literate.
[2] In 2008, there was no parent questionnaire in PSP and we use socioeconomic variables from 2007.

models according to Bramoulle, Djebbari e Fortin (2009) and the sample 2010-2007 is used to implement the strategy suggested by Jackson e Bruegmann (2009).

In table 1.1, we present a general description of sample 2010-2009. Column PSP 2010 1 refers to all teachers of students who took the test in 2010 and had basic information. In column PSP 2010 2, teachers are restricted to those with at least 5 students. The cutoff imposition reduces the number of teachers, students, schools and classes as expected, but it slightly increases the percent of teachers in 2 schools, a very important feature to identify peer effects, and it does not change the rest.

In columns PSP 2010 2 and PSP 2010-2009 1 of table 1.1, we see that almost all teachers with at least 5 students in 2010 are found in 2009 PSP edition. When the sample is restricted to those pupils with basic plus parent information (preschool, parental education, and family income and size; column PSP 2010-2009 2), we lost a considerable number of teachers and students. It can be noted in table 1.2 that the missing students are those with slightly lower scores in both subjects. The students' sample (PSP 2010-2009 1) are very similar to the population of students who took the test in 2010 (PSP 2010 1) and the students whose teachers have at least 5 pupils who took the test in 2010 (PSP 2010 2).

Table 1.1 – Sample 2010-2009 descriptive statistics

	PSP 2010 1	PSP 2010 2	PSP 2010-2009 1	PSP 2010-2009 2
Number of teachers	3430	2983	2887	1249
Number of pupils	35409	34379	31010	8948
Number of schools	537	536	535	477
Number of RBs	13	13	13	13
Number of classes	4697	3140	3026	1355
Mean of pupils per teacher	10.3	11.5	10.7	7.2
Minimum of pupils per teacher	1	5	5	5
Maximum of pupils per teacher	44	44	42	22
% of teachers in more than one school	2.88	3.39	2.15	4.32
Minimum of schools per teacher	1	1	1	1
Maximum of schools per teacher	2	2	2	2
% of teachers in only one DRE for those in 2 schools	81.89	87.13	80.65	85.19
Number of schools that share teachers	154	141	106	94

Notes: PSP 2010 1 contains all teachers whose students had basic information (test score, gender, age, PIC, shift, turnover and repetition) and took the test in 2010; PSP 2010 2 is similar to PSP 2010 1, but it restricts teachers to those with at least 5 students; PSP 2010-2008 1 contains teachers with at least 5 students with basic information and that took the test in both years; PSP 2010-2008 2 contains teachers with at least 5 students with basic and parent information (besides the features already mentioned, parents‘ education and family’s income and size) and took the test in both years. RB means Regional Board.

Table 1.2 – Descriptive statistics of students - sample 2010-2009

	PSP 2010 1		PSP 2010 2		PSP 2010-2009 1		PSP 2010-2009 2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
2010 raw Reading score	160.28	44.63	160.62	44.60	162.13	44.52	172.71	43.23
2010 raw Mathematics score	167.14	44.20	167.62	44.11	169.03	44.13	179.92	43.38
Age	10.04	1.04	10.05	1.04	10.02	1.01	10.05	0.94
2nd grade in 2010	0.01	0.09	0.00	0.03	0.00	0.01	0.00	0.01
3rd grade in 2010	0.38	0.49	0.38	0.49	0.38	0.49	0.32	0.47
4th grade in 2010	0.61	0.49	0.62	0.49	0.62	0.49	0.68	0.47
Proportion of girls	0.47	0.50	0.46	0.50	0.47	0.50	0.51	0.50
Prop. of students with age-grade distortion	0.10	0.29	0.10	0.29	0.09	0.28	0.07	0.26
Prop. of pupils in IPE	0.13	0.34	0.13	0.34	0.13	0.33	0.10	0.30
Prop. of students in morning shift	0.51	0.50	0.51	0.50	0.51	0.50	0.51	0.50
... intermediate shift	0.05	0.21	0.05	0.21	0.04	0.21	0.03	0.16
... afternoon shift	0.44	0.50	0.44	0.50	0.45	0.50	0.46	0.50
2009 raw Reading score					152.11	46.03	162.12	46.27
2009 raw Mathematics score					152.38	44.33	162.07	45.67
2nd grade in 2009					0.38	0.49	0.32	0.47
3rd grade in 2009					0.44	0.50	0.51	0.50
4th grade in 2009					0.18	0.38	0.17	0.38
Proportion of school movers					0.04	0.19	0.03	0.17
Prop. of students that repeated 2009's grade					0.18	0.38	0.17	0.38
Observations	35409		34379		31010		8948	

Notes: PSP 2010 1 contains all students with basic information (test score, age, gender, age-grade distortion, IPE, shift) and took the test in 2010; PSP 2010 2 is similar to PSP 2010 1, but it restricts to those with teachers with at least 15 pupils; PSP 2010-2008 1 contains students from teachers with at least 15 pupils that had basic information plus turnover and repetition and took the test in both years; PSP 2010-2008 2 is similar to PSP 2010-2008 1, but pupils had basic plus preschool, parents' education and family's income and size. SD means standard deviation and MW minimum wage. Intensive Project for Elementary Students (IPE) aims to help those not fully literate.

Table 1.19 in the appendix shows descriptive statistics of students used to estimate past teacher value added to sample 2010-2007. The population evaluated in 2008 has a better academic profile than students merged between 2007 and 2008 PSP editions. This seems to be related to the fact that, in 2007, third-grade students were not evaluated, implying that promoted students of second-grade in 2007 cannot be found in 2008 at all, thus increasing age-grade distortion and IPE participants. It can be noted that the proportion of students that repeated their grade in 2007 is higher than 2009's proportion from table 1.2.

Table 1.3 shows teacher descriptive statistics. In order to maximize teacher sample size, we merge PSP with School Census from the National Institute for Educational Studies and Research *Anísio Teixeira* to recover missing data on demographic characteristics (age and gender). To recover missing data of professional features, we input values for covariates, as experience and education, in the following way: we substituted missing values by the mode of the teachers with non-missing values by age group and gender.

Comparing the columns of table 1.3, it can be noted that sample teacher profile (PSP 2010-2009 1) is very similar to those in the population (PSP 2010) and to teachers whose pupils have also parental information (PSP 2010-2009 2). The sample differs, however, from sample 2010-2007, which has older and more experienced teachers. Teachers are in majority women, white, do not have post graduation and their mothers have not succeeded to achieve more than elementary school. More than half has less than 45 years old and almost half have more than 20 years of experience in sample 2010-2009; sample 2010-2007 practically does not have teachers under 35 years old and with less than 10 years of teaching experience.

Despite that almost half of the teachers reported working in 2 or more schools, sample 2010-2009 has less than 2% (see table 1.1) of teachers in this situation for two reasons. The other schools that a teacher works may belong to another system (state or private) and a teacher can teach different grades in different schools. In this case, even if a teacher works in 2 or more municipal schools, if in one school her classroom contains kids from a grade not included in the PSP design, we will not know that she works there.

Table 1.3 – Descriptive statistics of teachers

Proportion of	PSP 2010		PSP 2010-2009 1		PSP 2010-2009 2		PSP 2010-2007	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Woman	0.97	0.18	0.96	0.19	0.96	0.19	0.96	0.20
Less than 35 years old	0.17	0.38	0.17	0.38	0.18	0.38	0.05	0.21
35 to 45 years old	0.39	0.49	0.39	0.49	0.37	0.48	0.40	0.49
45 to 55 years old	0.29	0.45	0.29	0.45	0.29	0.46	0.35	0.48
55 years old or more	0.15	0.35	0.15	0.36	0.16	0.37	0.20	0.40
Fewer than 10 years of experience	0.16	0.37	0.16	0.37	0.16	0.37	0.05	0.22
11 to 20 years of experience	0.38	0.49	0.39	0.49	0.38	0.49	0.38	0.49
20 years of experience and over	0.46	0.50	0.45	0.50	0.46	0.50	0.57	0.50
Another school system or profession	0.56	0.50	0.56	0.50	0.54	0.50	0.46	0.50
Teaches in 2 or more schools	0.48	0.50	0.47	0.50	0.44	0.50	0.36	0.48
High school	0.00	0.06	0.00	0.06	0.00	0.06	0.01	0.08
Higher education	0.78	0.41	0.78	0.42	0.76	0.43	0.79	0.41
Specialization	0.20	0.40	0.21	0.41	0.23	0.42	0.19	0.40
Master or PhD	0.01	0.10	0.01	0.09	0.01	0.09	0.01	0.10
White	0.72	0.45	0.72	0.45	0.75	0.43	0.72	0.45
Black	0.27	0.44	0.27	0.44	0.23	0.42	0.26	0.44
Asian	0.02	0.12	0.02	0.13	0.02	0.13	0.02	0.13
Indigenous	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.00
Mother's education - elementary school	0.67	0.47	0.67	0.47	0.65	0.48	0.70	0.46
Mother's education - middle school	0.13	0.33	0.13	0.33	0.13	0.34	0.12	0.33
Mother's education - high school	0.11	0.31	0.11	0.31	0.11	0.32	0.09	0.29
Mother's education - higher education	0.09	0.29	0.10	0.30	0.10	0.30	0.09	0.28
Observations	4411		2887		1249		503	

Notes: PSP 2010-2009 shows all teachers whose students took the test in both years; PSP 2010-2009 1 contains teachers whose students had basic information (gender, age, PIC, shift, preschool, turnover and repetition) and took the test in both years; PSP 2010-2009 2 contains teachers whose students had basic and parent information (besides the features already mentioned, parents' education and family's income and size) took the test in both years. SD means standard deviation. Except for PSP 2010, it is was considered only teachers with at least 5 pupils.

1.3 Empirical strategy

Before presenting the peer effects models, we will present the teacher value-added model used to estimate our outcome of interest - the teacher quality.[3]

1.3.1 Teacher value added

Assuming that a teacher's quality is captured as her contribution to student achievement, holding other inputs constant, we must develop a learning production function model. Todd e Wolpin (2003) set the student performance at a specific period as a function of the cumulative history of inputs applied by families and schools as well as of children's inherited endowments. They shown that, with our data restrictions, we need to assume that a baseline achievement measure is a sufficient statistic for unobserved input histories and child's innate ability[4]. Additionally, our specification is additively separable with parameters that do not vary with time:

$$y_{isjt} = \alpha_0 + y_{i,t-1}\alpha_1 + \boldsymbol{x}'_{isjt}\boldsymbol{\alpha}_2 + \boldsymbol{s}'_{isjt}\boldsymbol{\alpha}_3 + \boldsymbol{t}'_{isjt}\boldsymbol{\gamma} + e_{isjt} \tag{1.1}$$

y_{isjt} refers to standardized test score[5] of student i at school s with teacher j in year t, $y_{i,t-1}$ is the baseline achievement, $\boldsymbol{x}_{isjt}$ is a vector of student characteristics, such as gender and age, s_{isjt} is a vector of school characteristics (aggregates of the student-level variables or school fixed effects), and $\boldsymbol{t}_{isjt}$ is a vector of teacher indicator variables[6]. Equation 1.1 is consistently estimated by least squares if e_{isjt} is an independent and identically distributed shock (TODD; WOLPIN, 2003). To estimate the teacher quality, we define:

$$\eta_{isjt} = \boldsymbol{t}'_{isjt}\boldsymbol{\gamma} + e_{isjt} \tag{1.2}$$

And we estimate the $\boldsymbol{\gamma}$ in two steps, as presented by Koedel, Mihaly e Rockoff (2015), estimating 1.1 without teacher fixed-effects and then regressing the residuals against $\boldsymbol{t}_{isjt}$. We chose a two-step value-added model (VAM), because it has potential to over-correct for context, and there is no clearly-preferred modeling structure, since the literature reports that one and two steps models have little bias (KOEDEL; MIHALY; ROCKOFF, 2015). Our sample uses teachers with at least 5 students, the maximum possible threshold that our data permitted to obtain a minimally stable quality measure.

[3] Everywhere in this paper, vectors are denoted with bold lower case letters, matrices with bold capital letters and all vectors are in column shape.

[4] In other words, the effects of all prior family and school inputs and ability endowment decay geometrically with the time between the application of the input and the measurement of achievement at the same rate.

[5] For each year and grade, we subtracted the mean from the raw score and then divided it by its standard deviation.

[6] We are aware of our methodology's philosophical inconsistency. We assume that the teacher value-added is a fixed effect, and we aim to estimate the effect of varying features on this immutable measure. We understand that the actual teacher effectiveness changes over time and that our strategy to estimate it captures its persistent component.

The bigger the number of pupils per teacher in a cross-section, the more stable over time is the value added and more accurately we capture the persistent component of teacher quality (KOEDEL; MIHALY; ROCKOFF, 2015).

Tables 1.20 and 1.21 in the appendix report the VAMs with different controls for sample 2010-2009 and tables 1.22 and 1.23 report the models for teacher past VA of sample 2010-2007. We can see that the models are robust and they present coefficients with expected signs and significance - prior achievement has the bigger partial effect, girls are better in Reading and boys in mathematics, parental education and income have a positive effect in child score. Also, as pointed out before, many students and teachers are lost when parental characteristics are added to the regressions. For this reason, in the following exercises, we have chosen to use VAMs with basic student controls.

Correlations between value-added measures with different controls are reported in tables 1.24 and 1.25 in the appendix. For the same subject, the correlations are high in size and significant, showing that quality estimates based only in student basic covariates generate very similar teacher rankings when compared with a more complete model. This is intuitive because the most important student control - prior achievement - is present in all models.

To estimate the adjusted standard deviation of teacher effects, we implemented the procedure suggested by Aaronson, Barrow e Sander (2007) in our two-step model. These authors assume that $\hat{\gamma}_j = \gamma_j + \zeta_j$ and $Cov(\gamma_j, \zeta_j) = 0$, that is, the estimated teacher fixed effect is the sum of the real teacher effect and some error and both are not correlated. This assumption imply that $Var(\hat{\gamma}_j) = Var(\gamma_j) + Var(\zeta_j)$. An estimator for $Var(\hat{\gamma}_j)$ is $\hat{\sigma}^2_{\hat{\gamma}} = \frac{\sum_{j=1}^{J}(\hat{\gamma}_j - \bar{\hat{\gamma}})^2}{J-1}$ and for $Var(\zeta_j)$ is $\hat{\sigma}^2_{\zeta} = \frac{\sum_{j=1}^{J} se^2_{\hat{\gamma}_j}}{J}$, where $se_{\hat{\gamma}_j}$ is the standard error of $\hat{\gamma}_j$, obtained from equation 1.2.

Table 1.4 – Distribution of estimated teacher effects - sample 2010-2009

	No school FE		With school FE	
	Reading	Math	Reading	Math
10th percentile	-.41	-.389	-.417	-.391
25th percentile	-.2	-.208	-.202	-.211
50th percentile	.004	-.006	-.001	-.006
75th percentile	.207	.201	.209	.206
90th percentile	.388	.402	.391	.408
Standard deviation	.33	.335	.332	.336
Adjusted standard deviation	.245	.253	.247	.254
F-statistic teacher FE	2.912	3.163	2.942	3.207
F-statistic student variables	2649.783	2608.484	1048.081	1212.251

Notes: FE means fixed effects. Value added estimated with basic student covariates.

Table 1.5 – Distribution of past value added - sample 2010-2007

	No school FE		With school FE	
	Reading	Math	Reading	Math
10th percentile	-.448	-.447	-.447	-.466
25th percentile	-.226	-.247	-.23	-.247
50th percentile	-.008	-.013	-.015	-.013
75th percentile	.22	.216	.231	.228
90th percentile	.415	.456	.422	.464
Standard deviation	.354	.381	.36	.385
Adjusted standard deviation	.269	.303	.278	.307
F-statistic teacher FE	3.68	3.658	3.847	3.765
F-statistic student variables	1877.777	1653.694	698.377	569.991

Notes: FE means fixed effects. Value added estimated with basic student covariates.

Standard deviations for VAM models with basic controls are presented in tables 1.4 and 1.5. We note that the adjusted standard errors are sizable, what is aligned with the literature stylized fact that teachers vary drastically in their effectiveness. Comparing the standard deviations with and without adjustment in sample 2010-2009, we can conclude that 26% and 24% of the standard deviation of value-added in Reading and Math, respectively, are due to sampling error. For sample 2010-2007, the percentages are 24% and 20%.

In what follows, we will use only basic VA, since the correlations between basic models with and without school fixed effects are high (around 0.99), distributions of estimated effects are similar and, according to Koedel, Mihaly e Rockoff (2015), there is evidence that VAM without school fixed effects perform well and that its inclusion may reduce stability.

1.3.2 Peer effects

In order to identify peer effects in the linear-in-means model, Bramoulle, Djebbari e Fortin (2009) propose to relax the assumption of group interactions - individuals partitioned in groups and affected by all others in the group and by none outside of it. A more general interaction pattern permits that each individual has her own specific reference group, understood as the people whose mean outcome and characteristics affect her own outcome, and interactions are thus structured through a social network.

Here the reference group of a teacher j, denoted by P_j, is broadly defined as all teachers in the same school-grade - therefore it is assumed that their value-added or socioeconomic characteristics affect j's value-added. The peers' value-added can affect a teacher's students achievement at a given period for three reasons already briefly discussed at the introduction. First, they share school resources and duties with her. High quality peers can use more school facilities and supplies, leaving less pedagogic resources to others;

or, if their quality in the classroom is positively associated with their productivity in other areas, better colleagues may reduce the burden of shared tasks, which releases time to teaching activities.

Second, colleagues' quality can affect a teacher effort positively - by motivating her, for example, to study and read more, and pay more attention to struggling students - or negatively - if accountability is present, good peers may stimulate free-riding. The third reason is peer learning, specially at the beginning of the career. Learn how to teach better requires investment, so teachers with greater labor-force attachment and less experience are more likely to invest in learning and more sensitive to peer quality.

We believe that the peers' experience can influence a teacher value-added because a beginner teacher is probably less effective in teaching, as well as in other tasks and, then, learn and be motivated by them seems less likely. Regarding other teacher socioeconomic characteristics, as they do not explain the quality of the teacher herself, we believe that they do not influence the peers' quality, but, in this particular instance, we find worth to test our prediction.

To depart from a group-wise structure, we allow a teacher to belong to more than one group. This is possible because the Sao Paulo City Education Department permits a teacher works in more than one school during a school year. The collection of P_j defines an undirected network between teachers.

Define $v_j = \hat{\gamma}_j$ as the value-added of teacher j, where $\hat{\gamma}_j$ is obtained from equation 1.2. Using Bramoulle, Djebbari e Fortin (2009)'s approach and assuming that there is no self selection into reference groups, peer effects can be estimated from the equation below:

$$v_j = \beta_0 + \frac{\sum_{h \in P_j} v_h}{n_j}\beta_1 + \boldsymbol{z}_j'\boldsymbol{\beta}_2 + \bar{\boldsymbol{z}}_j'\boldsymbol{\beta}_3 + \epsilon_j \tag{1.3}$$

$\frac{\sum_{h \in P_j} v_h}{n_j}$ is the mean value-added of j's peers. The reference group P_j contains n_j teachers and it is assumed that $j \notin P_j$. $\boldsymbol{z}_j$ is a vector of teachers characteristics, such as experience and gender. Each component of $\bar{\boldsymbol{z}}_j$ is defined similar to the mean value added of peers, that is, $\bar{z}_{j,c} = \frac{\sum_{h \in P_j} z_{h,c}}{n_j}$, $c = 1, ..., dim(\boldsymbol{z})$. It is required that $|\beta_1| < 1$. β_1 is the endogenous effect and $\boldsymbol{\beta}_3$ are the exogenous effects, following a nomenclature created by Manski (1993). An endogenous effect captures the influence of peers outcomes and an exogenous (or contextual) effect captures the influence of a peer characteristic. In matrix notation,

$$\boldsymbol{v} = \beta_0\boldsymbol{l} + \boldsymbol{G}\boldsymbol{v}\beta_1 + \boldsymbol{Z}\boldsymbol{\beta}_2 + \boldsymbol{G}\boldsymbol{Z}\boldsymbol{\beta}_3 + \boldsymbol{\epsilon} \tag{1.4}$$

n is the total number of teachers, $\boldsymbol{v}$ is a $n \times 1$ vector of teachers' value-added, $\boldsymbol{l}$ is

a $n \times 1$ vector of ones and $\boldsymbol{G}$ is a $n \times n$ matrix with $G_{jh} = 1/n_j$ if h is a colleague of j and $\mathrm{G}_{jh} = 0$ otherwise. Bramoulle, Djebbari e Fortin (2009) shown that $\boldsymbol{\beta} = (\beta_0, \beta_1, \boldsymbol{\beta}_2', \boldsymbol{\beta}_3')'$ is identified if $\mathbb{E}[\boldsymbol{\epsilon}|\boldsymbol{Z}] = 0$ and, assuming that there is $\beta_{2,c}$ and $\beta_{3,d}$ such that $\beta_1\beta_{2,c} + \beta_{3,d} \neq 0$ for $c = 1, ..., dim(\boldsymbol{\beta}_2)$ and $d = 1, ..., dim(\boldsymbol{\beta}_3)$, the matrices identity, $\boldsymbol{G}$ and $\boldsymbol{G}^2$ are linearly independent.

In this context, natural instruments for $\boldsymbol{Gv}$ arise, such as mean covariates of colleagues' colleagues. To see why, consider three teachers, A, B and C. Suppose that A teaches at school 1 and 2, B teaches at school 1 and C at school 2. C's covariates affect directly A and indirectly B, via her effect in A's value added. By construction, C's covariates are not correlated with the error term and are correlated with the peers' mean value added. In fact, the variables $(\boldsymbol{G}^2, \boldsymbol{G}^3, ...)$ might be used as instruments.

The model discussed so far assumes that there are no correlated effects, which can be defined, according to Manski (1993), as the propensity of individuals in the same reference group to behave similarly just because they are similar or share the same institutional environments. In other words, correlated effects account for self-selection (unobserved to the econometrician) into reference groups (SACERDOTE, 2011). Bramoulle, Djebbari e Fortin (2009) deal with correlated effects by means of network fixed effects. For a teacher j belonging to network r:

$$v_{j,r} = \mu_r + \frac{\sum_{h \in P_j} v_{h,r}}{n_j}\beta_1 + \boldsymbol{z}_{j,r}'\boldsymbol{\beta}_2 + \bar{\boldsymbol{z}}_{j,r}'\boldsymbol{\beta}_3 + \epsilon_{j,r} \tag{1.5}$$

μ_r is the network fixed effect and it accounts for unobserved variables with common effects on teacher effectiveness within the network, such as the same administration and similar preferences for school and students characteristics. We assumed that each Regional Board (RB) forms a network with a stochastic and strictly exogenous interaction matrix $\boldsymbol{G}_r$, as in the empirical application of Bramoulle, Djebbari e Fortin (2009). The matrix $\boldsymbol{G}$ is now block diagonal with $\boldsymbol{G}_r$, for $r = 1, ..., R$, in the diagonal.

It makes sense to consider a RB as a network because most teachers in two schools are in only one board (see table 1.1). A Regional Board is responsible for coordinating the educational policy in a small geographic area (contiguous districts[7]); it is the intermediary management between the Education Department and the schools and it responds for the provision of human resources (SAO PAULO, 2018). So another reference group can be considered: P_j can contain all colleagues that share the same school and grade with j and, for those in more than one school, it is not considered those from schools from different RBs. We explore both definitions of reference groups.

$\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2', \boldsymbol{\beta}_3')'$ is identified if $\mathbb{E}[\boldsymbol{\epsilon}_r|\boldsymbol{Z}_r, \mu_r] = 0$ and, assuming that there are $\beta_{2,c}$ and $\beta_{3,d}$ such that $\beta_1\beta_{2,c} + \beta_{3,d} \neq 0$ for $c = 1, ..., dim(\boldsymbol{\beta}_2)$ and $d = 1, ..., dim(\boldsymbol{\beta}_3)$, the

[7] A district is an administrative unit within a city (IBGE, 2013).

matrices identity, $\boldsymbol{G}$, $\boldsymbol{G}^2$ and $\boldsymbol{G}^3$ are linearly independent.

In order to eliminate unobserved variables, we can transform the data as in linear panel models. Bramoulle, Djebbari e Fortin (2009) suggest two kinds of within transformation, local (deviation from the mean of peers) and global (deviation from the mean of the network). We are able to use any transformation since our data meet the conditions for identification described in the earlier paragraph. Consider the equation 1.5 averaged over all teacher's j colleagues and subtract it from j's equation. In matrix notation, we have

$$(\boldsymbol{I}-\boldsymbol{G})\boldsymbol{v}_r = (\boldsymbol{I}-\boldsymbol{G})\boldsymbol{G}\boldsymbol{v}_r\beta_1 + (\boldsymbol{I}-\boldsymbol{G})\boldsymbol{Z}_r\boldsymbol{\beta}_2 + (\boldsymbol{I}-\boldsymbol{G})\boldsymbol{G}\boldsymbol{Z}_r\boldsymbol{\beta}_3 + (\boldsymbol{I}-\boldsymbol{G})\boldsymbol{\epsilon}_r \tag{1.6}$$

Again, natural instruments for $(\boldsymbol{I}-\boldsymbol{G})\boldsymbol{G}\boldsymbol{v}_r$ arise, namely, the variables $((\boldsymbol{I}-\boldsymbol{G})\boldsymbol{G}^2\boldsymbol{Z}_r, (\boldsymbol{I}-\boldsymbol{G})\boldsymbol{G}^3\boldsymbol{Z}_r, ...)$. The same argument used for the case without correlated effects applies here.

Another way to identify social effects in the linear-in-means model is proposed by Jackson e Bruegmann (2009). They explore the existence of peer effects in the same context as us in a different manner - they verified if changes in a teacher's peers characteristics affect the proficiency growth of her own students. To deal with the endogeneity of peer quality and with collinearity of exogenous and endogenous effects, they estimated teacher value added with pre-sample data. The idea is that this value added was not determined by contact with the contemporaneous peers. We intend to estimate equations 1.3 and 1.5 using pre-sample teacher value added (with data from 2008 and 2007 PSP) to compute mean peers' value added.

The two approaches described in this section propose different solutions to the endogeneity of mean peers' value-added, where the first one uses instrumental variables (IV) and the second one uses a proxy. To compare the estimates from the different methods, consider the equation 1.3 as the structural model. The instrumental variables solution estimates a consistent $\boldsymbol{\beta}$ once the IV assumptions are satisfied by construction. The instruments are both not correlated with $\epsilon_{j,r}$, since $\mathbb{E}[\boldsymbol{\epsilon}_r|\boldsymbol{Z}_r] = 0$, and correlated with the endogenous variable, because each exogenous feature of peers' peers affects the value-added of the peers, which in turn affects only indirectly the teacher value-added.

The proxy solution estimates consistently a vector $\boldsymbol{\beta}' \neq \boldsymbol{\beta}$ under the assumptions that the proxy variable is redundant[8], and

$$\frac{\sum_{h\in P_j} v_h}{n_j} = \theta_0 + \frac{\sum_{h\in P_j} v_{h,pre}}{n_j}\theta_1 + \boldsymbol{z}'_{j,r}\boldsymbol{\theta}_2 + \bar{\boldsymbol{z}}'_{j,r}\boldsymbol{\theta}_3 + u_{j,r} \tag{1.7}$$

Where $\mathbb{E}[u_{j,r}] = 0$, $Cov\left(u_{j,r}, \frac{\sum_{h\in P_j} v_{h,pre}}{n_j}\right) = 0$ and $\theta_1 \neq 0$. Substituting 1.7 in 1.3, it is

[8] We will test this condition, as well as the next proxy assumptions, in the future.

possible to show that $\boldsymbol{\beta}^{'} = (\theta_1\beta_1, (\boldsymbol{\beta}_2+\beta_1\boldsymbol{\theta}_2)', (\boldsymbol{\beta}_3+\beta_1\boldsymbol{\theta}_3)')'$. If our proxy is good, instead of imperfect, $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_3 = 0$, $u_{j,r}$ are not correlated with any $z_{j,c}$ or $\bar{z}_{j,c}$ and $\boldsymbol{\beta}^{'} = (\theta_1\beta_1, \boldsymbol{\beta}_2', \boldsymbol{\beta}_3')'$. Either way, the endogenous effect is multiplied by the partial correlation between the proxy and the average peers' quality.

All models have generated regressors since the value added is estimated in a previous step. So, the standard error coefficients were computed by bootstrapping. We drew 500 samples of 750 students with replacement, and with each sample we estimated teacher value-added and then the betas of peer effects models. In addition, to estimate average characteristics of peers, all teachers with any missing information or isolated ($P_j = \emptyset$) were excluded from regression samples.

1.4 Results

We begin presenting reduced form results. In table 1.6, we see that experience, color and mother's education of peers have an effect in teacher quality both in Reading and Math scores, indicating that these characteristics may have some direct or indirect effect on a teacher‘s effectiveness[9].

Next we present results from Bramoulle, Djebbari e Fortin (2009)'s approach. Tables 1.7 to 1.10 present first stages of our peer effects models to Reading and Mathematics subjects. Until table 1.8, the first model, named IV, assume no existence of correlated effects; the other two (IV *school FE* and IV *RB FE*) contains fixed effects and end up considering as a network the school and the Regional Board (RB), respectively, and applying a global transformation. It is worth mention that the adjacent matrix ($\boldsymbol{G}$) of these models considers that a teacher has a link with another one if they are in the same school, even if the school is in another RB. Models considering explicitly each RB as a network, that is, where a teacher has a link with another one if they are in the same school and RB, and applying a local transformation, are in tables 1.9 and 1.10. The first network has diameter[10] 3 while the second has diameter 2.

The results show that the instruments are weak - there is no instrument with significant partial effect on the endogenous variable and the statistics used to judge the first stage quality (tables' last five lines) are very low. We believe that this is related to the network density[11] and level of intransitivity[12]. Bramoulle, Djebbari e Fortin (2009) analyzed the impact of these measures on the strength of the identification. They concluded that the parameters' precision are higher and bias is smaller with lower density and, when the density is low, the precision rises with intransitivity. Our teacher networks have extremely small density, around 0.002, and no intransitivity at all. Nonetheless, the better first stage models are those that included RB fixed effects, which provides evidence of their importance as a factor to be considered in the estimation of peer effects in Sao Paulo city.

---

[9] The coefficients of peers' attributes in the reduced form are functions of exogenous and endogenous effects, as well as the effects of own individual's characteristics, in the structural model.

[10] The diameter of a network is the maximal distance between two nodes in the network; the distance between nodes i and j is the number of links connecting them in the shortest chain of nodes (BRAMOULLE; DJEBBARI; FORTIN, 2009).

[11] Density is the ratio of the number of links over the total number of possible links (BRAMOULLE; DJEBBARI; FORTIN, 2009).

[12] The level of intransitivity can be defined as the proportion of intransitive triads in the total number of triads. An intransitive triad is set of three nodes i, j and k such that i is affected by j and j is affected by k, but i is not affected by k (BRAMOULLE; DJEBBARI; FORTIN, 2009).

Table 1.6 – Reduced form - Exogenous peer effects

	Reading			Mathematics		
	(1)	(2)	(3)	(4)	(5)	(6)
Peers proportion (PP) of women	0.0101	0.107	6.70e-05	-0.0412	-0.0808	-0.0397
	(0.0551)	(0.163)	(0.0554)	(0.0557)	(0.143)	(0.0559)
PP with fewer than 10 years of experience	-0.0271	-0.193	-0.0137	0.000850	-0.347**	0.000656
	(0.0342)	(0.142)	(0.0350)	(0.0345)	(0.143)	(0.0351)
PP with 11 to 20 years of experience	-0.0538**	-0.0728	-0.0482*	-0.0504*	-0.114	-0.0499*
	(0.0272)	(0.104)	(0.0276)	(0.0274)	(0.0970)	(0.0277)
PP with postgraduate	0.0176	0.00762	0.0124	0.0370	0.137	0.0283
	(0.0296)	(0.110)	(0.0304)	(0.0306)	(0.122)	(0.0317)
PP other job or school system	0.0154	0.128	0.0287	-0.0379	0.0558	-0.0174
	(0.0245)	(0.0955)	(0.0250)	(0.0245)	(0.0860)	(0.0253)
White peers proportion	0.106***	-0.190*	0.0882***	0.104***	-0.0932	0.0716**
	(0.0272)	(0.108)	(0.0281)	(0.0273)	(0.121)	(0.0282)
PP mothers with middle school	-0.00408	-0.0436	-0.0111	-0.0273	-0.0467	-0.0301
	(0.0395)	(0.132)	(0.0399)	(0.0399)	(0.131)	(0.0398)
PP mothers with high school	0.0895**	0.288*	0.0617	-0.00604	0.295*	-0.0322
	(0.0407)	(0.149)	(0.0414)	(0.0433)	(0.163)	(0.0443)
PP mothers with higher education	0.0687	-0.0312	0.0492	0.0337	-0.0689	0.0144
	(0.0478)	(0.176)	(0.0472)	(0.0476)	(0.181)	(0.0477)
Observations	2,939	2,939	2,939	2,939	2,939	2,939
R-squared	0.015	0.008	0.028	0.025	0.014	0.036
F-test	2.597	1.167	3.025	4.246	2.276	3.811
Number of schools		525			525	
School FE		Yes			Yes	
RB FE			Yes			Yes

Notes: Robust standard errors in parentheses. All regressions control for teachers' own characteristics. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.7 – First stage of peer effects models - Reading - Basic VA

Mean characteristics of peers' peers	IV		IV school FE		IV RB FE	
	Coef	SE	Coef	SE	Coef	SE
Women	.02	.356	-.034	2.553	.021	.359
Less than 10 years of experience	-.094	.203	-.25	1.587	-.084	.204
11 to 20 years of experience	-.028	.146	-.067	1.028	-.028	.147
Postgraduate	-.061	.164	-.007	1.095	-.067	.164
Other job or school system	.107	.136	.04	1.037	.119	.137
White	.127	.155	-.015	1.126	.109	.154
Mother with middle school	-.067	.217	-.152	1.513	-.069	.218
Mother with high school	.141	.219	.265	1.689	.125	.221
Mother with higher education	-.041	.247	-.122	1.808	-.047	.246
Observations	2939	.	2939	.	2939	.
R-squared	.041	.	.032	.	.08	.
Adjusted R-squared	.032	.	.023	.	.068	.
Partial R-squared	.005	.	.071	.	.005	.
F-statistic	4.942	.	1.387	.	6.66	.
Partial F-statistic	1.516	.	1.761	.	1.457	.

Notes: Standard errors (SE) were calculated by bootstrapping. In the model with school fixed effects (FE), the R-squared is the within and the partial R-squared is the R-squared from the regression of residualized peers' mean VA versus residualized instruments, where each variable was regressed against exogenous variables. All regressions control for teachers' own characteristics.

Table 1.8 – First stage of peer effects models - Mathematics - Basic VA

Mean characteristics of peers' peers	IV		IV school FE		IV RB FE	
	Coef	SE	Coef	SE	Coef	SE
Women	-.002	.383	.003	2.74	.012	.385
Less than 10 years of experience	-.069	.214	-.435	1.534	-.073	.215
11 to 20 years of experience	0	.154	-.092	1.089	-.008	.155
Postgraduate	-.074	.163	.117	1.149	-.085	.164
Other job or school system	.059	.136	.01	.953	.075	.136
White	.087	.146	-.051	1.082	.055	.147
Mother with middle school	-.081	.208	-.229	1.404	-.077	.209
Mother with high school	-.032	.22	.443	1.702	-.048	.221
Mother with higher education	.137	.255	-.169	1.718	.138	.255
Observations	2939	.	2939	.	2939	.
R-squared	.058	.	.063	.	.091	.
Adjusted R-squared	.049	.	.055	.	.079	.
Partial R-squared	.003	.	.048	.	.003	.
F-statistic	6.278	.	4.018	.	7.646	.
Partial F-statistic	.89	.	4.878	.	.875	.

Notes: Standard errors (SE) were calculated by bootstrapping. In the model with school fixed effects (FE), the R-squared is the within and the partial R-squared is the R-squared from the regression of residualized peers' mean VA versus residualized instruments, where each variable was regressed against exogenous variables. All regressions control for teachers' own and average peers characteristics.

Table 1.9 – First stage of peer effects models with each RB as a network - Reading - Basic VA

Mean characteristics of peers' peers	IV		IV school FE	
	Coef	SE	Coef	SE
Women	-.195	2.61	-.323	4.916
Less than 10 years of experience	-.477	1.625	-.526	3.146
11 to 20 years of experience	.065	1.045	.085	2.191
Postgraduate	-.04	1.128	-.007	2.274
Other job or school system	.134	1.061	.141	1.941
White	.138	1.168	.19	2.187
Mother with middle school	-.508	1.577	-.549	3.143
Mother with high school	.402	1.747	.528	3.502
Mother with higher education	-.434	1.832	-.482	3.594
Observations	2939	.	2939	.
R-squared	.069	.	.082	.
Adjusted R-squared	.06	.	.074	.
Partial R-squared	.032	.	.032	.
F-statistic	5.088	.	6.629	.
Partial F-statistic	5.698	.	7.243	.

Notes: Standard errors (SE) were calculated by bootstrapping. In the model with school fixed effects (FE), the R-squared is the within and the partial R-squared is the R-squared from the regression of residualized peers' mean VA versus residualized instruments, where each variable was regressed against exogenous variables. All regressions control for teachers' own and average peers characteristics.

Table 1.10 – First stage of peer effects models with each RB as a network - Mathematics - Basic VA

Mean characteristics of peers' peers	IV		IV school FE	
	Coef	SE	Coef	SE
Women	-.215	2.819	-.299	5.134
Less than 10 years of experience	-.643	1.554	-.685	3.148
11 to 20 years of experience	.058	1.129	.065	2.236
Postgraduate	.091	1.174	.079	2.364
Other job or school system	.052	.999	.026	1.919
White	-.024	1.117	-.007	2.123
Mother with middle school	-.369	1.458	-.341	2.995
Mother with high school	.467	1.783	.547	3.353
Mother with higher education	-.23	1.801	-.227	3.476
Observations	2939	.	2939	.
R-squared	.076	.	.08	.
Adjusted R-squared	.067	.	.072	.
Partial R-squared	.033	.	.033	.
F-statistic	5.991	.	6.352	.
Partial F-statistic	6.874	.	7.95	.

Notes: Standard errors (SE) were calculated by bootstrapping. In the model with school fixed effects (FE), the R-squared is the within and the partial R-squared is the R-squared from the regression of residualized peers' mean VA versus residualized instruments, where each variable was regressed against exogenous variables. All regressions control for teachers' own and average peers characteristics.

Table 1.11 – Peer effects models - Reading - Basic VA

	OLS		IV		IV School FE		IV RB FE	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Mean VA of peers	.428	.08	-.05	.323	-.961	.481	-.177	.325
Peers proportion (PP) of women	-.004	.201	.012	.301	.101	.67	.004	.311
PP 10 years of experience or less	-.012	.112	-.029	.181	-.136	.358	-.017	.183
PP 11 to 20 years of experience	-.029	.086	-.057	.135	-.092	.282	-.057	.139
PP with postgraduate	.019	.09	.017	.147	.046	.3	.011	.153
PP other job or school system	.02	.073	.015	.119	.025	.248	.029	.128
White peers proportion	.064	.088	.11	.137	-.205	.27	.102	.145
PP mothers with middle school	.004	.119	-.005	.197	-.01	.405	-.016	.204
PP mothers with high school	.073	.127	.091	.206	.182	.391	.064	.213
PP mothers with higher education	.042	.136	.072	.213	.047	.414	.057	.219
Observations	2939	.	2939	.	2939	.	2939	.
F or Wald Chi2 -statistic	10.185	.	46.516	.	32.225	.	87.003	.

Notes: Standard errors (SE) were calculated by bootstrap. All regressions control for teachers' own and average peers characteristics.

The next tables (1.11 to 1.14) report the second stage of estimations. The first model of all tables are ordinary least squares estimates that not considers the endogeneity of mean VA of peers. When the network considers links between RBs, the OLS endogenous effect are around 0.5 and significant for both disciplines; using the average characteristics of peers' peers as instruments for mean VA of peers turn negative the endogenous effects in Reading and makes almost all coefficients be statically zero. When the network ignores links between teachers from different RBs, the endogenous effect in OLS is significant and around -3.5 for Reading and -3.6 for Math. Using IV estimation does not change the effect sign, but reduces its absolute value.

We know from table 1.4 that moving one standard deviation up the distribution of teacher fixed effects is expected to raise both Reading and Math test scores by about 0.25 standard deviations (11 points) on the PSP scale. In economic terms, the endogenous effect for VA seems to be relevant at least to the Reading subject because if the mean VA of peers increases in 0.25, the teacher VA decreases in 0.044 or 17.7% of teacher contribution to student learning[13]. In Math, the endogenous effect explains only 2.9% of teacher total effect in the achievement.

Regarding the exogenous effects, neither of them are statistically different from zero in all models. The lack of statistical significance of our coefficients, as much endogenous as exogenous, contrasts with the results of the reduced form, which lead us to conclude that this is caused by the weak first stage. Nevertheless, most coefficient signs are intuitive - less experienced peers would reduce teacher quality and more peers with postgraduate and better family background, as measured by mother's education, would raise value-added. More peers with another job, inside or outside teaching, seems to improve quality for reading but seems to reduce it for mathematics.

If the proportion of teachers with 11 to 20 years of experience decrease by 0.1 and the proportion with 20 years or more increase by the same amount, the teacher VA to Reading would raise 0.0057, representing only 2.3% of teacher contribution to student learning. The biggest effect comes from white peers proportion (0.102) and a 0.1 increase in this variable represent 4.1% of teacher standard deviation. Based in the results discussed so far, we may conclude that the exogenous effects do not have a significant economic impact on teacher quality.

[13] We used the coefficient from our preferred model *IV RB FE* from table 1.11. This model has the best first stage.

Table 1.12 – Peer effects models - Mathematics - Basic VA

	OLS		IV		IV School FE		IV RB FE	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Mean VA of peers	.453	.081	.164	.321	.043	.487	.029	.336
Peers proportion (PP) of women	-.041	.201	-.041	.334	-.081	.696	-.04	.34
PP 10 years of experience or less	.022	.118	.008	.184	-.347	.375	.002	.192
PP 11 to 20 years of experience	-.02	.087	-.039	.131	-.111	.287	-.048	.134
PP with postgraduate	.037	.094	.037	.15	.134	.314	.029	.152
PP other job or school system	-.016	.078	-.03	.122	.06	.257	-.017	.128
White peers proportion	.044	.082	.082	.129	-.095	.29	.069	.135
PP mothers with middle school	-.011	.111	-.021	.171	-.048	.367	-.029	.179
PP mothers with high school	-.001	.127	-.004	.214	.293	.395	-.031	.222
PP mothers with higher education	.025	.148	.031	.233	-.063	.446	.014	.243
Observations	2939	.	2939	.	2939	.	2939	.
F or Wald Chi2 -statistic	13.906	.	80.061	.	33.893	.	116.548	.

Notes: Standard errors (SE) were calculated by bootstrap. All regressions control for teachers' own and average peers characteristics.

Table 1.13 – Peer effects models with each RB as a network - Reading - Basic VA

	OLS		IV		IV School FE	
	Coef	SE	Coef	SE	Coef	SE
Mean VA of peers	-3.525	.033	-1.152	.417	-1.481	.284
Peers proportion (PP) of women	.048	.313	.03	.681	.035	.434
PP 10 years of experience or less	.37	.175	-.042	.368	.034	.267
PP 11 to 20 years of experience	-.002	.122	-.014	.285	-.016	.188
PP with postgraduate	.158	.147	.013	.306	.066	.207
PP other job or school system	-.264	.121	.024	.251	.009	.175
White peers proportion	-.165	.127	-.158	.28	-.153	.193
PP mothers with middle school	.255	.187	.005	.408	.022	.267
PP mothers with high school	-.408	.195	.094	.402	.018	.281
PP mothers with higher education	.253	.215	.001	.434	-.051	.311
Observations	2939	.	2939	.	2939	.
F or Wald Chi2 -statistic	114.859	.	160.386	.	94.532	.

Notes: Standard errors (SE) were calculated by bootstrap. All regressions control for teachers' own and average peers characteristics.

Table 1.14 – Peer effects models with each RB as a network - Mathematics - Basic VA

	OLS		IV		IV School FE	
	Coef	SE	Coef	SE	Coef	SE
Mean VA of peers	-3.621	.031	-.773	.422	-.961	.269
Peers proportion (PP) of women	.061	.318	-.067	.71	-.097	.472
PP 10 years of experience or less	.342	.187	-.245	.382	-.176	.268
PP 11 to 20 years of experience	-.137	.134	-.046	.295	-.065	.2
PP with postgraduate	.131	.143	.048	.318	.045	.203
PP other job or school system	-.212	.123	.025	.265	.025	.181
White peers proportion	.166	.133	-.082	.291	-.111	.208
PP mothers with middle school	.289	.177	-.025	.375	.051	.276
PP mothers with high school	-.116	.196	.184	.41	.217	.292
PP mothers with higher education	-.377	.209	-.015	.464	-.043	.31
Observations	2939	.	2939	.	2939	.
F or Wald Chi2 -statistic	103.893	.	116.439	.	57.764	.

Notes: Standard errors (SE) were calculated by bootstrap. All regressions control for teachers' own and average peers characteristics.

A negative endogenous peer effect is compatible with a jointed production and shared resources story, where a peer with higher quality increases duties (they can value more time expend with activities related to teaching) and reduces educational resources available to other teachers (they can use them more often). We do not believe that these results are due to a motivation and effort story, because there is no hard accountability in Sao Paulo city schools and, in such an environment, this source of spillovers would increase teacher quality. To investigate the source of teacher spillovers, we estimated the following equations:

$$\frac{\sum_{h \in P_j} v_{h,r}}{n_j} = \rho_r + \rho_{1,k} R_{j,r,k} + \boldsymbol{z}'_{j,r}\boldsymbol{\rho}_2 + \bar{\boldsymbol{z}}'_{j,r}\boldsymbol{\rho}_3 + \omega_{j,r} \tag{1.8}$$

The dependent variable is the mean VA of peers of teacher j that belongs to network r, ρ_r can be school or RB fixed effects and, in the absence of network effects, it is the constant. As above, $\boldsymbol{z}_{j,r}$ is a vector of teacher characteristics and $\bar{\boldsymbol{z}}_{j,r}$ is a vector of average peers' teacher features. A teacher was considered to have a link with another one if they were in the same school, regardless of school's board (network with higher density).

$R_{j,r,k}$ can be one of the following variables: time spent with teaching activities (planning, test correction etc., $k = 1$), duties (high frequency of bureaucratic activities, $k = 2$) and use of equipment and school facilities (computers, DVDs, newspapers and magazines, laboratories, reading room and books, $k = 3, ..., 8$). If the negative endogenous effects are explained by jointed production and shared results, $\rho_{1,1} < 0$, $\rho_{1,2} > 0$ and $\rho_{1,l} < 0$, $l = 3, ..., 8$.

As there are no evidence that bigger mean peers' VA is associated with more duties, less time spent with pedagogic activities and less use of resources, we conclude that the negative endogenous effects are not due to the investigated source.

Table 1.15 – Source of teacher spillovers - Reading

	OLS		School FE		RB FE	
	Coef	SE	Coef	SE	Coef	SE
> 8 week-hours spent with pedagogic activities	.01	.01	-.001	.004	.004	.009
High frequency of bureaucratic activities	-.02	.013	.005	.005	-.022	.013
Computer	.011	.008	-.006	.004	.016	.008
DVD	.02	.018	-.003	.008	.02	.018
Newspaper and magazines	.054	.023	.012	.01	.051	.023
Laboratories	.009	.009	.004	.004	.011	.009
Reading room	.025	.015	.003	.008	.028	.015
Books for children	.037	.028	-.001	.01	.045	.029

Notes: Standard errors (SE) are robust.

Table 1.16 – Source of teacher spillovers - Mathematics

	OLS		School FE		RB FE	
	Coef	SE	Coef	SE	Coef	SE
> 8 week-hours spent with pedagogic activities	.017	.009	0	.004	.012	.009
High frequency of bureaucratic activities	-.007	.013	-.001	.006	-.01	.013
Computer	.003	.008	-.004	.004	.009	.008
DVD	.032	.019	.008	.009	.03	.019
Newspaper and magazines	-.001	.024	-.013	.01	-.001	.024
Laboratories	.013	.009	.005	.004	.017	.009
Reading room	.01	.016	-.001	.007	.01	.016
Books for children	.043	.028	-.001	.01	.051	.028

Notes: Standard errors (SE) are robust.

Tables 1.17 to 1.18 report the models that use past instead of contemporaneous mean VA of peers and they show that there are no social effects at all. It is interesting to contrast the *OLS* and *RB FE* models with the *School FE*. When we compare teachers in general or those in the same RB, the endogenous effects are positive; when we compare only teachers in same school, the sign is reversed. The same happens for some contextual effects for Reading, as gender, other job or school system and, mother's education. The effect of postgraduate is negative in the *OLS* and *RB FE* models but positive in the *School FE* in both subjects. The single explanation that we can think of is small sample size that makes our results too sensible to the inclusion of school fixed effects. The number of teachers extremely reduced occurred because, as described above, we merge data from four editions of *PSP*.

Table 1.17 – Peer effects models with past VA of peers - Reading - Basic VA

	OLS		School FE		RB FE	
	Coef	SE	Coef	SE	Coef	SE
Mean past VA of peers	.122	.241	-.167	.43	.122	.283
Peers proportion (PP) of women	.024	.451	-.109	2.003	.028	.533
PP 10 years of experience or less	-.042	.368	-.194	2.081	-.038	.429
PP 11 to 20 years of experience	-.025	.163	-.223	1.351	-.024	.198
PP with postgraduate	-.045	.186	.215	1.653	-.053	.205
PP other job or school system	.05	.166	-.055	1.561	.045	.19
White peers proportion	-.002	.181	-.069	1.542	-.015	.209
PP mothers with middle school	-.018	.24	-.11	2.215	-.029	.275
PP mothers with high school	.094	.276	-.321	2.243	.079	.321
PP mothers with higher education	.104	.326	-.364	1.986	.096	.406
Observations	340	.	340	.	340	.
F-statistic	.978	.	.981	.	.922	.

Notes: Standard errors (SE) were calculated by bootstrap. All regressions control for teachers' own characteristics.

Table 1.18 – Peer effects models with past VA of peers - Mathematics - Basic VA

	OLS		School FE		RB FE	
	Coef	SE	Coef	SE	Coef	SE
Mean past VA of peers	.041	.219	-.077	.355	.029	.245
Peers proportion (PP) of women	-.05	.449	-.169	2.192	-.04	.544
PP 10 years of experience or less	.004	.429	-.113	2.26	.016	.484
PP 11 to 20 years of experience	-.014	.178	-.215	1.503	-.008	.203
PP with postgraduate	-.065	.179	.073	1.678	-.074	.198
PP other job or school system	-.029	.171	-.023	1.74	-.009	.194
White peers proportion	-.006	.171	-.062	1.857	-.027	.204
PP mothers with middle school	-.007	.266	-.261	2.36	-.005	.291
PP mothers with high school	.021	.279	.11	2.13	.032	.334
PP mothers with higher education	.053	.325	-.42	2.175	.05	.394
Observations	340	.	340	.	340	.
F-statistic	.688	.	.793	.	.936	.

Notes: Standard errors (SE) were calculated by bootstrap. All regressions control for teachers' own characteristics.

Despite their statistical insignificance, the Reading endogenous effect is economically relevant, as it stands for 12.2% of teacher total effect. The Math effect stands for 2.9% again. Some exogenous effects of the *School FE* model have notable size, between 6.8% and 16.8%, like experience, mother's education for both disciplines, postgraduate for Reading and, gender for Reading.

As discussed above, the second approach estimates the endogenous effect times the coefficient of a linear projection of the mean peers' VA on the proxy. Thus, the sign change indicates that $\theta_1 < 0$.

1.5 Final remarks

Taking into account the aspects of the educational production process - output generated by the combined effort of many employees and the presence of difficulties to identify and reward the contribution made by each worker - it is reasonable to believe that peer effects must matter to explain student learning. This paper investigated the role of direct peer effects to teacher quality, that is, the impact of the average peers' features, including effectiveness, on teacher quality (and not on her students), taking out spillovers sources not related to the contact between teachers.

The teacher quality was measured by their value-added, assuming that a baseline achievement measure is a sufficient statistic for unobserved input histories and child's innate ability. To deal with the issues associated with peer effects estimation, we use two approaches that could solve the reflection problem and the perfect collinearity between

mean outcome and mean characteristics of peers - instrumental variables and out of sample estimates of teacher quality - and treat the self-selection into reference groups using network or schools fixed effects. Using data from *Prova Sao Paulo* and School Census, our samples are composed of teachers of elementary students from Sao Paulo City system, a big Brazilian educational system.

Unfortunately, the first stages of the IV approach were not good enough to guarantee a clean identification and we believe that the weakness of the instruments are related to the extremely low network density and the lack of intransitive triads. We think that is the reason why we were not able to capture statistical significance. Besides, the use of past value added of peers as a proxy to average peer quality reduced excessively the teacher sample in the second approach.

Our first empirical strategy provides evidence that more effective peers reduce teacher contribution to the student learning (in an economically relevant amount for Reading but not for Mathematics), and that other peers' features do not matter in statistical and economic terms. We investigated if the negative endogenous effects were explained by a jointed production and shared resources story, the only feasible source between the ones considered in this article. We concluded that this was not the source of endogenous effects.

The second approach brought different results. The endogenous effects are positive in most models and, for Reading, they are sizable in economic terms. Some contextual effects presented notable economic sizes, such as mothers' education and experience, when we included school fixed effects. The difference between the two methods can be understand when we reconize that the Jackson e Bruegmann (2009)'s approach uses a proxy variable to deal with the reflection problem and estimates the endogenous effect times the (negative) partial correlation between the proxy and the average peers' quality.

Our chose to use a linear-in-means model has a clear disadvantage. As discussed by Sacerdote (2011), this model constrains the net effect from reassignment of peers to different reference groups to be zero. This restriction prevent us to discuss how to allocate teachers to improve mean teacher quality, although we could discuss how to reduce dispersion. Thus, a next step would be to estimate non-linear models to address social welfare questions that our linear models do not allow.

Additionally, we will search ways to improve the features of the teachers network - increase the network density and try to create reasonable intransitive triads - and to raise sample size of teachers with students in the various editions of PSP. We would like to improve the treatment gave to the endogenous formation of the reference groups. The use of fixed effects has a clear limit since it controls only for time fixed unobserved variables.

1.A Appendix A

Table 1.19 – Descriptive statistics of students used to estimate past teacher value added to sample 2010-2007

	PSP 2008 1		PSP 2008 2		PSP 2008-2007 1		PSP 2008-2007 2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
2008 raw Reading score	150.45	46.50	150.46	46.55	140.35	43.77	144.25	44.38
2008 raw Mathematics score	162.27	44.04	162.26	44.09	153.79	45.51	156.84	46.83
Age	9.54	1.31	9.53	1.31	10.30	1.17	10.14	1.14
2nd grade in 2008	0.39	0.49	0.40	0.49	0.00	0.05	0.00	0.04
3rd grade in 2008	0.12	0.33	0.11	0.32	0.68	0.47	0.72	0.45
4th grade in 2008	0.49	0.50	0.49	0.50	0.32	0.47	0.28	0.45
Proportion of girls	0.49	0.50	0.49	0.50	0.43	0.50	0.45	0.50
Prop. of students with age-grade distortion	0.07	0.26	0.07	0.25	0.19	0.39	0.15	0.36
Prop. of pupils in IPE	0.08	0.26	0.08	0.26	0.45	0.50	0.50	0.50
Prop. of students in morning shift	0.49	0.50	0.49	0.50	0.52	0.50	0.54	0.50
... intermediate shift	0.13	0.33	0.13	0.33	0.09	0.28	0.08	0.28
... afternoon shift	0.38	0.49	0.38	0.49	0.39	0.49	0.38	0.48
2007 raw Reading score					122.80	37.50	125.30	38.23
2007 raw Mathematics score					131.51	39.78	134.39	40.84
2nd grade in 2007					0.68	0.47	0.72	0.45
4th grade in 2007					0.32	0.47	0.28	0.45
Proportion of school movers					0.02	0.15	0.02	0.15
Prop. of students that repeated 2007's grade					0.32	0.47	0.28	0.45
Observations	114050		112959		14114		4583	

Notes: PSP 2008 1 contains all students with basic information (test score, age, gender, age-grade distortion, IPE, shift) and took the test in 2008; PSP 2008 2 is similar to PSP 2010 1, but it restricts to those with teachers with at least 5 pupils; PSP 2008-2007 1 contains students from teachers with at least 5 pupils that had basic information plus turnover and repetition and took the test in both years; PSP 2008-2007 2 is similar to PSP 2008-2007 1, but pupils had basic plus preschool, parents' education and family's income and size. SD means standard deviation and MW minimum wage. Intensive Project for Elementary Students (IPE) aims to help those not fully literate.

Table 1.20 – Value added model - Reading

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Basic		Basic+Parent		Basic+Parent+School
2009 Reading score	0.654***	0.640***	0.609***	0.609***	0.603***
	(0.00543)	(0.00872)	(0.00994)	(0.0127)	(0.00997)
Age	-0.0289***	-0.0222***	-0.0163	-0.00267	-0.0145
	(0.00573)	(0.00701)	(0.0110)	(0.0131)	(0.0110)
Girl	0.0992***	0.0989***	0.0819***	0.0779***	0.0817***
	(0.00837)	(0.00851)	(0.0150)	(0.0150)	(0.0150)
Intensive Project for Elementary Students (IPE)	-0.114***	-0.116***	-0.0216	-0.00696	-0.0212
	(0.0143)	(0.0265)	(0.0283)	(0.0480)	(0.0302)
Intermediate shift	-0.0721***	-0.0556	0.00913	0.0496	0.0232
	(0.0192)	(0.0350)	(0.0460)	(0.0574)	(0.0460)
Afternoon shift	0.0320***	0.0125	0.0394***	0.0700**	0.0388**
	(0.00846)	(0.0180)	(0.0150)	(0.0335)	(0.0151)
School mover	-0.0613***	-0.0279	0.00661	0.00389	0.0215
	(0.0231)	(0.0302)	(0.0447)	(0.0602)	(0.0448)
Repeated last grade	0.155***	0.134***	0.0953***	0.0740**	0.0860***
	(0.0155)	(0.0205)	(0.0288)	(0.0337)	(0.0290)
At least one parent studied at least one year of middle school			0.0559**	0.0516**	0.0488**
			(0.0241)	(0.0257)	(0.0241)
... at least one year of high school			0.137***	0.131***	0.126***
			(0.0237)	(0.0258)	(0.0237)
... at least one year of higher education			0.237***	0.206***	0.206***
			(0.0320)	(0.0344)	(0.0322)
Family income between 2 and 5 MW			0.0756***	0.0584***	0.0618***
			(0.0160)	(0.0157)	(0.0162)
Family income over 5 MW			0.136***	0.103***	0.106***
			(0.0306)	(0.0289)	(0.0308)
4 people			-0.0144	-0.0124	-0.00985
			(0.0222)	(0.0234)	(0.0221)
5 or more people			-0.0729***	-0.0642***	-0.0658***
			(0.0205)	(0.0207)	(0.0205)
Preschool			0.00940	0.00139	0.00709
			(0.0148)	(0.0146)	(0.0147)
Proportion of girls at school					0.260
					(0.220)
Prop. of PICs students at school					0.0980
					(0.131)
Prop. of students (PS) with age-grade distortion at school					0.110
					(0.259)
PS whose parent studied at least one year of middle school					0.184
					(0.258)
PS whose parent studied at least one year of high school					0.0224
					(0.179)
PS whose parent studied at least one year of higher education					0.559**
					(0.263)
PS with family income between 2 to 5 MW at school					0.164
					(0.134)
PS with family income 5 MW and over at school					0.472*
					(0.257)
Observations	31,010	31,010	8,948	8,948	8,948
R-squared	0.479	0.456	0.482	0.456	0.486
F-test	2649.783	1048.081	417.127	243.701	283.06
Number of schools		535		477	
School FE		Yes		Yes	

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.21 – Value added model - Mathematics

VARIABLES	(1) Basic	(2)	(3) Basic+Parent	(4)	(5) Basic+Parent+School
2009 Mathematics score	0.680***	0.667***	0.634***	0.632***	0.629***
	(0.00525)	(0.00735)	(0.00946)	(0.0114)	(0.00946)
Age	-0.0496***	-0.0433***	-0.0349***	-0.0269*	-0.0335***
	(0.00558)	(0.00726)	(0.0108)	(0.0147)	(0.0109)
Girl	-0.00811	-0.0106	-0.0328**	-0.0317**	-0.0339**
	(0.00823)	(0.00830)	(0.0151)	(0.0153)	(0.0151)
Intensive Project for Elementary Students (IPE)	-0.112***	-0.119***	-0.0314	-0.0347	-0.0439
	(0.0139)	(0.0244)	(0.0284)	(0.0425)	(0.0298)
Intermediate shift	-0.0652***	-0.0350	0.00433	0.0347	0.0211
	(0.0192)	(0.0346)	(0.0453)	(0.0575)	(0.0455)
Afternoon shift	-0.00192	-0.0250	-0.000492	-0.00402	-0.00340
	(0.00842)	(0.0189)	(0.0152)	(0.0365)	(0.0153)
School mover	-0.0907***	-0.0689**	0.0121	0.0377	0.0203
	(0.0209)	(0.0327)	(0.0399)	(0.0600)	(0.0401)
Repeated last grade	0.264***	0.247***	0.204***	0.180***	0.197***
	(0.0152)	(0.0203)	(0.0288)	(0.0360)	(0.0290)
At least one parent studied at least one year of middle school			0.0545**	0.0464*	0.0501**
			(0.0246)	(0.0237)	(0.0246)
... at least one year of high school			0.127***	0.114***	0.122***
			(0.0242)	(0.0235)	(0.0243)
... at least one year of higher education			0.211***	0.190***	0.189***
			(0.0320)	(0.0320)	(0.0324)
Family income between 2 and 5 MW			0.0650***	0.0500***	0.0496***
			(0.0163)	(0.0156)	(0.0165)
Family income over 5 MW			0.132***	0.102***	0.106***
			(0.0312)	(0.0335)	(0.0314)
4 people			-0.0190	-0.0253	-0.0143
			(0.0228)	(0.0239)	(0.0228)
5 or more people			-0.0813***	-0.0792***	-0.0745***
			(0.0210)	(0.0218)	(0.0210)
Preschool			0.0177	0.0168	0.0168
			(0.0149)	(0.0151)	(0.0149)
Proportion of girls at school					-0.00783
					(0.225)
Prop. of PICs students at school					0.227*
					(0.129)
Prop. of students (PS) with age-grade distortion at school					0.393
					(0.265)
PS whose parent studied at least one year of middle school					0.0733
					(0.273)
PS whose parent studied at least one year of high school					-0.155
					(0.177)
PS whose parent studied at least one year of higher education					0.390
					(0.267)
PS with family income between 2 to 5 MW at school					0.447***
					(0.136)
PS with family income 5 MW and over at school					0.272
					(0.259)
Observations	31,010	31,010	8,948	8,948	8,948
R-squared	0.483	0.462	0.485	0.462	0.487
F-test	2608.484	1212.251	403.821	255.998	272.612
Number of schools		535		477	
School FE		Yes		Yes	

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.22 – Past value added model - Reading

VARIABLES	(1) Basic	(2)	(3) Basic+Parent	(4)	(5) Basic+Parent+School
2007 Reading score	0.556***	0.526***	0.501***	0.477***	0.483***
	(0.00789)	(0.0112)	(0.0142)	(0.0195)	(0.0143)
Girl	0.0956***	0.0946***	0.0914***	0.0893***	0.0973***
	(0.0126)	(0.0134)	(0.0221)	(0.0236)	(0.0221)
Age	-0.0966***	-0.0804***	-0.0661***	-0.0709***	-0.0598***
	(0.00870)	(0.00946)	(0.0163)	(0.0179)	(0.0163)
Intensive Project for Elementary Students (IPE)	-0.333***	-0.371***	-0.401***	-0.481***	-0.445***
	(0.0161)	(0.0315)	(0.0305)	(0.0571)	(0.0320)
Intermediate shift	-0.0578**	-0.0263	-0.0821*	-0.114	-0.0916**
	(0.0237)	(0.0527)	(0.0430)	(0.106)	(0.0431)
Afternoon shift	0.0245*	0.0707**	0.0381*	0.112*	0.0315
	(0.0130)	(0.0354)	(0.0229)	(0.0592)	(0.0230)
Proportion of school movers	-0.0613	0.0269	-0.0729	0.00315	-0.0463
	(0.0424)	(0.0613)	(0.0761)	(0.104)	(0.0770)
Repeated last grade	-0.00291	0.0114	0.0156	0.0761	0.0248
	(0.0233)	(0.0327)	(0.0456)	(0.0637)	(0.0453)
At least one parent studied at least one year of middle school			-0.00720	-8.28e-05	-0.00290
			(0.0346)	(0.0348)	(0.0346)
... at least one year of high school			0.0798**	0.0658*	0.0774**
			(0.0363)	(0.0353)	(0.0365)
... at least one year of higher education			0.141***	0.127***	0.140***
			(0.0370)	(0.0382)	(0.0374)
Family income between 2 and 5 MW			0.0396	0.0320	0.0126
			(0.0282)	(0.0260)	(0.0285)
Family income over 5 MW			0.187***	0.117*	0.142**
			(0.0651)	(0.0698)	(0.0651)
4 people			-0.0358	-0.0175	-0.0284
			(0.0337)	(0.0334)	(0.0335)
5 or more people			-0.122***	-0.0917***	-0.112***
			(0.0306)	(0.0305)	(0.0304)
Preschool			0.134***	0.116***	0.125***
			(0.0271)	(0.0274)	(0.0270)
Proportion of girls at school					-0.0873
					(0.347)
Prop. of PICs students at school					0.796***
					(0.184)
Prop. of students (PS) with age-grade distortion at school					-0.579
					(0.375)
PS whose parent studied at least one year of middle school					-0.510
					(0.439)
PS whose parent studied at least one year of high school					-0.573
					(0.365)
PS whose parent studied at least one year of higher education					-0.827***
					(0.277)
PS with family income between 2 to 5 MW at school					0.907***
					(0.240)
PS with family income 5 MW and over at school					1.247***
					(0.360)
Observations	14,114	14,114	4,583	4,583	4,583
R-squared	0.505	0.439	0.523	0.413	0.530
F-test	1877.7	698.4	334.5	134.7	230.7
Number of schools		468		352	
School FE		Yes		Yes	

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.23 – Past value added model - Mathematics

VARIABLES	(1) Basic	(2)	(3) Basic+Parent	(4)	(5) Basic+Parent+School
2007 Mathematics score	0.541***	0.523***	0.503***	0.483***	0.491***
	(0.00798)	(0.0109)	(0.0151)	(0.0181)	(0.0152)
Girl	-0.0434***	-0.0483***	-0.0554**	-0.0626***	-0.0531**
	(0.0134)	(0.0138)	(0.0238)	(0.0215)	(0.0238)
Age	-0.0832***	-0.0663***	-0.0532***	-0.0441**	-0.0498***
	(0.00957)	(0.0111)	(0.0184)	(0.0207)	(0.0184)
Intensive Project for Elementary Students (IPE)	-0.415***	-0.437***	-0.416***	-0.468***	-0.446***
	(0.0177)	(0.0333)	(0.0343)	(0.0599)	(0.0356)
Intermediate shift	-0.00731	0.0545	0.0150	-0.0769	-0.00542
	(0.0255)	(0.0618)	(0.0474)	(0.142)	(0.0479)
Afternoon shift	0.0337**	0.0769	0.0564**	0.0941	0.0529**
	(0.0141)	(0.0489)	(0.0251)	(0.0794)	(0.0253)
Proportion of school movers	-0.103**	0.0356	-0.0762	0.162	-0.0536
	(0.0407)	(0.0670)	(0.0700)	(0.110)	(0.0704)
Repeated last grade	-0.0346	-0.0433	-0.0525	-0.0316	-0.0432
	(0.0261)	(0.0376)	(0.0528)	(0.0696)	(0.0523)
At least one parent studied at least one year of middle school			-0.0405	-0.0319	-0.0477
			(0.0377)	(0.0344)	(0.0379)
... at least one year of high school			0.0393	0.0450	0.0339
			(0.0394)	(0.0378)	(0.0397)
... at least one year of higher education			0.0923**	0.0609	0.0758*
			(0.0396)	(0.0398)	(0.0403)
Family income between 2 and 5 MW			0.00111	0.0169	-0.0209
			(0.0290)	(0.0306)	(0.0292)
Family income over 5 MW			0.215***	0.212***	0.185***
			(0.0562)	(0.0518)	(0.0570)
4 people			-0.00737	-0.0119	-0.00215
			(0.0360)	(0.0356)	(0.0361)
5 or more people			-0.0752**	-0.0799**	-0.0708**
			(0.0332)	(0.0333)	(0.0333)
Preschool			0.160***	0.149***	0.152***
			(0.0301)	(0.0307)	(0.0302)
Proportion of girls at school					0.0482
					(0.367)
Prop. of PICs students at school					0.683***
					(0.196)
Prop. of students (PS) with age-grade distortion at school					-0.389
					(0.409)
PS whose parent studied at least one year of middle school					0.377
					(0.469)
PS whose parent studied at least one year of high school					-0.713*
					(0.388)
PS whose parent studied at least one year of higher education					0.243
					(0.299)
PS with family income between 2 to 5 MW at school					0.268
					(0.259)
PS with family income 5 MW and over at school					0.651*
					(0.370)
Observations	14,114	14,114	4,583	4,583	4,583
R-squared	0.469	0.407	0.478	0.379	0.482
F-test	1653.6	570	293.7	130.5	202.5
Number of schools		468		352	
School FE		Yes		Yes	

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.24 – Correlations between value added estimates - sample 2010-2009

Model		Reading 1	2	3	4	5	Mathematics 1	2	3	4	5
Reading	1	1	.999	.832	.83	.807	.545	.548	.448	.451	.427
	2	.999	1	.832	.83	.806	.548	.553	.451	.455	.429
	3	.832	.832	1	.998	.988	.45	.453	.473	.475	.461
	4	.83	.83	.998	1	.985	.45	.453	.473	.475	.46
	5	.807	.806	.988	.985	1	.425	.428	.462	.462	.468
Math	1	.545	.548	.45	.45	.425	1	.999	.858	.86	.839
	2	.548	.553	.453	.453	.428	.999	1	.858	.86	.838
	3	.448	.451	.473	.473	.462	.858	.858	1	1	.991
	4	.451	.455	.475	.475	.462	.86	.86	1	1	.989
	5	.427	.429	.461	.46	.468	.839	.838	.991	.989	1

Notes: Reading 1 refers to model 1 from table 1.20 and so on; Math 1 refers to model 1 from table 1.21 and so on. All correlations are significant at 5%.

Table 1.25 – Correlations between past value added estimates - sample 2010-2007

Model		Reading 1	2	3	4	5	Mathematics 1	2	3	4	5
Reading	1	1	.993	.889	.874	.859	.574	.577	.517	.512	.493
	2	.993	1	.89	.887	.857	.578	.59	.521	.525	.496
	3	.889	.89	1	.991	.974	.521	.524	.531	.532	.511
	4	.874	.887	.991	1	.963	.518	.529	.529	.544	.508
	5	.859	.857	.974	.963	1	.502	.503	.521	.521	.529
Math	1	.574	.578	.521	.518	.502	1	.995	.895	.888	.874
	2	.577	.59	.524	.529	.503	.995	1	.893	.892	.871
	3	.517	.521	.531	.529	.521	.895	.893	1	.993	.988
	4	.512	.525	.532	.544	.521	.888	.892	.993	1	.979
	5	.493	.496	.511	.508	.529	.874	.871	.988	.979	1

Notes: Reading 1 refers to model 1 from table 1.20 and so on; Math 1 refers to model 1 from table 1.21 and so on. All correlations are significant at 5%.

2 Impacts of a large-scale literacy program on student learning

Abstract

This article intends to study a large-scale literacy intervention - the Literacy Program at the Right Age (LPRA) - implemented in Ceara, a low-income state of Brazil. This program seems to gather many of the essential features pointed by the literature as mediators to policy success, such as professional development, assessment, accountability, and an institutional structure to support local governments. Using the natural experiment method alone and combined with matching, we found significant effects on mean test scores from cohorts of fifth-graders for both subjects. The estimated overall effects did not rise with time as expected in view of the increase in treatment intensity, but we found an increasing effect for low-achievement students, that took more time of exposure to the program to catch up with the rest. The program did not affect differently schools with distinct proportions of female, black and poorer students or those who attended preschool.

Keywords: literacy, policy evaluation, Literacy Program at the Right Age.

2.1 Introduction

Nowadays, literacy is not defined as knowing how to read and write anymore[1], it became more complex and advanced than these cognitive skills that result from the simple acquisition of the written system. In Brazil, Soares (2004) defines literacy[2] as the social practice of reading and writing. This definition is closely related to the UNESCO (2005)'s that understands (functionally) literacy as crucial to the acquisition of life skills that enable people to address the challenges they can face in life, and represents a necessary step in basic education, which is an indispensable means for effective participation in today's societies and economies. Literacy and schooling[3] are associated to different kinds of benefits, from intrinsically valuable and instrumental effects, as improved self-esteem and empowerment, to political and economic impacts, as increased political participation and individual earnings (UNESCO, 2005).

The main route to achieving literacy is through elementary school (UNESCO, 2005), and this schooling cycle can be split into two periods regarding the stages of reading and writing acquisition, according to Slavin et al. (2009). These authors named these

1 What is called beginning literacy in English and *alfabetização* in Brazil (SOARES, 2004).
2 The term in Portuguese is *Letramento*.
3 As literacy and schooling are deeply connected, it is hard to separate their benefits.

stages as the beginning and the upper elementary reading and placed them, respectively, from kindergarten to first grade (K-1), and from second to fifth-grade. In the K-1 period, children must learn the basic skills of turning print into meaning, that is, know the sounds of all letters, form them into words, recognize the most common sight words, read and comprehend simple texts; in the upper elementary stage, the focus is to consolidate and extend the basic skills, as well as, build fluency, comprehension, and vocabulary for reading more and more complex texts from different genres (SLAVIN et al., 2009).

Notwithstanding, what happens before kindergarten matters. NRC (1998) note that the less prior knowledge and certain skills the children have when they begin school - particularly letter knowledge, phonological awareness, familiarity with the basic purposes and mechanisms of reading, and language ability -, the more difficulty they have learning to read in the primary grades. The majority of the pupils that struggles to learn come from low-income families, who also live in low-income areas, who attend schools with a high concentration of low achievement students and whose parents have a history of reading problems (NRC, 1998).

The literature that investigates reading ability found that the performance differences between good and poor readers may increase over time (CAIN; OAKHILL, 2011). This cumulative phenomenon, called Matthew effects[4], on the one hand, makes children who are reading very well increase their vocabulary what, in turns, facilitates their reading comprehension, and further improve their reading; on the other hand, children who read slowly and without enjoyment maintain inadequate vocabulary and comprehension, inhibiting further growth in reading ability (STANOVICH, 2009). A previously existing knowledge base that is rich and elaborated facilitates further learning because new educational experiences are used more efficiently, which explains why Matthew effects exist (STANOVICH, 2009).

Slavin et al. (2009) reviewed research on the learning impacts of reading programs for elementary students and identified common characteristics of interventions likely to improve reading achievement, including those in high-poverty schools. These authors found that successful programs give extensive professional development and follow-up with a focus on specific instructional methods to teachers[5]. Interventions that adopt phonetic curricula alone, for example, but do not provide teachers with practical strategies for teaching reading had small impacts on students outcomes.

Summing up, it is tough to improve literacy of all children regardless of their socioeconomic status, and literacy is crucial for lifelong learning. A successful intervention should not just be based on the scientific knowledge about how and when children learn

[4] This name came "after the Gospel according to Matthew: “For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath” (XXV: 29)" (STANOVICH, 2009).

[5] Such as teaching phonics and phonemic awareness (SLAVIN et al., 2009).

to read, but it also must incorporate other operational features such as professional development, assessment, and choice of textbooks (SLAVIN et al., 2009). Institutional and political features even matter, especially when many actors (federal, state e local governments, schools, teachers) are involved and the policy is formulated by different actors from those that implement it, implicating that the policy implementation must be carefully thought (SEGATTO, 2012). We believe that the failure in properly teach literacy at the right age, considering and compensating the disadvantages of low-income children, can trigger Matthew effects, compromise the developmental course of reading progress, and thus the quality of life.

This article intends to study a large-scale literacy intervention - the Literacy Program at the Right Age (LPRA)[6] - implemented in Ceara, a low-income state of Brazil. This policy seems to gather many of the essential features described above, as professional development, assessment, accountability, and an institutional structure to support local governments, especially those lacking the technical and financial capacity to manage their educational systems. Besides, Ceara has excelled in educational outcomes, as showed by the national evaluations of student learning. The state's student achievement has continually improved in third, fifth and ninth grades and overcame Brazil's and the other Northeastern states altogether in the most recent assessments.

The literature had evaluated the effect of LPRA on the first treated cohort - students that took part in the program in 2008 at second grade and participated in national evaluation system in 2011 at fifth grade - and found out a positive impact on reading (COSTA; CARNOY, 2015; LAVOR; ARRAES, 2014; KASMIRSKI; GUSMAO; RIBEIRO, 2015), and mathematics achievement (COSTA; CARNOY, 2015). All these papers take advantage of policy variation across time (it started in 2007) and space (only Ceara adopted a universal literacy intervention between the northeastern states) to identify its effects. Costa e Carnoy (2015) also used the variation by grade to evaluate the LPRA and applied triple differences to assure that the improvement in fifth-grade students was not due to other reforms in Ceara. Their results showed significant gains in the average Reading performance and even greater ones in Mathematics of fifth graders. They reported that students with higher test scores experimented larger increases in learning and those who had had or not had early childhood education differed little.

Kasmirski, Gusmao e Ribeiro (2015) investigated if the program is related to the equity[7] improvement detected in the state with a difference-in-difference approach. Their results showed that LPRA was responsible for at least 40% of equity increase observed for the last grade of elementary school in Ceara. Lavor e Arraes (2014) used the synthetic

6 Programa Alfabetização na Idade Certa (Paic).

7 Equity was defined as the situation in which all students, regardless of their socioeconomic status, achieve adequate levels of learning. This concept is based on the justice principle proposed by Marcel Crahay called basic equality (KASMIRSKI; GUSMAO; RIBEIRO, 2015).

control method and also found a positive effect of the program on 5th-grade reading achievement. Thus the impact of the first year of treatment is already documented, but no study verified the effect of LPRA on the subsequent treated cohorts. This paper aims to estimate precisely these impacts in addition to spillover and heterogeneous effects using the difference-in-difference method alone and combined with matching methods.

Although the initial program focus was only literacy, we believe there are spillover effects into mathematics. Greater reading ability can lead to improvements in mathematics, and Education, Psychology and Neuroscience literature investigate the link between these subjects. Grimm (2008) presents explanations that support reading ability as a precursor in the learning of mathematics: pre-reading vocabulary skills could shape the development of number concepts and influence on at least some aspects of numeracy, and successful mathematical problem solving may depend on the reading comprehension.

We found significant and positive effects on mean test scores of all investigated cohorts of fifth-graders for both subjects. The estimated overall effects did not rise with time as expected in view of the increase in treatment intensity, but we found an increasing effect for low-achievement students, that took more time of exposure to the program to catch up with the rest. The program did not affect differently schools with distinct proportions of female, black and poorer students or those who attended preschool.

This paper is organized as follows: next section describes the program, the third section presents data and samples used, the fourth section presents our empirical strategy, part 5 shows the results, section 6 discuss the identification hypothesis of the methods applied and the final section concludes.

2.2 Program's description

The Literacy Program at the Right Age (LPRA), created in 2007[8] and implemented in all Ceara public schools, is defined as a collaboration policy between the state and the municipalities supported by Ceara State government and partners[9]. Its original aim was to teach literacy to all public school students in the Ceara until they are seven years old (until the second grade of the elementary school in the system with nine years) (SEDUC-CE, 2012). In 2011, the government expanded the intervention with the creation of the Learning Program at the Right Age (LPRA+5) that included third, fourth and fifth grades besides mathematics subject for these grades; later, in 2015, the public administration launched the More Learning Program at the Right Age (More LPRA) to include all grades of middle school (SEDUC-CE, 2019).

8 The program's origin is dated 2004 with the work developed by Ceara Committee on the Elimination of School Illiteracy. For more information see SEDUC-CE (2012).

9 United Nations Children's Fund (UNICEF) and civil society institutions (SEDUC-CE, 2012).

The state government created central and regional departments in 2007 to cooperate with the municipalities[10] and established staffs to work exclusively with the program inside all these departments and municipal Education Offices. Through the city teams, the state gives financial and technical aid to implement and monitor the LPRA (SEDUC-CE, 2012).

The program started with five overall goals: eliminate child illiteracy, provide an external evaluation of literacy, reading promotion, expand the preschool quality and attendance, and strengthening municipal management (SEDUC-CE, 2012). The state and municipalities offices developed a teacher training program articulated with the curriculum, teacher and student materials to accomplish the first goal. Based on a previous diagnosis, the state office prepared the first-grade and selected the second-grade materials among those already available by publishing companies in 2007 and 2008 (SEDUC-CE, 2012; SEDUC-CE, 2015).

Concerning the external evaluation, in 2007, two tests were created to diagnose the second-grade student level of begging reading: the LPRA test, applied at the beginning of the school year, aiming to allow time for educational interventions to take place, and the standardized *Spaece Alfa* test that occurs at the ending of the school year (SEDUC-CE, 2012). This last test joined the state system of learning evaluation to assess reading proficiency (SEDUC-CE, 2012). There are two additional policies to incentive municipal and school management to prioritize child literacy in 2009. Based on *Spaece Alfa* scores, the state office calculates indices[11] that determine the distribution share of the state tax on the goods and services circulation to cities and prizes to schools[12] (SEDUC-CE, 2012).

As regarding the reading promotion, all preschool, first and second grades classrooms gained literary collections to stimulate the taste for reading (SEDUC-CE, 2012), which helps pupils to become fluent readers and master skills such as build a vocabulary of words and concepts (SLAVIN et al., 2009). The state has three strategies to expand the preschool quality and attendance - training of municipal office technicians, who should structure training of child education professionals; help the city offices to develop a pedagogical proposal for children until five years old; and providing financial help to build schools. The federal government and another Ceara state office also had initiatives to open more schools reinforcing this last program strategy[13] (SEDUC-CE, 2012).

The strengthening municipal management goal aims to spread a systemic and effective management culture to the cities and develop actions to consolidate the ability

[10] Local governments must provide primarily preschool and elementary education according to Brazil's Constitution, and many cities face financial and technical issues to manage their educational systems.

[11] The indices are presented in CEARA (2009).

[12] Prêmio Escola Nota Dez.

[13] The federal intervention is the National Program for the Restructuring and Acquisition of Equipment for Public Schools of Early Childhood Education (*Proinfancia*), and the state intervention is the Program to Support Social Reforms (*Proares*), from Labor and Social Development Office of Ceara.

to diagnose, planning and monitoring, and rearrange processes in municipal offices. Of great importance are the actions that intend to raise instructional time, such as better control of student and teacher absences, and fulfillment of school calendar (SEDUC-CE, 2012).

After the LPRA expansions - LPRA+5 and More LPRA - the goal eliminate child illiteracy was reformulated to accommodate the specificity of each educational level and was split into elementary and middle school. The LPRA actions arrived in the classroom effectively in 2008, as said by SEDUC-CE (2015). In 2007, the government implemented the institutional structure required by the policy such as the creation of the cooperation departments and the establishment of the literacy evaluations and then, in 2008, it started training teachers, applying the LPRA test, buying books and textbooks. Due to the lack of public information, we assume that something similar occurred when the program was expanded in 2011 and 2015 - in the first year the government mainly rearranged the institutional design, and in the second year the actions could be implemented in the classrooms.

To sum up, this complex policy yielded an institutional structure that allows monitoring processes at many levels (classrooms, schools, city and state offices) and assesses the results, mainly through external evaluation. It was implemented gradually starting at the beginning of elementary school with more straightforward goals, such as put an end in child illiteracy, and then it evolved adding objectives for higher grades that became possible as the fundamental issues were solved.

In 2005 and 2006, there was a first phase of the LPRA at 56 cities and, despite this phase be sometimes called pilot, during this period the main action developed was the literacy diagnosis evaluation (SEDUC-CE, 2012). The program actually begin in 2007 as described above and table 2.28 in the appendix shows no effects of this phase on students from these cities[14].

Finally, we must remember that, in 2012, the federal government launched the National Pact of Literacy at the Right Age, whose classroom interventions started in 2013 in 5,420 out of 5,570 Brazilian cities (MEC, 2015). Which means that a national reading program treated other Brazilian states after 2012 and this could have affected the control states time trend.

[14] Second-grade students in 2006 from the first-phase cities arrived to fifth-grade in 2009.

2.3 Data and samples

We utilize data from the National Evaluation System of Basic Education (NES) provided by the National Institute for Educational Studies and Research *Anísio Teixeira* (INEP). According to INEP (2019), the NES started in 1990 and since 1995 uses Item Response Theory - a methodology to construct the tests and analyze their results that allows comparability over time. It always takes place in odd years, assesses Reading and Mathematics subjects and applies contextual questionnaires to students, teachers, and principals.

Until 2003, the NES applied its instruments only in a sample of Brazilian students in the target grades (final grades of the basic education cycles) and had limited coverage - its lowest level of geographic representativeness was the states. In 2005, the evaluation became universal to students from public schools with at least 20 pupils per class, making to yield results by school possible (INEP, 2019). Notwithstanding, the 2005 census evaluation is not publicly available, and we only have data per school from 2007 on.

In figure 2.1 we can visualize the treatment and comparison cohorts of Ceara over time and the periods with data available. For example, the control cohort U1 never received treatment, because its students were in sixth grade when LPRA started and LPRA+5 had not begun yet. The treated cohort U1 received only one year of treatment, in the second grade, and we can observe their proficiency in 2011 for those who managed to achieve the fifth grade and in 2015 for those in ninth grade. The cohorts with both more years of treatment and information available are the treated U3 and U4 as LPRA and LPRA+5 thoroughly treated them.

The cohorts in bordering states did not receive any treatment that we are aware of until 2013 when National Pact of Literacy at the Right Age begun. Which means that the parallel trend assumption may not be valid when we compare cohorts U3 and U4 with the control groups, because control fifth-grade students evaluated in 2015 (2017) probably received National Pact treatment in 2013 (2013 to 2015).

As we have data before and after the policy implementation for the Ceara and all other Brazilian states and following the literature, we decided to implement a natural experiment method. We choose to use only bordering states to Ceara (Piaui, Pernambuco, Paraiba and Rio Grande do Norte) as a comparison group since they belong to the northeast region, which is geographically closer and very similar in economic and social terms to Ceara. We also considered using all other northeastern states, but the placebo DID discussed in section 2.6 let us conclude that the parallel paths assumption was not valid for this control group for both grades.

Our main exercises employ data for the two (four) pre-intervention years and four (two) post-treatment periods with universal data for the fifth (ninth) grade. We construct

panels of schools that participated in all six and two editions of NES, one before and other after the LPRA. We also use the data before 2007 to assess the identification hypothesis of the method chosen to estimate the program effect.

Over the years, the definition of NES universe changed for technical and cost reasons, and to maintain time comparability we restricted our sample to urban schools. Also, we only use data from students that reported never failed in past grades, always attended public schools, and that attend daytime classes. State schools were excluded because practically all Ceara's elementary enrollments are in local schools; for middle students, this percentage was 77% in 2007, and nowadays it exceeds 90%. Finally, Sobral students were not considered, since this city implemented a similar program before 2007.

We created trimmed samples of the comparison group to improve overlap and, therefore, the robustness of estimates, using a procedure proposed by Imbens (2015). In a given pre-treatment year and panel type[15], first, we assess overlap in covariate distributions based on normalized differences, that are differences in average covariate values by treatment status, scaled by a measure of the standard deviation of the covariates. They provide a scale and sample size free way of assessing overlap (IMBENS, 2015). Next, we match treated with control observations on the propensity score without replacement. The propensity score was estimated using the algorithm[16] described in Imbens (2015). We tested only for quadratic terms and include in the linear part all covariates listed in table 2.1 or 2.2.

To perform the matching and generate the trimmed samples, we selected student socioeconomic variables that affect test scores and are available in all periods studied, like gender and socioeconomic status[17], and then average them by schools. A crucial control is the proportion of children that attended preschool, because, as discussed in section 2.2, state and Federal governments intended to increase its supply, and LPRA intended to raise its quality. We believe that preschool attendance would increase the knowledge base of children and boost the learning in subsequent educational levels.

[15] We use two types of panels: schools that participated in NES for two or six years.
[16] This algorithm is implemented in Stata on the psestimate command by Alvaro Carril.
[17] We applied Brazil Criteria to estimate the purchasing power of student families. This criterion is based on ownership of comfort items and head of household education. For details see ABEP (2011).

Figure 2.1 – Control and treatment cohorts of Ceara

Cohort		2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	TT*
Control	S1	9th															0
	S2	7th	8th	9th													0
	S3	5th	6th	7th	8th	9th											0
	S4	3th	4th	5th	6th	7th	8th	9th									0
	U1	1st	2nd	3th	4th	5th	6th	7th	8th	9th							0
	U2			1st	2nd	3th	4th	5th	6th	7th	8th	9th					0
Treated	U1					1st	2nd	3th	4th	5th	6th	7th	8th	9th			1
	U2							1st	2nd	3th	4th	5th	6th	7th	8th	9th	4
	U3									1st	2nd	3th	4th	5th	6th	7th	5
	U4											1st	2nd	3th	4th	5th	5

NES data available
Non treated
Treated by LPRA
Treated by LPRA+5
Treated by More LPRA

Notes: *Indicates total years of treatment until fifth grade. The control cohorts identified by S1 to S4 have only sample data available, except for 9th grade in 2007 and 2009. Cohorts identified by U (1 and 2 for control and 1 to 4 for treatment) have universal data.

Table 2.1 – Summary statistics of school sample - 5th-grade - 2007 - Panel with 6 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.54	0.16	0.55	0.19	1.33	-0.07	0.54	0.16	0.38	-0.02
Black	0.68	0.16	0.62	0.20	-5.62	0.30	0.66	0.17	-1.81	0.11
Distortion	0.15	0.14	0.14	0.17	-0.25	0.01	0.15	0.15	0.05	-0.00
Work	0.15	0.13	0.15	0.13	-0.90	0.05	0.15	0.12	-0.28	0.02
K-Preschool	0.71	0.17	0.72	0.19	2.01	-0.11	0.70	0.17	-0.35	0.02
EC A or B	0.06	0.07	0.08	0.10	3.29	-0.18	0.07	0.07	1.20	-0.07
EC C1	0.14	0.10	0.17	0.14	4.72	-0.26	0.15	0.10	1.61	-0.09
EC C2	0.24	0.12	0.29	0.16	6.48	-0.35	0.26	0.12	2.64	-0.16
EC D	0.41	0.14	0.38	0.19	-3.87	0.21	0.41	0.15	-0.29	0.02
EC E	0.15	0.14	0.09	0.13	-8.84	0.46	0.12	0.13	-3.92	0.23
Observations	579		1002				579			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

Tables 2.1 and 2.2 reports descriptive statistics of our full and trimmed samples of schools in 2007, the first pre-treatment period with universal data, by panel type. We can observe that the characteristics of Ceara schools are very similar by panel type and that the normalized differences between control and treatment groups are substantially smaller in the trimmed sample. For example, in the full sample of table 2.2, the normalized difference in the proportion of students from economic class E was 0.43, and in the trimmed sample it is 0.08. Something similar occur for the other pre-treatment periods and ninth-grade (see tables 2.5 to 2.14 in the appendix).

Table 2.2 – Summary statistics of school sample - 5th-grade - 2007 - Panel with 2 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.54	0.18	0.55	0.20	2.21	-0.07	0.54	0.19	0.44	-0.02
Black	0.66	0.18	0.62	0.21	-7.28	0.24	0.66	0.19	-0.84	0.03
Distortion	0.14	0.15	0.15	0.18	1.73	-0.06	0.15	0.16	1.84	-0.07
Work	0.15	0.14	0.15	0.14	-0.95	0.03	0.16	0.14	0.63	-0.02
K-Preschool	0.72	0.18	0.73	0.20	2.10	-0.07	0.72	0.18	-0.37	0.01
EC A or B	0.06	0.08	0.08	0.10	5.14	-0.17	0.07	0.08	1.01	-0.04
EC C1	0.13	0.11	0.17	0.15	8.09	-0.27	0.13	0.11	0.56	-0.02
EC C2	0.24	0.14	0.28	0.17	8.78	-0.29	0.24	0.14	1.16	-0.04
EC D	0.42	0.16	0.38	0.20	-5.78	0.19	0.42	0.17	0.00	0.00
EC E	0.15	0.15	0.09	0.13	-14.13	0.43	0.14	0.13	-2.14	0.08
Observations	1455		3294				1447			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

2.4 Empirical strategy

We take advantage of the variation across time and space of the program's implementation in 2008 and use the natural experiment method. We estimate the impact of the LPRA on student learning using difference-in-difference alone (DID) and combined with matching methods (MDID). We begin applying the DID estimator to a balanced panel of schools between 2007 to 2017 to allow the LPRA's impact to be heterogeneous according to cohort's time of exposure:

$$y_{ist}^{g} = \lambda^{g} + \theta_{CE,t}^{g} + \gamma_{t}^{g} + \delta_{s}^{g} + \boldsymbol{x}_{i}^{'}\boldsymbol{\alpha} + \varepsilon_{ist}^{g} \tag{2.1}$$

y_{ist}^{g} is the mean test score of school i at grade g in state s at period t, γ_{t}^{g} is a time fixed effect, δ_{s}^{g} is a state fixed effect, ε_{ist}^{g} is a idiossincratic error and λ^{g} is a constant term. The parameter $\theta_{CE,t}^{g}$ give us the policy effects at grade $g = 5$ or 9 by total years of treatment received until fifth-grade. As depicted in figure 2.1, 2007 and 2009 are pre-treatment years and the control cohorts U1 and U2 were not exposed to the treatment, in 2011 the cohort U1 was exposed to one year of treatment, in 2013 the cohort U2 participated during 4 years and so on.

To improve precision we control for the school covariates presented in table 2.2, denoted by $\boldsymbol{x}_i$. The variance-covariance matrix is clustered by cities to correct problems that appear when data from different levels are combined (school and states in our case), such as serial correlation. We identify the average treatment effect on the treated (ATT), and we assume, as in Blundell e Dias (2009), that the idiossincratic error is a transitory chock that is exogenous conditional on the treatment status.

Next, we employ the equation 2.1 on the samples of schools that participated of the NES in two years $t = t_0, t_1$, where $t_0, t_1 \in \{2007, 2009, ..., 2017\}$ and $t_0 < t_1$. When t_1 is a pre-treatment year, we are performing a placebo test, and when $t_0 < 2008$ and $t_1 > 2008$, we are estimating the program's effect. Comparing two years at a time, we maximize the sample size, especially in the trimmed sample.

Besides the classic DID, we combine this approach with matching methods to further assure that the comparison group is similar to the treatment group in terms of its characteristics and, therefore, would respond to shocks in similar ways (BLUNDELL et al., 2004). We pair treated schools with others from bordering states based on the attributes of table 2.2.

In order to identify the ATT with MDID, we maintain the DID assumption about the error and add more two hypothesis: $y_{it_1}^{0} - y_{it_0}^{0}$ is independent of d_i conditional on $\boldsymbol{x}_i$ and $Pr[d_i = 1|\boldsymbol{x}_i] < 1$ (BLUNDELL; DIAS, 2009). Define d as a dummy variable that assumes value one if the school is treated and zero otherwise. Then, we apply matching estimators to the differences between groups before and after the program, as proposed

by Heckman, Ichimura e Todd (1997):

$$\hat{\alpha}^{ATT} = \sum_{d_i=1} \omega(i) \left[q_{1i} - \sum_{d_j=0} W(i,j) q_{0j} \right] \tag{2.2}$$

We use the nearest-neighbour matching estimator with bias-adjustment recommended by Imbens (2015). This estimator sets the weight $\omega(i) = 1/N_T$, $N_T = \sum d$, with each i in treatment group ($d_i = 1$), and picks the match $C(\boldsymbol{x}_i) = \min \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$, j in the comparison group ($d_j = 0$), $\|\cdot\|$ is the mahalanobis norm. $q_{1i} = \hat{y}^{1,adj}_{it_1} - \hat{y}^{0,adj}_{it_0}$, $q_{0j} = \hat{y}^{0,adj}_{jt_1} - \hat{y}^{0,adj}_{jt_0}$ and $\hat{y}^{d,adj}_{it}$ is defined like in Imbens (2015). The number of matches per observation is four and therefore $W(i,j)$ equals a quarter if j was matched to i and zero otherwise.

Also, we use the kernel propensity score[18] matching estimator that sets $\omega(i) = 1/N_T$, $W(i,j) = \frac{\kappa(p_i - p_j)}{\sum_{d_j=0} \kappa(p_i - p_j)}$ to each control j, and p_i is the estimated propensity score. The kernel function $\kappa(\cdot)$ chosen is the epanechnikov, $q_{1i} = y^1_{it_1} - y^0_{it_0}$ and $q_{0j} = y^0_{jt_1} - y^0_{jt_0}$.

We apply models 2.1 and 2.2 for each grade g and subject analyzed. The models by grade intend to capture the program's short-run (fifth) and long-run impacts (ninth). The models by discipline intend to capture LPRA's direct and spillovers effect. For the cohort U1, an impact in Reading is direct, once the main original goal is to teach literacy and therefore must improve notably this discipline, and an effect on Mathematics is a spillover, because this subject was included in the policy only in 2011 for grades three and on. The effect on Mathematics for the cohorts U2 to U4 can be direct or indirect.

To investigate possible heterogeneous effects of the LPRA, we estimate the following DID model:

$$y^g_{ist} = \lambda^g + \theta^g_{CE,t} + (d_i \times \boldsymbol{z}^g_{ist_0})' \boldsymbol{\beta}^g + \gamma^g_t + \delta^g_s + \varepsilon^g_{ist} \tag{2.3}$$

This equation is similar to 2.1 except for the new interaction term of the dummy for Ceara in year $t = t_1$ with some school attribute in the period before treatment. $\boldsymbol{z}^g_{ist_0}$ are binary variables that indicates quintiles of initial average proficiency, proportion of students that declared to be black, proportion of students from economic classes D or E and only E, and proportion of students that attended K-preschool by school. Our main objective is to verify if the most disadvantage students benefit equally or more from the policy.

[18] The propensity score was estimated with the algorithm proposed by Imbens (2015).

2.5 Results

In this section, we present and discuss the estimates of the LPRA's effects on student tests scores, including placebo tests with universal NES data[19], by grade.

2.5.1 Fifth-grade results

We begin presenting the results for schools that participated in the national evaluation from 2007 to 2017 (see figure 2.2). In this sample, that is five times smaller than the 2-years ones, the program did not contribute to raising student's achievement in 2011 but contributed equally from 2013 and on for both subjects. Considering that the first treated cohort received the intervention only at the upper elementary reading stage, that should build on basic skills from the beginning reading stage, this result makes sense. These students probably did not fully develop the necessary skills pointed by Slavin et al. (2009).

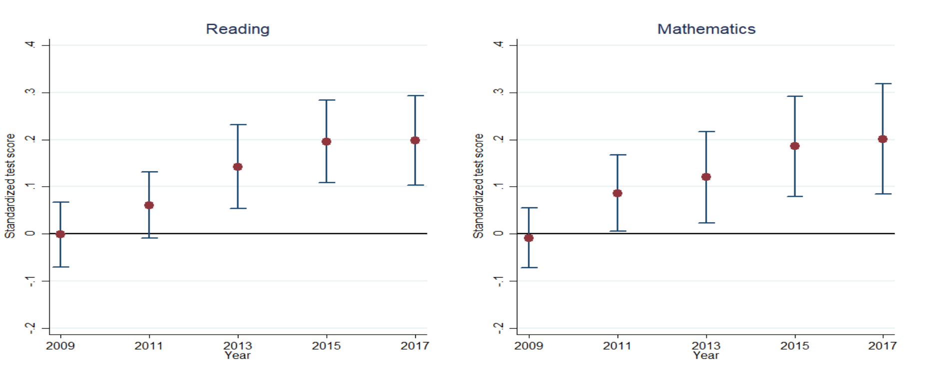
Once we increase the sample size, merging schools that took part in NES at least twice, we observe, in figures 2.3 and 2.4, a positive effect of the policy already in 2011. We summarize all the results in table 2.3 that shows mean effects in units of standard deviation and points in the NES pedagogic scale. In Reading, the average impact is about 0.12 of standard deviation, which can be considered a small effect size compared to the mean effect size of 0.21 reported by Slavin et al. (2009) for programs that provide teachers with extensive professional development to implement specific instructional methods in the U.S. and some European countries. Again, this makes sense considering that students may not fully develop the basic skills in the beginning reading stage.

The Reading average effects for 2013, 2015 and 2017 are consistent in size with those of similar programs in developed countries, with the increased years of treatment for these cohorts and with the higher chance of experiencing better early childhood education, as established in the LPRA's design. Despite the higher point estimates, these overall impacts do not differ from 2011, excepting some estimates for 2017. We are going to see in section 2.5.3 that the program only started properly improving the achievement of students with low test score from 2015 when LPRA and LPRA+5 fully treated the pupils (see figures 2.1 and 2.8). Additionally, as discussed earlier, the time trend of the control group might have changed as a consequence of the National Pact and might have become steeper from 2013 on. We believe that the effects for 2015 and 2017 would be higher in the absence of the federal intervention on the neighboring states.

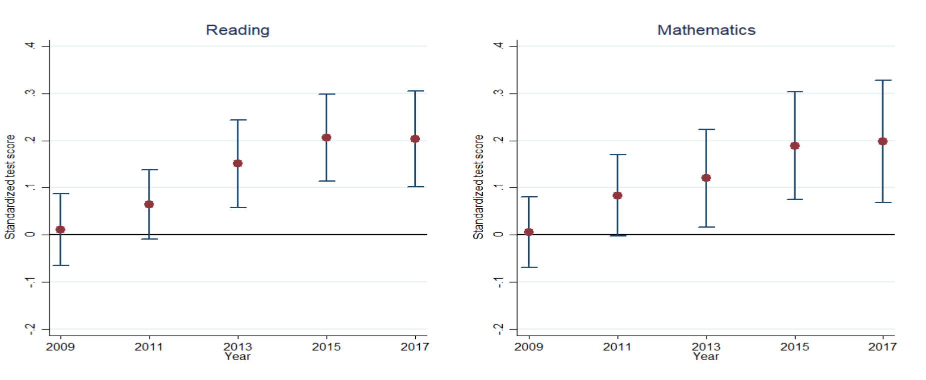
The existence of spillover effects is corroborated by the program positive impacts on Mathematics in 2011, whose mean is around 0.15 standard deviation. In 2013, the mean impact is 0.17, 0.24 in 2015 and 0.27 in 2017. McEwan (2015) reviewed randomized

19 Placebo tests for data before 2007 are reported in section 2.6.

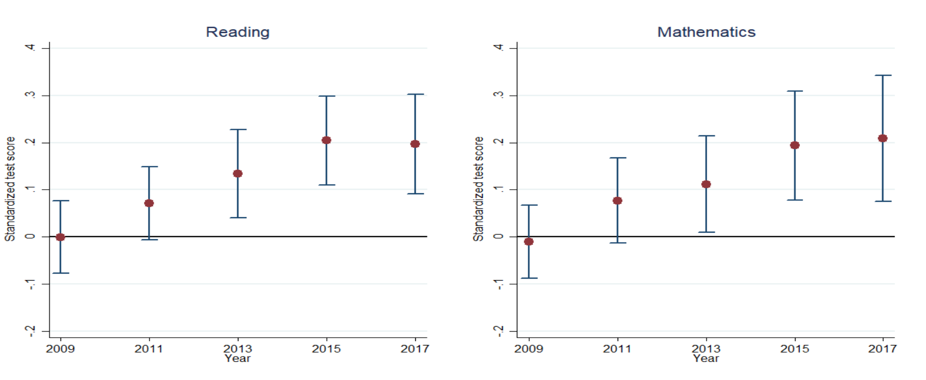
Figure 2.2 – DID estimates of LPRA impact on 5th-grade test scores - school panel from 2007 to 2017



(a) Full sample



(b) Trimmed in 2007



(c) Trimmed in 2009

Notes: All estimates reported control for student covariates.

experiments conducted in developing countries that evaluated the effects of school-based interventions on learning in primary schools and computed effect sizes of several kinds of treatments. The mean effect size for treatments that provided teacher training, combined or not with other inputs such as textbooks, is 0.12. If we use this effect size as a parameter and take into account that, for the first treated cohort, it is a spillover, LPRA's impact on Mathematics achievement can be considered large.

Adopting the standards defined by TPE (2007) for adequate proficiency for fifth-grade - 200 points for Reading and 225 points for Mathematics -, we can say that the mean student achievement, $\bar{n}_{s,5}$, must increase at least $\tilde{n}_{s,5} - \bar{n}_{s,5}$ in fifth-grade, where $\tilde{n}_{s,5}$ denotes the standard for subject s. In table 2.3 we can see these gaps, as well as, the mean achievement in points in the NES pedagogic scale. For example, the gaps between 200 and the mean test score of Ceara students in 2007 were about 18 points and in 2009 13 points for Reading. Taking this into account, the average effect of 6 points in 2011 and 10 in 2013 can be considered small or insufficient. Besides only in 2015 the mean test scores overcame 200 points. The Mathematics mean effects were also insufficient by this criteria, however, the average test score did not reach 225 points until 2017.

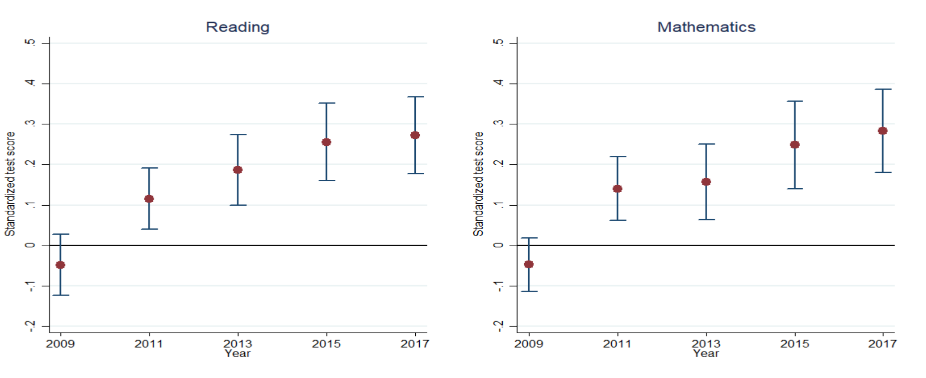
In general, for a given post-treatment year, the results for fifth-graders are robust between samples, methods, control covariates and pre-intervention periods. The common trend assumption is supported by the data - the policy had null effect on the average test score of both disciplines in 2009 (figures 2.4(a), 2.4(b), 2.5(a) and 2.5(b)). When we compare our effect sizes with the literature, we see that LPRA has a good performance in both subjects, but, pedagogically speaking, the effects started too low to bring the mean student to an adequate level of achievement.

Table 2.3 – Average test scores and estimates of the panels with 2 years - 5th-grade - Trimmed sample of Ceara students

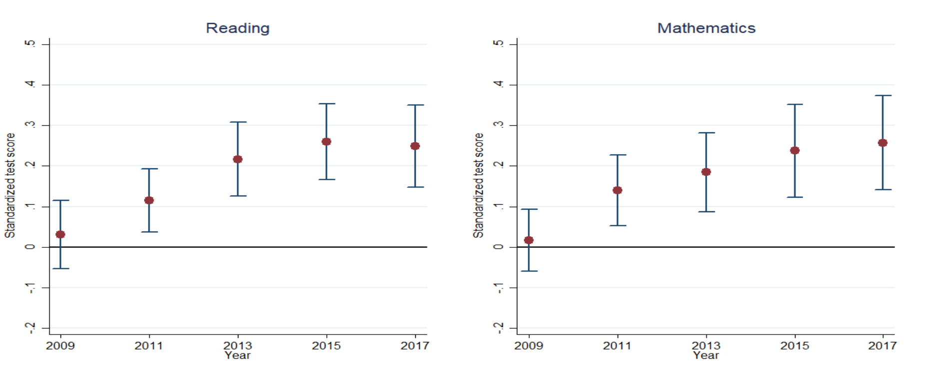
Year	Reading	Mean effects in		Gap*	Mathematics	Mean effects in		Gap*
		SD	Points			SD	Points	
2007	181.58			18.42	194.46			30.54
2009	187.42			12.58	198.55			26.45
2011	182.77	0.12	6	17.23	198.84	0.15	7	26.16
2013	188.55	0.19	10	11.45	200.01	0.17	8	24.99
2015	208.45	0.25	13	0.00	216.66	0.24	12	8.34
2017	217.01	0.25	13	0.00	223.93	0.27	14	1.07

Notes: * denotes the difference between the standard proficiency defined by TPE (2007) and the average test score. The NES standard deviation is 50 points in the pedagogic scale.

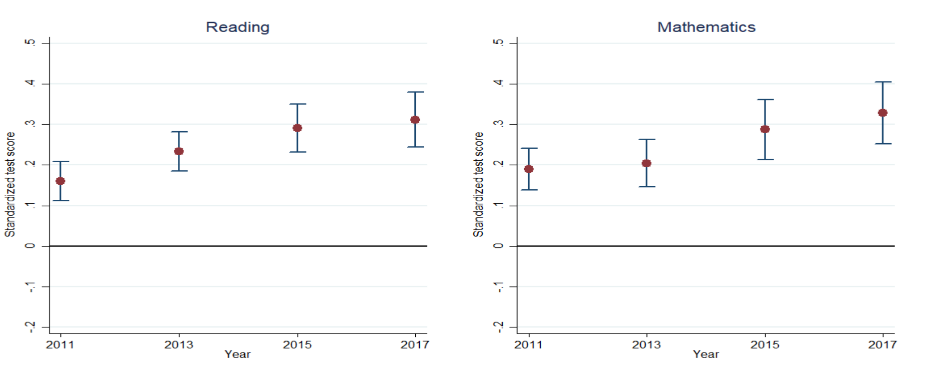
Figure 2.3 – DID estimates of LPRA impact on 5th-grade test scores



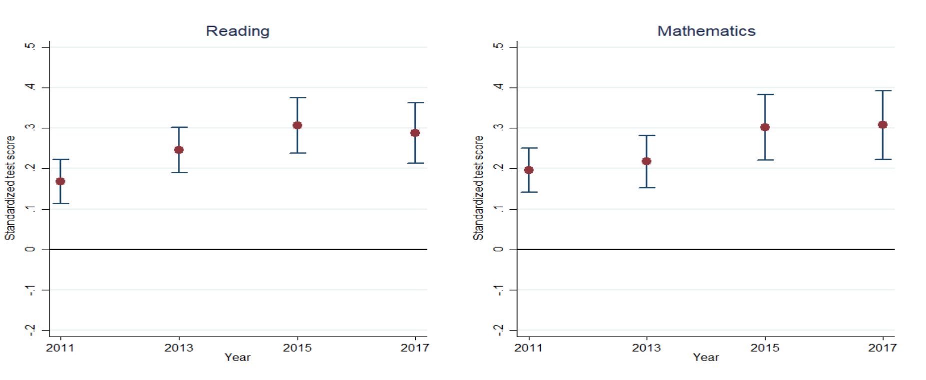
(a) Pre-treatment period: 2007 - Full sample



(b) Pre-treatment period: 2007 - Trimmed sample



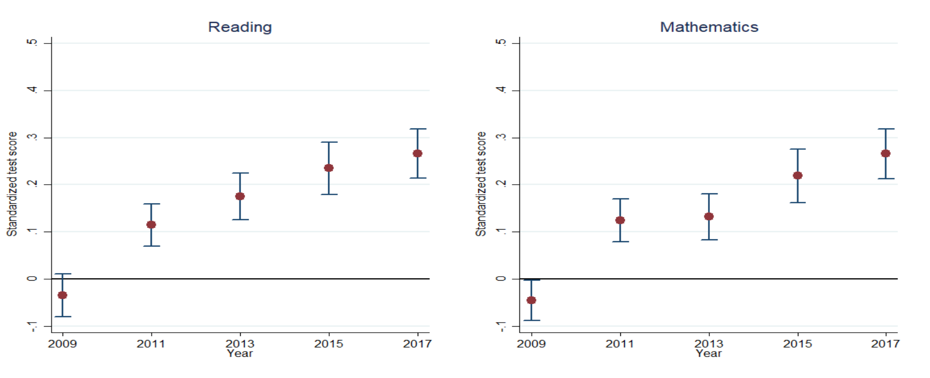
(c) Pre-treatment period: 2009 - Full sample



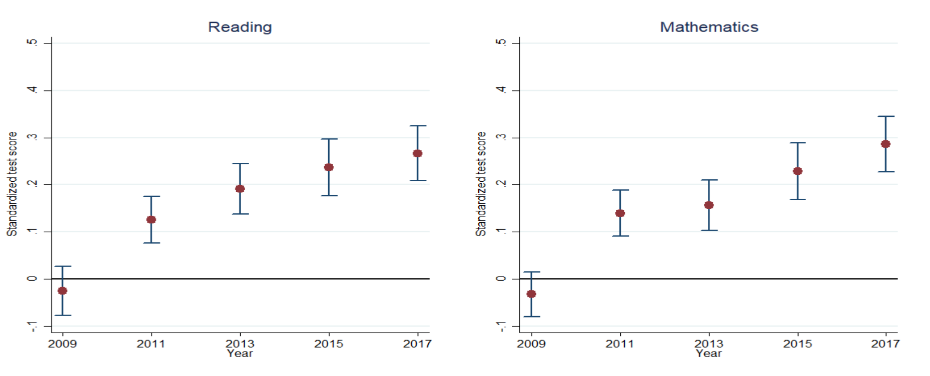
(d) Pre-treatment period: 2009 - Trimmed sample

Notes: All estimates reported control for student covariates.

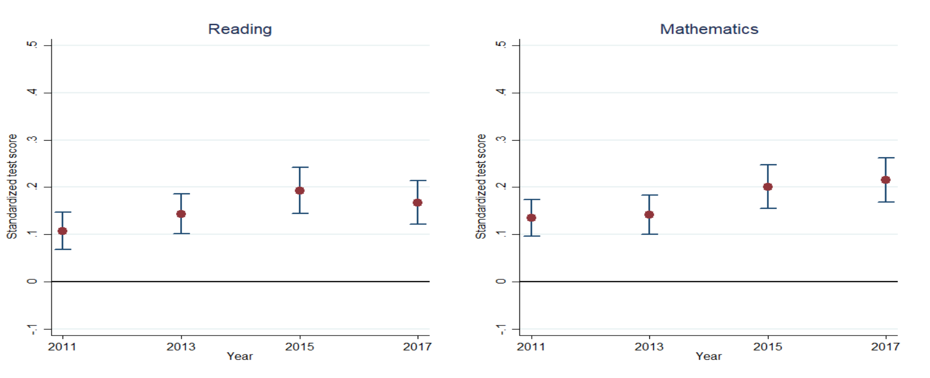
Figure 2.4 – MDID estimates of LPRA impact on 5th-grade test scores



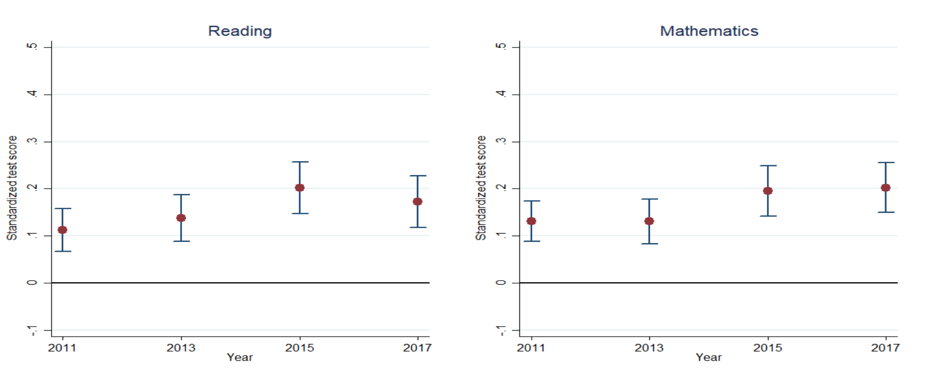
(a) Pre-treatment period: 2007 - Full sample



(b) Pre-treatment period: 2007 - Trimmed sample



(c) Pre-treatment period: 2009 - Full sample



(d) Pre-treatment period: 2009 - Trimmed sample

Notes: All estimates reported control for student covariates and were calculated with the nearest-neighbor matching estimator.

2.5.2 Ninth-grade results

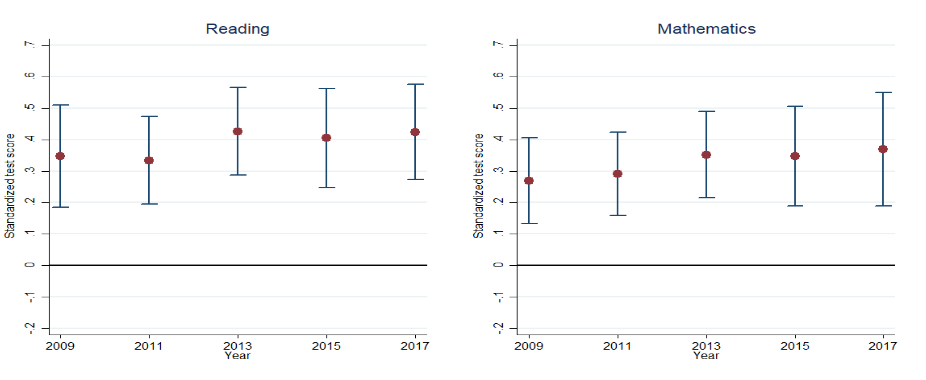
Again, we begin presenting the results for schools that participated in the national evaluation from 2007 to 2017 (see figure 2.5). All estimates for pre-treatment years differ from 2007, suggesting that the common trend assumption does not seem reasonable for 9th-grade for both disciplines once LPRA was launched. The results for the 2-years panels confirm it (see figures 2.6 to 2.7). Ceara's student learning grew faster than in bordering states' before the treated cohorts reached ninth-grade, between 2007 and 2009 and between 2011 and 2013.

Table 2.4 – Average test scores and estimates of the panels with 2 years - 9th-grade - Trimmed sample of Ceara students

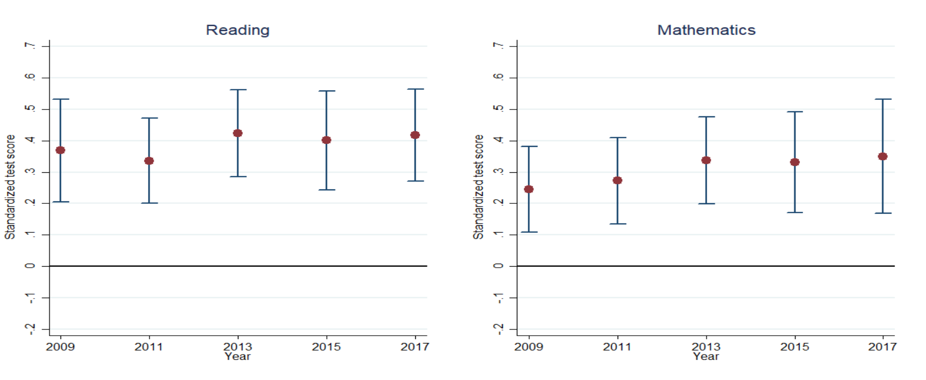
Year	Reading	Mean effects in		Gap*	Mathematics	Mean effects in		Gap*
		SD	Points			SD	Points	
2007	184.01			90.99	196.99			103.01
2009	210.88			64.12	216.05			83.95
2011	231.64			43.36	236.64			63.36
2013	239.01			35.99	240.37			59.63
2015	251.54	0.13	7	23.46	251.61	0.14	7	48.39
2017	260.92	0.14	7	14.08	256.20	0.14	7	43.80

Notes: * denotes the difference between the standard proficiency defined by TPE (2007) and the average test score. The NES standard deviation is 50 points in the pedagogic scale.

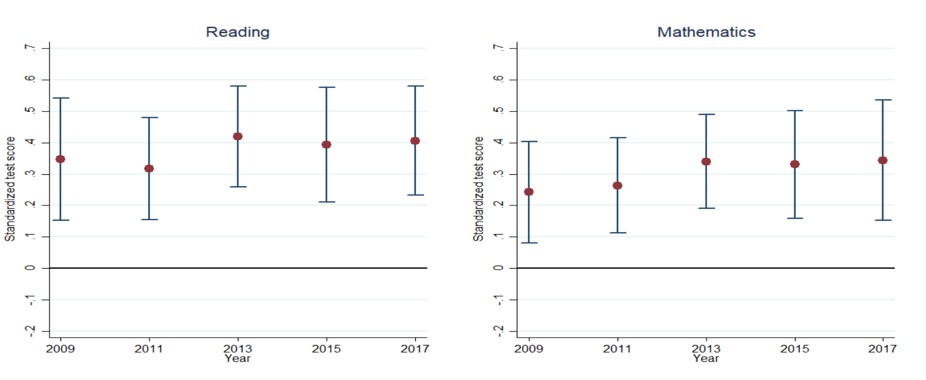
Figure 2.5 – DID estimates of LPRA impact on 9th-grade test scores - school panel from 2007 to 2017



(a) Full sample



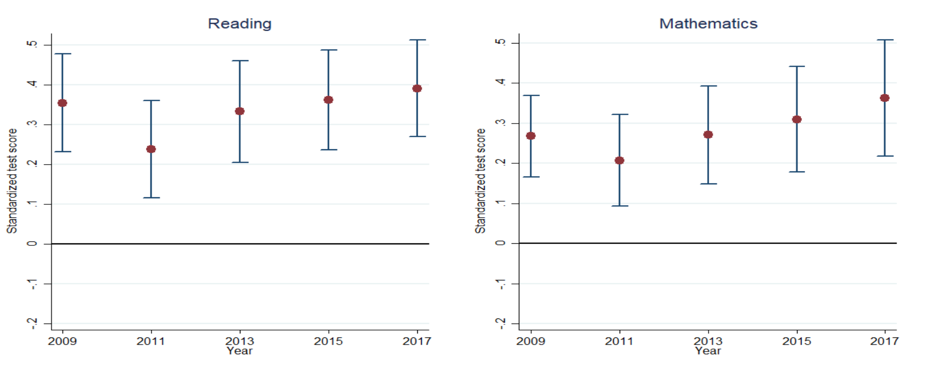
(b) Trimmed in 2007



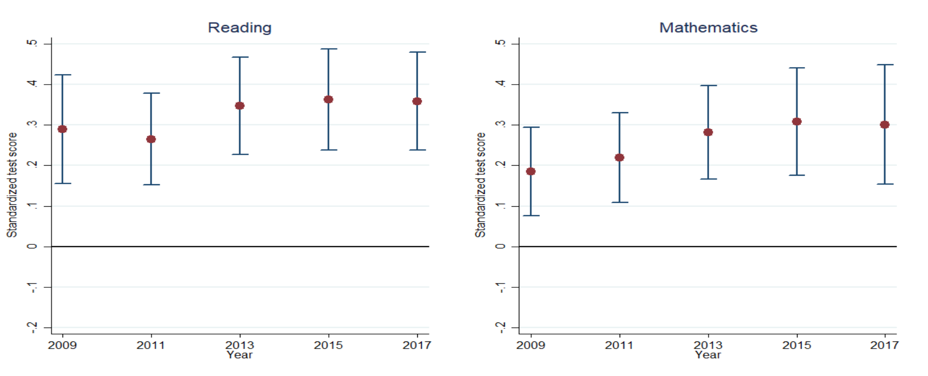
(c) Trimmed in 2009

Notes: All estimates reported control for student covariates.

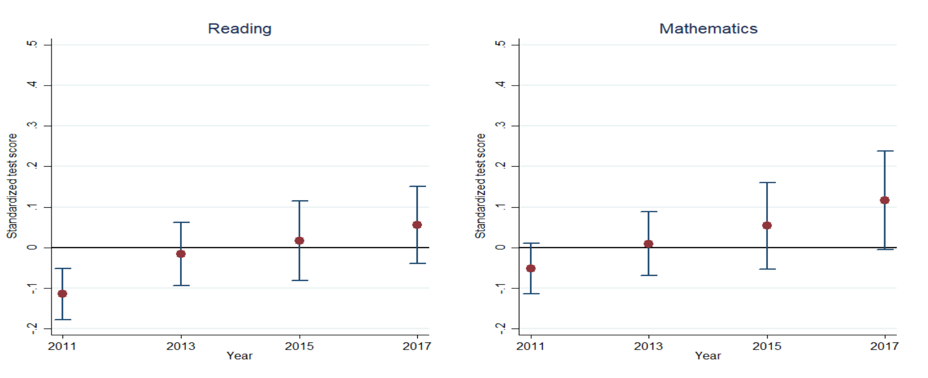
Figure 2.6 – DID estimates of LPRA impact on 9th-grade test scores - pre-treatment year 2007 and 2009



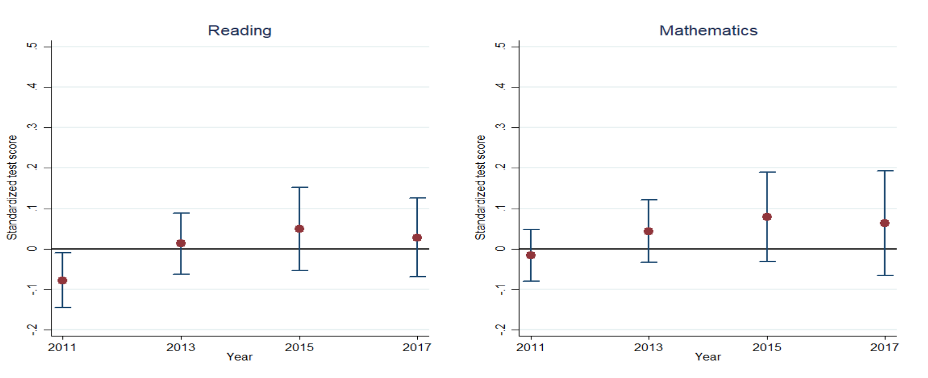
(a) Pre-treatment period: 2007 - Full sample



(b) Pre-treatment period: 2007 - Trimmed sample



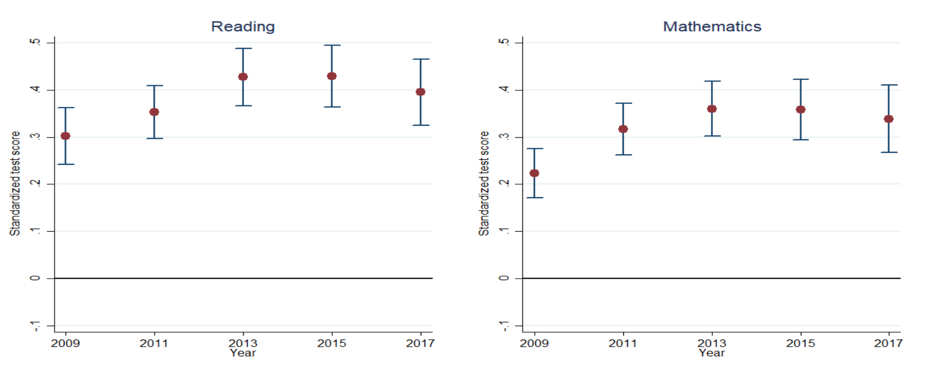
(c) Pre-treatment period: 2009 - Full sample



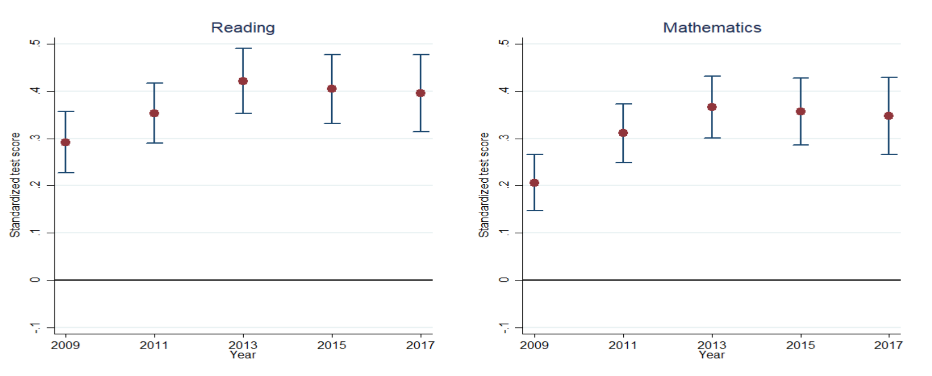
(d) Pre-treatment period: 2009 - Trimmed sample

Notes: All estimates reported control for student covariates.

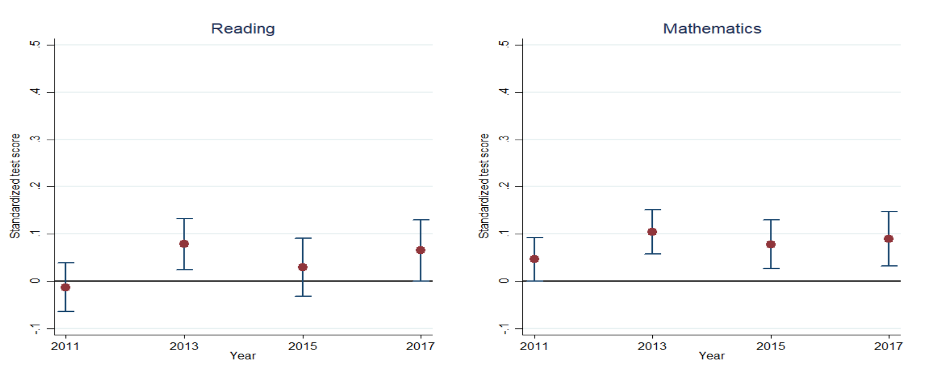
Figure 2.7 – MDID estimates of LPRA impact on 9th-grade test scores - pre-treatment year 2007 and 2009



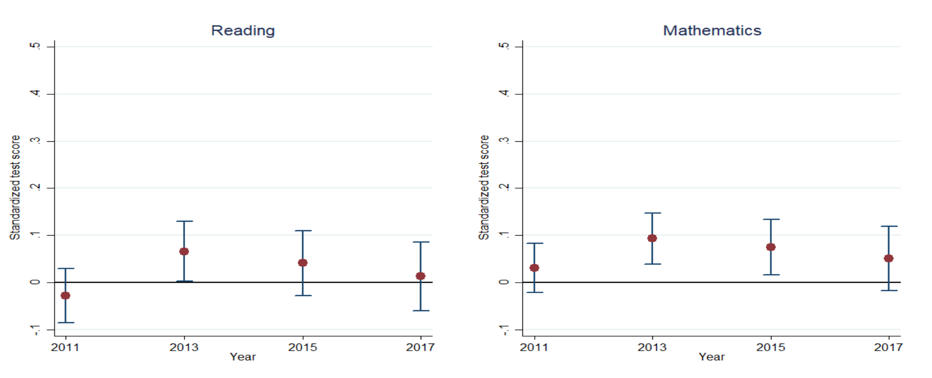
(a) Pre-treatment period: 2007 - Full sample



(b) Pre-treatment period: 2007 - Trimmed sample



(c) Pre-treatment period: 2009 - Full sample



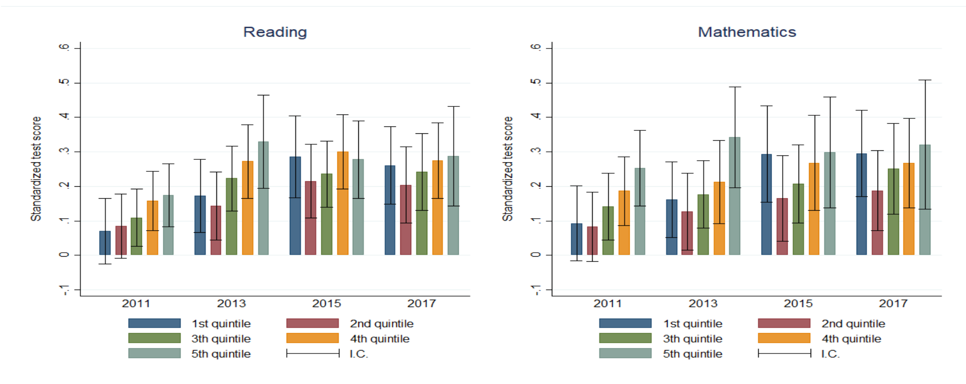
(d) Pre-treatment period: 2009 - Trimmed sample

Notes: All estimates reported control for student covariates and were calculated with the nearest-neighbor matching estimator.

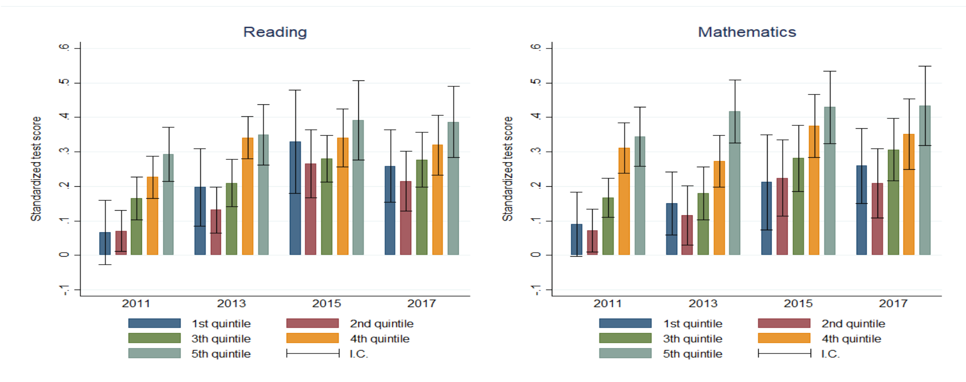
2.5.3 Heterogeneous effects

It is essential to verify if the program helps to improve the learning of all children, independently of their initial proficiency, demographic and socioeconomic status. A policy can increase an average outcome while deepening the gap between different groups simply by not addressing the heterogeneous prior knowledge and skills that children acquired before they begin school. In this section, we check if the LPRA succeed in this tough task. To facilitate the presentation, we created dummy variables that indicate in which quintile the school is by attribute. For example, in figure 2.8 the schools of the first quintile have the smallest average test scores, and those of the fifth quintile have the highest averages. We begin discussing the results for fifth-grade.

Figure 2.8 – DID estimates of LPRA heterogeneous impact on 5th-grade test scores - Initial average test score



(a) Pre-treatment period: 2007



(b) Pre-treatment period: 2009

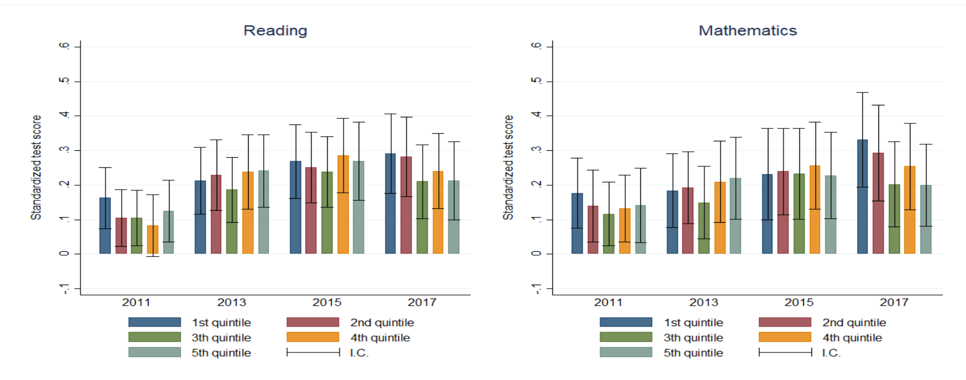
Notes: All estimates reported control for student covariates and use trimmed samples.

First, we check if the policy affected those who needed it the most - the students of low achievement. Our results show that fifth-grade students started fully benefiting from the program only from 2015 (figure 2.8). LPRA did not increase mean test scores of first and second-quintile schools in 2011, and, when the pre-treatment is 2009, the high-achievement pupils benefit from the program much more than the low-performance ones

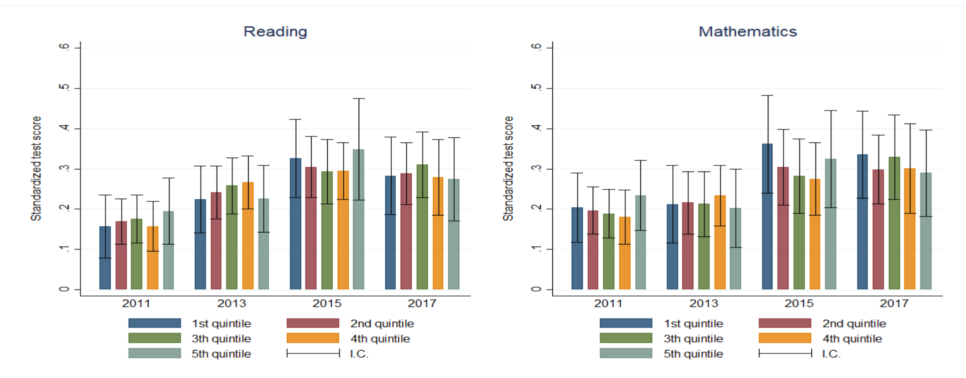
in 2011 and 2013. The impacts stop differing between quintiles in 2015 and 2017.

Here we can understand why the effect estimated for 2011 is relatively smaller. As noted above, the first-cohort students may not fully develop the basic skills in the beginning reading stage as the intervention started in the second grade. The low-performance students, that probably began school with less prior knowledge and pre-reading skills, were those who needed the most special attention at the begging reading stage and even after. Just when the LPRA intervened in the first, third, fourth and fifth grades, the students with low test score were able to advance with the rest. We can interpret these findings as evidence of the Matthews effects - if the treatment time of exposition has not raised, only those students with a certain level of proficiency were able to learn more.

Figure 2.9 – DID estimates of LPRA heterogeneous impact on 5th-grade test scores - Initial proportion of girls



(a) Pre-treatment period: 2007



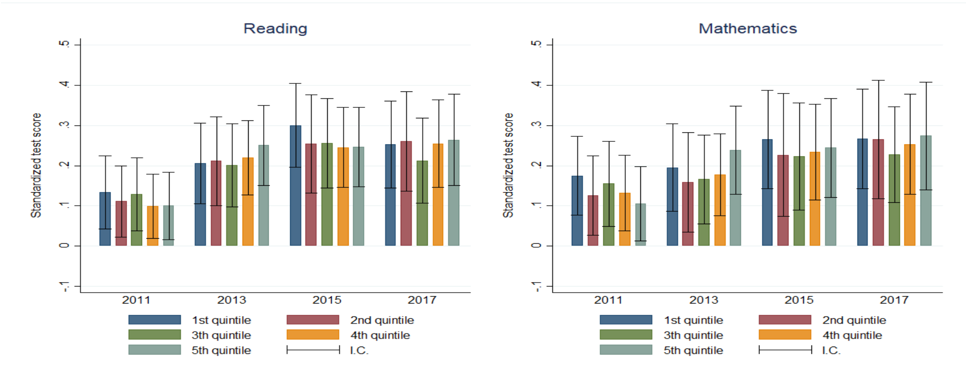
(b) Pre-treatment period: 2009

Notes: All estimates reported control for student covariates and use trimmed samples.

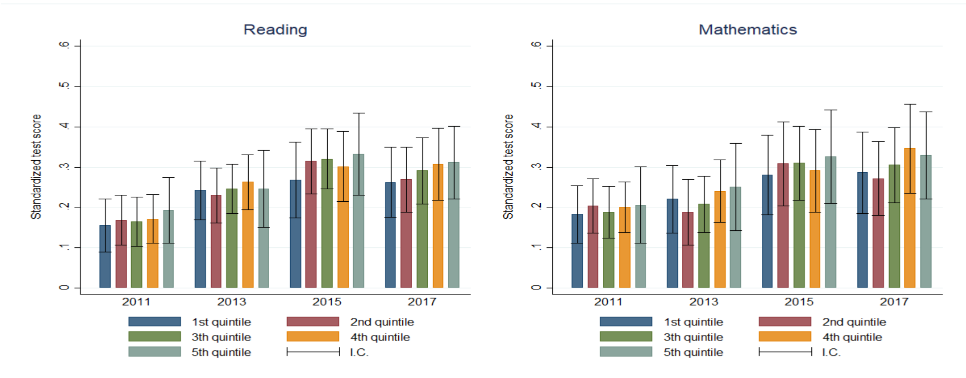
It is widespread in the literature that girls perform better in Reading, boys in Mathematics (see, for example, Marks (2008)), and that black students perform worse than white pupils in general. Given these proficiency gaps between genders and colors, we would expect that the program boosted more Reading achievement for girls, Mathematics performance for boys and achievement for both subjects for white children. Based on

figures 2.9 and 2.10, we conclude that there is no heterogeneous effects by gender and color.

Figure 2.10 – DID estimates of LPRA heterogeneous impact on 5th-grade test scores - Initial proportion of black students



(a) Pre-treatment period: 2007



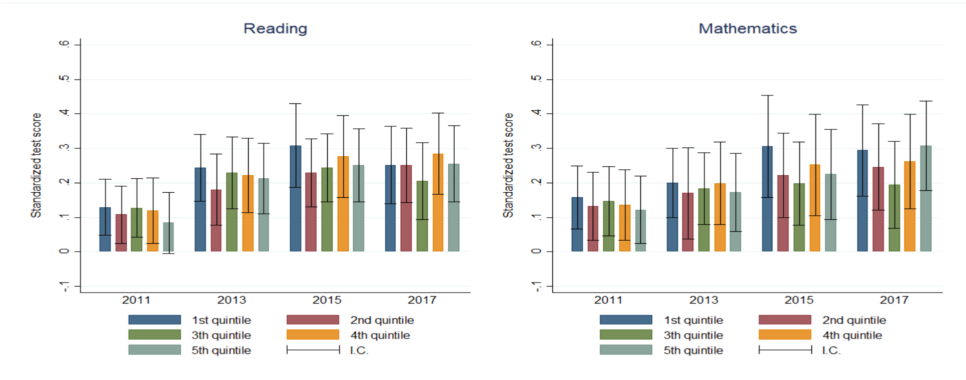
(b) Pre-treatment period: 2009

Notes: All estimates reported control for student covariates and use trimmed samples.

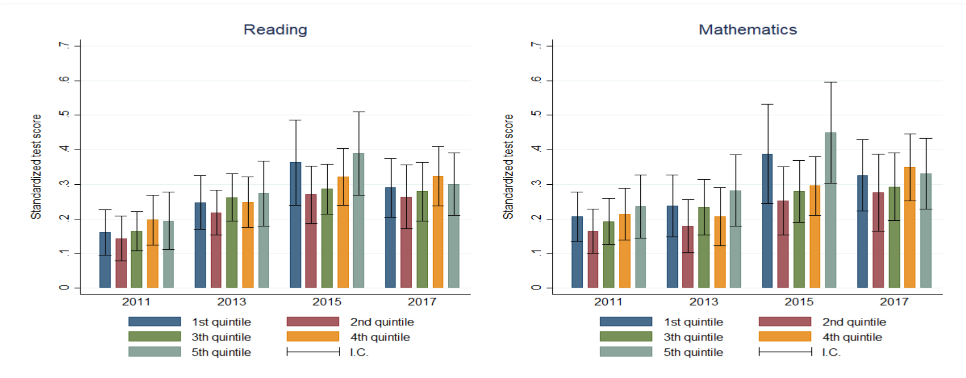
As discussed in section 2.2, the LPRA intended to expand the supply of preschool and improve its quality. If the program succeeded in achieving this goal, more children could enter elementary school with a better knowledge base and then learn at faster rates. Figure 2.11 shows that the program did not affect schools with distinct proportions of students who attended preschool differently in the fifth-grade. This result does not deny the importance of early childhood education, because control states received federal help to expand their preschool supply too. Besides, we do not know if preschool quality actually improved.

Concerning economic status, schools with a high concentration of poorer students (economic classes D and E) benefit equally from the policy since the beginning. Strangely, in figure 2.13(a), the program had a null effect in more heterogeneous schools in terms of economic status, but this result is not robust to the pre-treatment year. The lack of heterogeneous effects by gender, color, preschool attendance, and economic status suggests that

Figure 2.11 – DID estimates of LPRA heterogeneous impact on 5th-grade test scores - Initial proportion of students with K-preschool



(a) Pre-treatment period: 2007

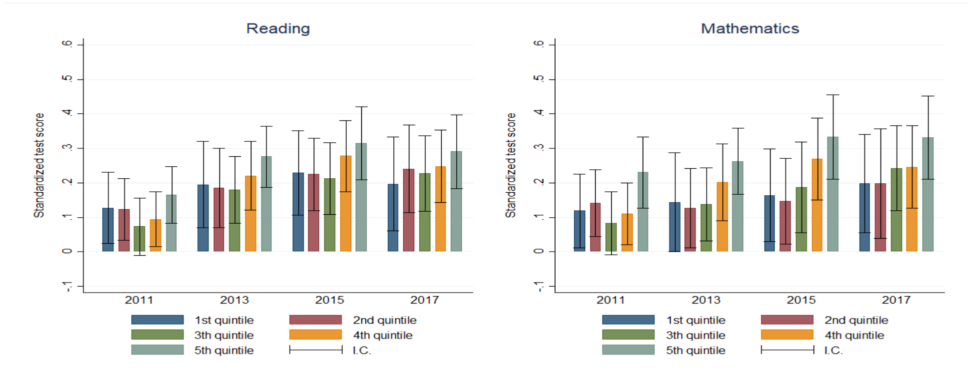


(b) Pre-treatment period: 2009

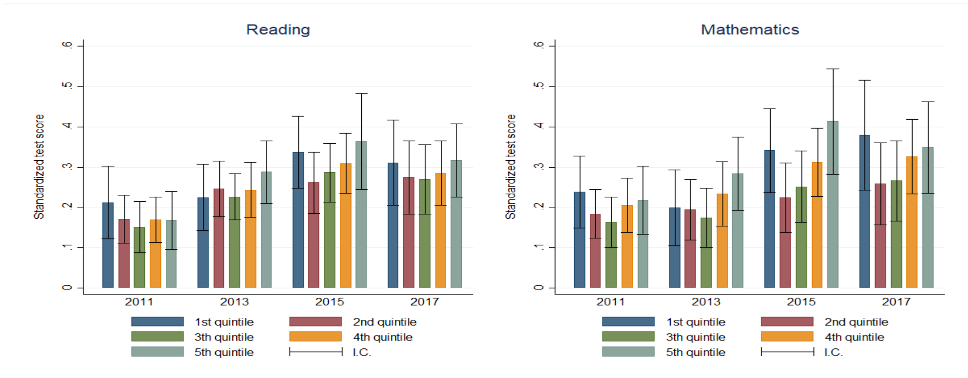
Notes: All estimates reported control for student covariates and use trimmed samples.

low-achievement children are relatively equally distributed between these socioeconomic groups.

Figure 2.12 – DID estimates of LPRA heterogeneous impact on 5th-grade test scores - Initial proportion of students of economic class D and E



(a) Pre-treatment period: 2007

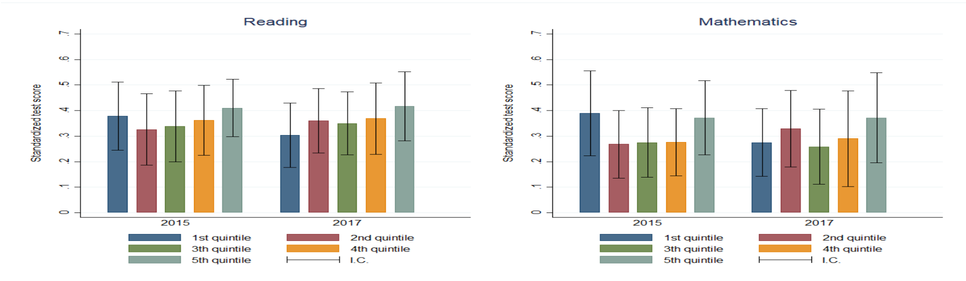


(b) Pre-treatment period: 2009

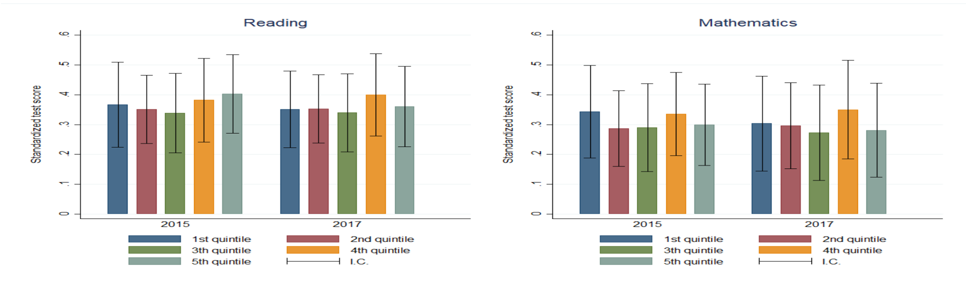
Notes: All estimates reported control for student covariates and use trimmed samples.

The results for ninth-grade are in figure 2.13 and show no difference between schools from distinct quintiles for any characteristic investigated. The average achievement of ninth-grade students started improved before the treated students arrived at the grade.

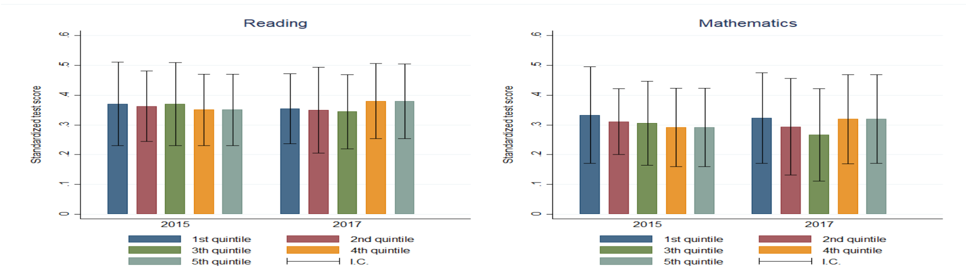
Figure 2.13 – DID estimates of LPRA heterogeneous impact on 9th-grade test scores - Pre-treatment period 2007



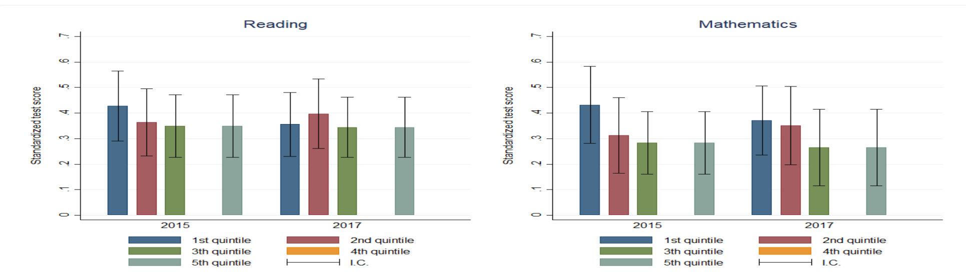
(a) Initial average test score



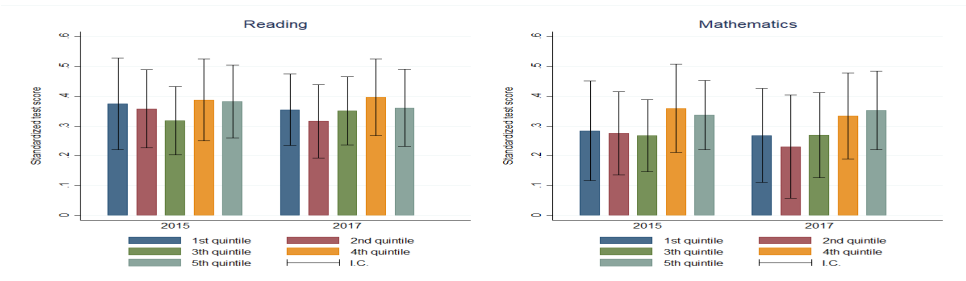
(b) Initial proportion of girls



(c) Initial proportion of black students



(d) Initial proportion of students with K-preschool



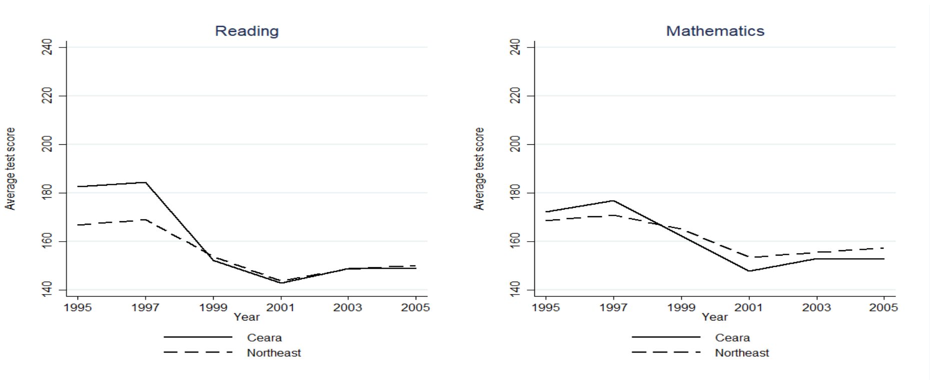
(e) Initial proportion of students of economic class D and E

Notes: All estimates reported control for student covariates and use trimmed samples.

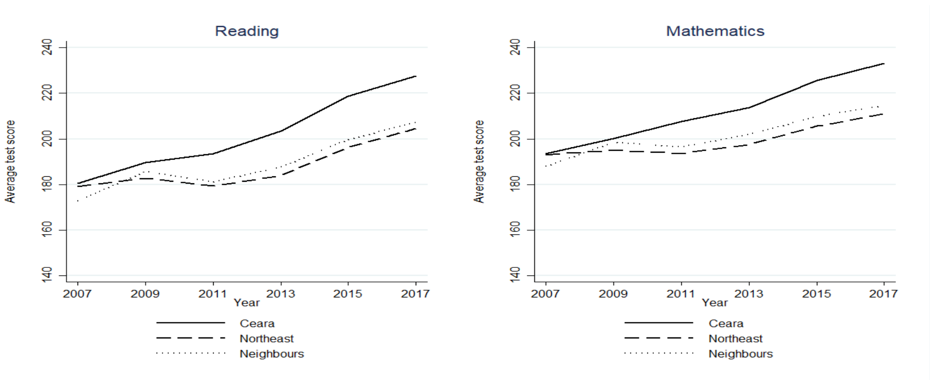
2.6 Specifications and falsification tests

To assess the validity of the parallel paths assumption, we present in figures 2.14 and 2.15 the mean outcomes for pre-treatment periods since 1995, the first year of the historical series. Panel (a) presents all Ceara and Northeast states students until 2005 and panel (b) presents our sample students for Ceara, Northeast and bordering states from 2007 on. For fifth-grade, excepting Reading from 1997 to 1999, the curves are nearly parallel for both subjects until 2009.

Figure 2.14 – Average test scores - 5th grade - municipal urban schools



(a) Sample NES



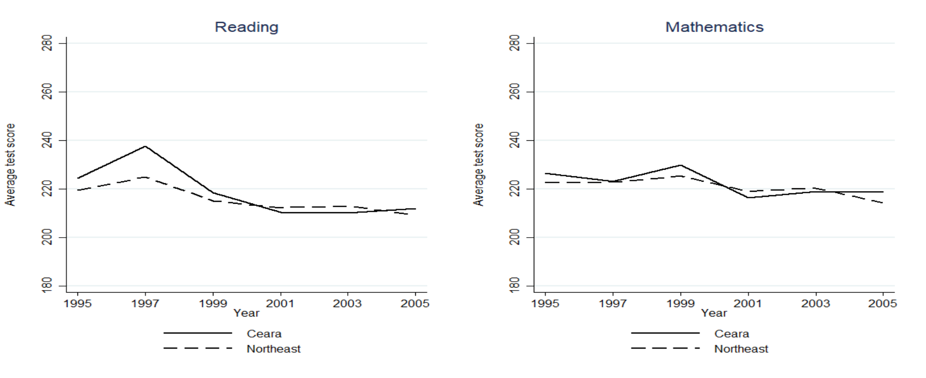
(b) Universal NES - Full sample

Source: Panel (a) - INEP (2007).

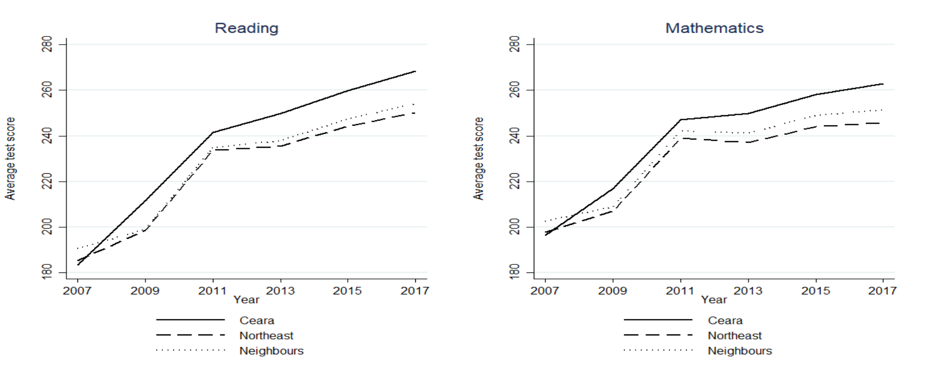
The curves are almost parallel for the ninth-grade until 2005, not considering the two series' first years for Reading. But, as noted in section 2.5, from 2007 to 2009, and from 2011 to 2013, the average test score of Ceara increased faster than bordering states. Ceara only seemed to grow together with the control group from 2009 to 2011 (figure 2.16(a)).

We also estimate placebo DID models similar to equation 2.1 using NES data for 2003 and 2005 at student level (see figure 2.16 and appendix's table 2.27). When the control group contains students from only bordering states, the placebo effect is statistically

Figure 2.15 – Average test scores - 9th grade - municipal urban schools



(a) Sample NES



(b) Universal NES - Full sample

Source: Panel (a) - INEP (2007).

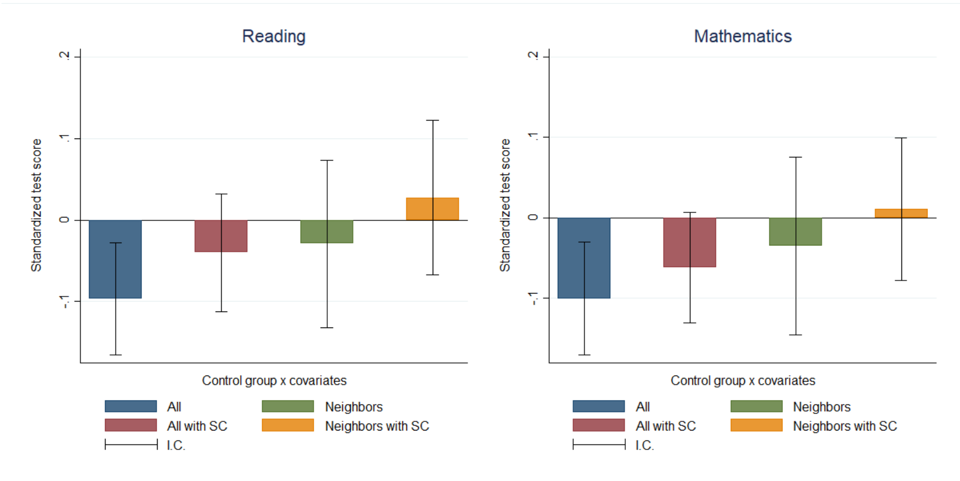
zero for both grades and subjects, but when we consider all northeastern states, there is a negative placebo impact for fifth-graders.

This test reveals that the parallel path assumption is reasonable for both grades before the intervention when the control group contains students from neighboring states. Nevertheless, once LPRA started the 9th-path seems to have changed, and the common trend assumption does not seem reasonable for 9th-grade after 2008.

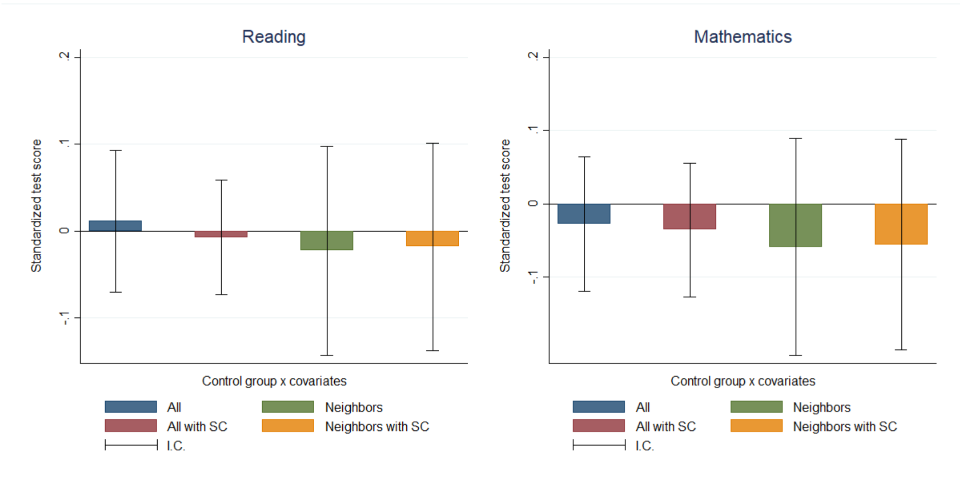
To further reinforce the validity of our main assumption, we should seek similar programs implemented in the control group between 2007 and 2017. We are not aware of any initiative carried by state governments. At the city level, we must search city official newspapers and office websites. However, there are 799 cities in the bordering states and such search is not completely viable on line. Many municipalities do not have their official newspaper online, and, even when they have, the content is limited[20].

[20] We contacted the National Union of Municipal Directors of Education (Undime), but we did not receive an answer so far. We intend to keep contacting them to have access to the literacy programs implemented at city level.

Figure 2.16 – Placebo DID



(a) 5th grade



(b) 9th grade

Notes: *All* refers to the control group formed by all northeastern states, *neighbors* refers to bordering states, SC means that the regression controls for student covariates and I.C. is the confidence interval of the coefficient at 95%.

2.7 Final remarks

Literacy is a critical set of abilities that every human being has the right to acquire. A literate person has the cognitive skills of reading and writing and can use them to address life's challenges of the contemporary world. Ideally, all people who conclude the basic education should be literate, but this is not happening in many countries, regardless of their development level, and there is room to interventions whose goal is to improve literacy, especially at the right age.

Children begin primary school with heterogeneous knowledge bases and pre-reading skills, which are highly associated with demographic and socioeconomic characteristics. The failure to account for this heterogeneity at the right moment can generate increasing reading achievement gaps, a phenomenon called Matthews effects, and prevent people from effectively participating in today's societies and economies.

The Literacy Program at the Right Age (LPRA) is an interesting and complex large-scale reading intervention in a disadvantaged context. It counts with the operational, institutional and political features that are related to policy success and can provide valuable lessons to future Reading programs. The improvement of Ceara's educational outcomes is frequently associated with this policy and, indeed, there is evidence of its effects on the first treated cohort. The availability of data for the subsequent cohorts became possible to verify if the intervention continues to yield positive results, which was the primary goal of this paper. Additionally, we investigated the existence of spillover and heterogeneous effects.

Using the natural experiment method, we found positive and significant program's effects on all investigated cohorts of fifth-graders for both subjects. Comparing the impacts in Reading with the effect sizes reported by the literature, the effect on the first-treated cohort can be considered small (0.12 standard deviation or 6 points in NES pedadogic scale), and this is explained by the fact that low-achievement students were not being benefited from the policy in 2011, possibly because they needed more time of exposure to the treatment to compensate their initial conditions and subsequent learning flaws derived from them. After the LPRA and LPRA+5 were fully implemented, students with all levels of achievement presented better test scores on average. The program did not differently affect students by gender, color, economic status and preschool attendance in all periods analyzed. The impacts in Mathematics in 2011 (0.15 s.d. or 7 points) can be considered large when compared to the literature and corroborates the existence of spillovers.

For the students from the second to fourth cohorts treated, the program's effects in Reading have a similar size to those of comparable interventions in developed countries and is consistent with the increased years of treatment. Nevertheless, pedagogically speaking, the effects were too low to bring the mean student to an adequate level of achievement until

2013. The impacts for Mathematics were large in view of the literature but pedagogically insufficient to make the mean student achieve a sufficient level of learning until 2017.

We could not investigate long-run impacts of the program. Our placebo tests indicated that the parallel trend hypothesis is not reasonable for ninth-grade and the effects for the post-treatment years did not differ from those for the years before.

As future steps, we intend to estimate the LPRA's impact on retention and then use Lee's bounds (LEE, 2009). Probably retention precludes the cohorts of treated students from arriving at the final grades of elementary and middle school. In other words, we do not observe the test score of treated students that fail one or more grades, but we can address them as missing values and use the mentioned bounds.

2.A Appendix A

Table 2.5 – Summary statistics of school sample - 5th-grade - 2009 - Panel with 6 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.51	0.16	0.53	0.20	1.80	-0.10	0.51	0.17	0.10	-0.01
Black	0.66	0.17	0.62	0.22	-3.39	0.18	0.66	0.17	0.13	-0.01
Distortion	0.14	0.13	0.16	0.17	2.98	-0.16	0.14	0.13	-0.13	0.01
Work	0.15	0.13	0.15	0.14	-0.10	0.01	0.15	0.13	0.13	-0.01
K-Preschool	0.71	0.17	0.72	0.18	1.83	-0.10	0.71	0.17	0.22	-0.01
EC A or B	0.08	0.08	0.10	0.12	4.18	-0.23	0.08	0.09	0.89	-0.05
EC C1	0.17	0.12	0.21	0.17	4.82	-0.26	0.17	0.12	1.09	-0.06
EC C2	0.30	0.15	0.31	0.18	1.12	-0.06	0.31	0.15	1.32	-0.08
EC D	0.36	0.16	0.31	0.19	-5.39	0.29	0.35	0.16	-0.85	0.05
EC E	0.10	0.12	0.08	0.13	-3.32	0.17	0.08	0.10	-2.46	0.14
Observations	585		1028				582			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

Table 2.6 – Summary statistics of school sample - 5th-grade - 2009 - Panel with 2 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.52	0.22	0.52	0.25	0.09	-0.00	0.52	0.22	0.50	-0.01
Black	0.66	0.22	0.63	0.25	-5.55	0.14	0.66	0.23	-0.20	0.01
Distortion	0.14	0.17	0.16	0.20	4.66	-0.12	0.14	0.17	-0.25	0.01
Work	0.15	0.16	0.15	0.18	1.01	-0.02	0.14	0.16	-1.33	0.04
K-Preschool	0.71	0.21	0.73	0.23	2.02	-0.05	0.71	0.21	-0.11	0.00
EC A or B	0.08	0.12	0.10	0.14	5.52	-0.14	0.07	0.12	-1.02	0.03
EC C1	0.17	0.16	0.21	0.20	9.29	-0.23	0.16	0.16	-0.62	0.02
EC C2	0.31	0.21	0.31	0.22	0.07	-0.00	0.31	0.20	-0.63	0.02
EC D	0.36	0.21	0.31	0.23	-7.97	0.20	0.37	0.21	1.78	-0.05
EC E	0.09	0.14	0.07	0.14	-4.97	0.12	0.09	0.14	-0.24	0.01
Observations	2522		5335				2472			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

Table 2.7 – Summary statistics of school sample - 9th-grade - 2007 - Panel with 6 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.59	0.21	0.62	0.25	1.85	-0.11	0.60	0.23	0.26	-0.02
Black	0.64	0.21	0.60	0.25	-2.55	0.15	0.63	0.23	-0.45	0.03
Distortion	0.03	0.08	0.06	0.15	3.63	-0.22	0.04	0.11	1.16	-0.07
Work	0.09	0.13	0.09	0.16	0.46	-0.03	0.08	0.14	-0.29	0.02
K-Preschool	0.88	0.17	0.86	0.20	-1.93	0.11	0.87	0.18	-0.99	0.06
EC A or B	0.09	0.14	0.13	0.18	4.17	-0.24	0.10	0.14	1.27	-0.08
EC C1	0.17	0.16	0.22	0.21	4.85	-0.29	0.17	0.15	-0.09	0.01
EC C2	0.30	0.19	0.32	0.23	1.73	-0.10	0.32	0.21	1.55	-0.10
EC D	0.38	0.23	0.28	0.25	-6.96	0.40	0.36	0.23	-1.35	0.09
EC E	0.07	0.10	0.05	0.12	-2.73	0.16	0.06	0.12	-1.52	0.10
Observations	500		800				499			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

Table 2.8 – Summary statistics of school sample - 9th-grade - 2007 - Panel with 2 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.58	0.25	0.61	0.27	3.04	-0.10	0.59	0.26	0.50	-0.02
Black	0.65	0.23	0.60	0.27	-5.55	0.18	0.64	0.25	-1.28	0.05
Distortion	0.04	0.13	0.06	0.15	3.43	-0.11	0.04	0.14	0.25	-0.01
Work	0.10	0.17	0.09	0.17	-2.85	0.09	0.10	0.17	0.07	0.00
K-Preschool	0.86	0.20	0.86	0.22	-1.00	0.03	0.86	0.22	-0.77	0.03
EC A or B	0.09	0.14	0.13	0.19	8.19	-0.28	0.09	0.15	0.33	-0.01
EC C1	0.16	0.17	0.21	0.21	8.09	-0.27	0.16	0.17	-0.59	0.02
EC C2	0.30	0.22	0.32	0.25	2.88	-0.09	0.30	0.23	0.36	-0.01
EC D	0.38	0.25	0.29	0.26	-11.15	0.36	0.39	0.26	0.83	-0.03
EC E	0.08	0.14	0.05	0.13	-6.48	0.21	0.07	0.16	-1.59	0.06
Observations	1410		3073				1406			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

Table 2.9 – Summary statistics of school sample - 9th-grade - 2009 - Panel of 6 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.60	0.19	0.63	0.23	2.49	-0.15	0.60	0.19	-0.06	0.00
Black	0.69	0.18	0.64	0.23	-3.52	0.21	0.67	0.19	-0.92	0.06
Distortion	0.03	0.08	0.06	0.12	4.76	-0.28	0.04	0.08	0.58	-0.04
Work	0.08	0.12	0.08	0.14	-0.26	0.02	0.08	0.14	0.24	-0.02
K-Preschool	0.85	0.15	0.84	0.18	-0.43	0.02	0.84	0.16	-0.61	0.04
EC A or B	0.12	0.14	0.16	0.18	3.70	-0.22	0.12	0.14	0.31	-0.02
EC C1	0.23	0.16	0.25	0.20	1.81	-0.11	0.24	0.17	0.86	-0.05
EC C2	0.33	0.18	0.31	0.21	-2.39	0.14	0.32	0.18	-0.83	0.05
EC D	0.27	0.18	0.24	0.22	-1.97	0.11	0.26	0.19	-0.04	0.00
EC E	0.05	0.10	0.05	0.09	-0.66	0.04	0.05	0.09	-0.29	0.02
Observations	498		811				496			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

Table 2.10 – Summary statistics of school sample - 9th-grade - 2009 - Panel with 2 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.60	0.26	0.63	0.29	4.18	-0.11	0.61	0.26	1.17	-0.04
Black	0.67	0.26	0.62	0.30	-6.94	0.18	0.67	0.27	-0.50	0.02
Distortion	0.04	0.12	0.06	0.16	6.24	-0.17	0.04	0.12	-0.55	0.02
Work	0.09	0.17	0.07	0.16	-4.66	0.12	0.09	0.17	-0.78	0.02
K-Preschool	0.84	0.20	0.85	0.22	1.15	-0.03	0.84	0.21	-0.91	0.03
EC A or B	0.11	0.17	0.16	0.22	8.37	-0.23	0.11	0.17	-1.20	0.04
EC C1	0.22	0.22	0.26	0.27	6.75	-0.18	0.22	0.23	-0.14	0.00
EC C2	0.33	0.25	0.31	0.27	-3.13	0.08	0.33	0.26	0.99	-0.03
EC D	0.29	0.26	0.23	0.26	-8.65	0.22	0.29	0.26	-0.08	0.00
EC E	0.05	0.14	0.04	0.13	-3.09	0.08	0.05	0.14	0.05	0.00
Observations	2142		5155				2123			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

Table 2.11 – Summary statistics of school sample - 9th-grade - 2011 - Panel with 6 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.58	0.11	0.63	0.13	6.23	-0.36	0.60	0.12	2.69	-0.17
Black	0.72	0.12	0.67	0.14	-6.93	0.40	0.70	0.12	-3.14	0.20
Distortion	0.02	0.04	0.02	0.04	-0.13	0.01	0.02	0.04	-1.01	0.06
Work	0.14	0.08	0.12	0.09	-5.15	0.30	0.13	0.08	-2.76	0.17
K-Preschool	0.87	0.09	0.82	0.13	-7.88	0.47	0.86	0.09	-2.68	0.17
EC A or B	0.10	0.08	0.14	0.12	6.84	-0.40	0.12	0.09	3.53	-0.22
EC C1	0.24	0.11	0.26	0.13	2.58	-0.15	0.26	0.12	2.65	-0.17
EC C2	0.37	0.11	0.34	0.12	-4.99	0.29	0.36	0.11	-1.66	0.10
EC D	0.26	0.14	0.23	0.14	-3.29	0.19	0.23	0.14	-2.93	0.19
EC E	0.03	0.04	0.03	0.05	0.14	-0.01	0.03	0.05	-0.50	0.03
Observations	500		811				500			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

Table 2.12 – Summary statistics of school sample - 9th-grade - 2011 - Panel with 2 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.58	0.12	0.63	0.15	8.82	-0.32	0.59	0.13	1.41	-0.06
Black	0.72	0.13	0.66	0.16	-11.05	0.41	0.71	0.13	-1.43	0.06
Distortion	0.02	0.04	0.02	0.05	1.53	-0.06	0.02	0.04	0.46	-0.02
Work	0.14	0.09	0.12	0.10	-7.84	0.29	0.14	0.09	-1.98	0.08
K-Preschool	0.87	0.09	0.83	0.14	-9.92	0.38	0.86	0.10	-2.69	0.11
EC A or B	0.10	0.08	0.13	0.12	10.15	-0.38	0.10	0.09	1.40	-0.06
EC C1	0.23	0.12	0.27	0.14	7.45	-0.27	0.24	0.12	1.32	-0.06
EC C2	0.37	0.11	0.34	0.13	-5.69	0.21	0.37	0.12	0.89	-0.04
EC D	0.27	0.14	0.23	0.15	-8.06	0.29	0.25	0.14	-2.38	0.10
EC E	0.03	0.05	0.03	0.06	-3.30	0.12	0.03	0.06	-0.85	0.04
Observations	1120		2698				1120			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

Table 2.13 – Summary statistics of school sample - 9th-grade - 2013 - Panel with 6 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.58	0.11	0.61	0.12	4.35	-0.25	0.60	0.11	2.04	-0.13
Black	0.72	0.13	0.67	0.14	-6.04	0.35	0.70	0.12	-2.51	0.16
Distortion	0.02	0.04	0.02	0.04	0.04	-0.00	0.02	0.04	-0.33	0.02
Work	0.11	0.08	0.09	0.07	-3.55	0.20	0.10	0.07	-2.57	0.16
K-Preschool	0.89	0.10	0.84	0.12	-6.84	0.40	0.87	0.09	-2.23	0.14
EC A or B	0.11	0.09	0.18	0.12	10.30	-0.60	0.15	0.10	5.67	-0.36
EC C1	0.25	0.13	0.30	0.12	7.58	-0.43	0.29	0.12	5.65	-0.36
EC C2	0.30	0.11	0.31	0.12	1.45	-0.08	0.32	0.11	2.89	-0.18
EC D	0.27	0.14	0.19	0.13	-10.68	0.60	0.21	0.13	-6.73	0.43
EC E	0.07	0.11	0.02	0.06	-9.96	0.52	0.03	0.06	-7.26	0.46
Observations	500		811				491			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

Table 2.14 – Summary statistics of school sample - 9th-grade - 2013 - Panel with 2 years

Covariate	Treated		Non Treated - Full				Non Treated - Trimmed			
	Mean	SD	Mean	SD	t-stat	nor-dif	Mean	SD	t-stat	nor-dif
Girl	0.58	0.13	0.61	0.14	6.65	-0.24	0.59	0.12	2.14	-0.09
Black	0.73	0.14	0.68	0.16	-8.85	0.32	0.72	0.14	-1.77	0.08
Distortion	0.02	0.05	0.02	0.05	0.73	-0.03	0.02	0.04	-0.16	0.01
Work	0.11	0.09	0.09	0.08	-4.66	0.16	0.10	0.08	-1.14	0.05
K-Preschool	0.89	0.10	0.85	0.12	-9.54	0.36	0.89	0.09	-1.13	0.05
EC A or B	0.11	0.09	0.17	0.13	16.06	-0.61	0.13	0.10	5.09	-0.22
EC C1	0.24	0.14	0.30	0.14	13.77	-0.50	0.27	0.12	6.17	-0.27
EC C2	0.30	0.12	0.32	0.14	2.92	-0.11	0.32	0.11	4.46	-0.19
EC D	0.28	0.15	0.19	0.14	-18.63	0.66	0.24	0.14	-5.89	0.25
EC E	0.08	0.12	0.02	0.06	-18.41	0.56	0.04	0.07	-9.37	0.41
Observations	1092		2736				1051			

Notes: SD denotes standard deviation, nor-dif is normalized differences in average, and EC is economic class.

Table 2.15 – DID estimates of LPRA impact on 5th-grade test scores - pre-treatment period 2007

Period	Reading				Mathematics			
	Full sample		Trim. sample		Full sample		Trim. sample	
2009*	-0.046	-0.048	0.025	0.031	-0.046	-0.048	0.013	0.016
	0.039	0.039	0.041	0.043	0.033	0.034	0.037	0.039
2011	0.124	0.116	0.106	0.115	0.151	0.140	0.128	0.139
	0.038	0.039	0.040	0.040	0.040	0.040	0.045	0.044
2013	0.194	0.187	0.201	0.217	0.166	0.157	0.169	0.184
	0.045	0.045	0.047	0.046	0.048	0.048	0.051	0.050
2015	0.264	0.255	0.258	0.259	0.259	0.249	0.237	0.238
	0.049	0.049	0.048	0.048	0.055	0.055	0.059	0.058
2017	0.289	0.273	0.246	0.249	0.306	0.284	0.257	0.257
	0.048	0.048	0.051	0.052	0.053	0.053	0.059	0.059
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: *Placebo test. Clustered standard errors by cities below coefficients. The period is the post treatment year.

Table 2.16 – DID estimates of LPRA impact on 5th-grade test scores - pre-treatment period 2009

Period	Reading				Mathematics			
	Full sample		Trim. sample		Full sample		Trim. sample	
2011	0.170	0.161	0.166	0.169	0.197	0.189	0.192	0.196
	0.024	0.025	0.028	0.028	0.027	0.026	0.029	0.028
2013	0.236	0.233	0.236	0.246	0.209	0.204	0.210	0.217
	0.026	0.024	0.030	0.029	0.032	0.030	0.035	0.033
2015	0.306	0.291	0.307	0.307	0.302	0.288	0.302	0.302
	0.031	0.030	0.036	0.035	0.040	0.038	0.043	0.042
2017	0.333	0.312	0.285	0.288	0.351	0.329	0.305	0.308
	0.033	0.035	0.037	0.038	0.039	0.039	0.044	0.043
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Clustered standard errors by cities below coefficients. The period is the post treatment year.

Table 2.17 – MDID estimates of LPRA impact on 5th-graders - pre-treatment period 2007

	Reading				Mathematics			
Period	NN		Kernel		NN		Kernel	
	Full	Trim.	Full	Trim	Full	Trim.	Full	Trim
2009*	-0.035	-0.021	-0.038	-0.015	-0.048	-0.026	-0.046	-0.022
	0.001	0.001	0.021	0.021	0.001	0.001	0.018	0.020
2011	0.128	0.131	0.110	0.111	0.130	0.148	0.128	0.129
	0.001	0.001	0.023	0.021	0.001	0.001	0.024	0.027
2013	0.185	0.196	0.179	0.197	0.135	0.160	0.149	0.165
	0.001	0.001	0.019	0.024	0.001	0.001	0.020	0.025
2015	0.254	0.239	0.267	0.271	0.232	0.229	0.247	0.248
	0.001	0.001	0.032	0.036	0.001	0.001	0.027	0.027
2017	0.268	0.258	0.276	0.282	0.269	0.280	0.274	0.285
	0.001	0.001	0.026	0.029	0.001	0.001	0.024	0.028
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: *Placebo test. Clustered standard errors by cities below coefficients. The period is the post treatment year. NN denotes nearest-neighbour matching estimator and kernel is the kernel propensity score matching estimator with epanechnikov kernel function. The period is the post treatment year.

Table 2.18 – MDID estimates of LPRA impact on 5th-graders - pre-treatment period 2009

	Reading				Mathematics			
Period	NN		Kernel		NN		Kernel	
	Full	Trim.	Full	Trim	Full	Trim.	Full	Trim
2011	0.118	0.112	0.079	0.082	0.142	0.130	0.111	0.105
	0.000	0.001	0.018	0.021	0.000	0.001	0.021	0.023
2013	0.149	0.143	0.122	0.128	0.144	0.138	0.123	0.119
	0.000	0.001	0.021	0.023	0.000	0.001	0.018	0.021
2015	0.196	0.201	0.189	0.202	0.196	0.194	0.195	0.198
	0.001	0.001	0.024	0.024	0.001	0.001	0.022	0.027
2017	0.174	0.170	0.167	0.169	0.210	0.199	0.206	0.199
	0.001	0.001	0.020	0.024	0.001	0.001	0.022	0.021
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: *Placebo test. Clustered standard errors by cities below coefficients. The period is the post treatment year. NN denotes nearest-neighbour matching estimator and kernel is the kernel propensity score matching estimator with epanechnikov kernel function. The period is the post treatment year.

Table 2.19 – DID estimates of LPRA impact on 9th-grade test scores - pre-treatment period 2007

Period	Reading				Mathematics			
	Full sample		Trim. sample		Full sample		Trim. sample	
2009*	0.354	0.354	0.279	0.290	0.265	0.267	0.176	0.185
	0.064	0.063	0.066	0.068	0.053	0.052	0.054	0.056
2011*	0.243	0.238	0.243	0.265	0.216	0.207	0.201	0.219
	0.063	0.062	0.059	0.057	0.059	0.058	0.059	0.056
2013*	0.323	0.332	0.319	0.347	0.270	0.270	0.260	0.282
	0.064	0.065	0.061	0.061	0.063	0.062	0.061	0.059
2015	0.364	0.362	0.350	0.363	0.315	0.309	0.298	0.309
	0.064	0.064	0.065	0.064	0.068	0.067	0.070	0.068
2017	0.397	0.391	0.345	0.359	0.371	0.362	0.288	0.301
	0.063	0.062	0.064	0.062	0.076	0.074	0.079	0.075
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: *Placebo test. Clustered standard errors by cities below coefficients. The period is the post treatment year.

Table 2.20 – DID estimates of LPRA impact on 9th-grade test scores - pre-treatment period 2009

Period	Reading				Mathematics			
	Full sample		Trim. sample		Full sample		Trim. sample	
2011*	-0.107	-0.115	-0.086	-0.078	-0.043	-0.052	-0.020	-0.016
	0.032	0.032	0.036	0.035	0.031	0.032	0.034	0.033
2013*	-0.023	-0.016	-0.008	0.013	0.012	0.010	0.032	0.044
	0.042	0.040	0.042	0.039	0.042	0.040	0.043	0.040
2015	0.021	0.017	0.044	0.049	0.059	0.054	0.076	0.079
	0.050	0.050	0.053	0.053	0.054	0.054	0.058	0.057
2017	0.061	0.056	0.022	0.028	0.124	0.116	0.060	0.063
	0.049	0.048	0.051	0.050	0.063	0.062	0.068	0.066
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: *Placebo test. Clustered standard errors by cities below coefficients. The period is the post treatment year.

Table 2.21 – DID estimates of LPRA impact on 9th-grade test scores - pre-treatment period 2011

Period	Reading				Mathematics			
	Full sample		Trim. sample		Full sample		Trim. sample	
2013*	0.083	0.122	0.066	0.125	0.055	0.086	0.051	0.088
	0.022	0.018	0.022	0.018	0.021	0.018	0.023	0.020
2015	0.127	0.126	0.093	0.116	0.103	0.106	0.086	0.098
	0.033	0.028	0.035	0.031	0.035	0.030	0.039	0.034
2017	0.167	0.159	0.073	0.100	0.166	0.165	0.068	0.086
	0.033	0.025	0.035	0.027	0.044	0.034	0.048	0.039
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: *Placebo test. Clustered standard errors by cities below coefficients. The period is the post treatment year.

Table 2.22 – DID estimates of LPRA impact on 9th-grade test scores - pre-treatment period 2013

Period	Reading				Mathematics			
	Full sample		Trim. sample		Full sample		Trim. sample	
2015	0.045	0.014	0.024	0.018	0.048	0.030	0.043	0.034
	0.021	0.020	0.025	0.024	0.023	0.022	0.026	0.026
2017	0.083	0.042	0.015	0.012	0.109	0.079	0.033	0.028
	0.021	0.020	0.024	0.023	0.031	0.028	0.035	0.033
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Clustered standard errors by cities below coefficients. The period is the post treatment year.

Table 2.23 – MDID estimates of LPRA impact on 9th-graders - pre-treatment period 2007

Period	Reading				Mathematics			
	NN		Kernel		NN		Kernel	
	Full	Trim.	Full	Trim	Full	Trim.	Full	Trim
2009*	0.312	0.288	0.328	0.301	0.222	0.200	0.242	0.209
	0.001	0.001	0.034	0.031	0.001	0.001	0.023	0.028
2011*	0.361	0.346	0.354	0.329	0.324	0.301	0.315	0.289
	0.001	0.001	0.030	0.032	0.001	0.001	0.024	0.028
2013*	0.428	0.420	0.427	0.407	0.367	0.361	0.364	0.345
	0.001	0.001	0.029	0.031	0.001	0.001	0.027	0.026
2015	0.439	0.416	0.435	0.406	0.377	0.365	0.376	0.359
	0.001	0.001	0.033	0.032	0.001	0.001	0.032	0.040
2017	0.408	0.410	0.420	0.403	0.363	0.360	0.377	0.356
	0.001	0.002	0.035	0.040	0.001	0.002	0.023	0.033
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: *Placebo test. Clustered standard errors by cities below coefficients. The period is the post treatment year. NN denotes nearest-neighbour matching estimator and kernel is the kernel propensity score matching estimator with epanechnikov kernel function. The period is the post treatment year.

Table 2.24 – MDID estimates of LPRA impact on 9th-graders - pre-treatment period 2009

Period	Reading				Mathematics			
	NN		Kernel		NN		Kernel	
	Full	Trim.	Full	Trim	Full	Trim.	Full	Trim
2011*	-0.022	-0.022	-0.051	-0.047	0.035	0.040	0.008	0.013
	0.001	0.001	0.025	0.024	0.001	0.001	0.022	0.022
2013*	0.079	0.068	0.038	0.042	0.106	0.103	0.068	0.075
	0.001	0.001	0.034	0.027	0.001	0.001	0.023	0.027
2015	0.036	0.039	0.016	0.021	0.079	0.073	0.063	0.063
	0.001	0.001	0.027	0.032	0.001	0.001	0.031	0.028
2017	0.054	0.025	0.021	-0.003	0.082	0.065	0.070	0.051
	0.001	0.001	0.034	0.031	0.001	0.001	0.031	0.031
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: *Placebo test. Clustered standard errors by cities below coefficients. The period is the post treatment year. NN denotes nearest-neighbour matching estimator and kernel is the kernel propensity score matching estimator with epanechnikov kernel function. The period is the post treatment year.

Table 2.25 – MDID estimates of LPRA impact on 9th-graders - pre-treatment period 2011

Period	Reading				Mathematics			
	NN		Kernel		NN		Kernel	
	Full	Trim.	Full	Trim	Full	Trim.	Full	Trim
2013*	0.059	0.061	0.063	0.064	0.041	0.040	0.052	0.046
	0.000	0.000	0.012	0.017	0.000	0.000	0.015	0.015
2015	0.081	0.090	0.088	0.085	0.073	0.077	0.095	0.083
	0.000	0.000	0.014	0.018	0.000	0.000	0.019	0.018
2017	0.052	0.040	0.087	0.069	0.052	0.043	0.092	0.073
	0.000	0.000	0.020	0.020	0.000	0.001	0.020	0.019
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: *Placebo test. Clustered standard errors by cities below coefficients. The period is the post treatment year. NN denotes nearest-neighbour matching estimator and kernel is the kernel propensity score matching estimator with epanechnikov kernel function. The period is the post treatment year.

Table 2.26 – MDID estimates of LPRA impact on 9th-graders - pre-treatment period 2013

	Reading				Mathematics			
Period	NN		Kernel		NN		Kernel	
	Full	Trim.	Full	Trim	Full	Trim.	Full	Trim
2015	0.037	0.032	0.033	0.025	0.055	0.054	0.052	0.043
	0.000	0.000	0.017	0.017	0.000	0.000	0.017	0.015
2017	-0.001	0.012	0.008	0.010	0.037	0.047	0.057	0.059
	0.000	0.000	0.018	0.021	0.000	0.000	0.022	0.022
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Clustered standard errors by cities below coefficients. The period is the post treatment year. NN denotes nearest-neighbour matching estimator and kernel is the kernel propensity score matching estimator with epanechnikov kernel function. The period is the post treatment year.

Table 2.27 – Placebo DID estimates of LPRA impact

Subject	5th grade				9th grade			
	All		Neighbors		All		Neighbors	
Reading	-0.097	-0.040	-0.029	0.027	0.011	-0.007	-0.023	-0.018
SE	0.035	0.037	0.053	0.048	0.042	0.034	0.061	0.061
Observations	5150	4692	2864	2609	2567	2503	1525	1493
Mathematics	-0.100	-0.062	-0.035	0.011	-0.028	-0.035	-0.059	-0.056
SE	0.036	0.035	0.056	0.045	0.047	0.047	0.075	0.073
Observation	5295	4955	2947	2755	2607	2531	1570	1526
Stu. covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Clustered standard errors by states below coefficients. The regression used student-data level and did not considered sample weights. All refers to the control group formed by all northeastern states, and neighbors refers to bordering states. The 5th-grade sample considers students that reported never repeated a grade; the 9th-grade sample considers students that never repeated and that only studied in public schools.

Table 2.28 – DID estimates of LPRA's first phase impact on 5th-grade test scores - pre-treatment period 2007

	Reading		Mathematics	
Coefficent	0.003	0.003	-0.013	-0.010
SE	0.047	0.045	0.044	0.044
Observations	3915	3868	3915	3868
Stu. covariates	No	Yes	No	Yes

Notes: Standard errors (SE) are clustered by cities. The the post treatment period is 2009. The sample contains only Ceara schools; we compare schools from the 56 first-phase-cities with those in the other cities from the state.

3 Spillover effect of a literacy program on Brazilian students achievement

Abstract

This paper intends to verify if the teacher training of Literacy Program at the Right Age (LPRA) enhanced the instruction quality and improved students' achievement in grades not target by the program. We use data from the Ceara State Evaluation System and School Census and estimated educational production functions to check if direct contact between trained or in-training teacher and students from grades 3 to 5 contributed to boost achievement in Reading and Mathematics. Our results indicate the absence of spillovers effects, suggesting that the training indeed focused in literacy instructional methods and the use of the program's materials, that probably apply better for young children, aged 6 to 7, not included teaching behaviors that are expected to apply generically to all school subjects.

Keywords: spillover effects, teacher training, teacher professional development.

3.1 Introduction

This paper intends to verify if the teacher training of Literacy Program at the Right Age (LPRA)[1] enhanced the instruction quality of Reading and Mathematics and improved students' achievement in grades not target by the program. In other words, we intend to check if the training improved teaching in general or only teaching for the beginning reading stage. We use data from the Ceara State Evaluation System and School Census for a non-treated cohort not evaluated by the National Evaluation System of Basic Education (NES) - pupils in second grade in 2007 that attained fifth grade in 2010 or 2011.

Briefly, the LPRA is a successful educational policy in improving the achievement of low-income students, at least, in the short run. In the previous chapter, we discussed the importance of being literate at the right age and showed that the intervention increased Reading and Mathematics abilities of all treated cohorts that were evaluated in fifth-grade by the NES. Our results are aligned with the results of Costa e Carnoy (2015), Kasmirski, Gusmao e Ribeiro (2015) and Lavor e Arraes (2014) that evaluated the program's impact on the first treated cohort.

Until 2010, the main goal of the program was to eliminate child illiteracy, and the

[1] Programa Alfabetização na Idade Certa (Paic).

state and municipalities offices developed a professional development (PD)[2] program with a focus on practical strategies for teaching reading (SEDUC-CE, 2012). If this in-service training improved teacher quality in general, the direct contact with trained teachers should not just increase Reading and Mathematics achievement of treated pupils, but also it could spillover for those not treated as well. Students from third to fifth grades might have direct contact with trained teachers because they are polyvalent, can instruct any class from the elementary cycle, and can teach for different grades in the same year.

Practically all levels of governments rely on PD as a way to improve teacher pedagogical and subject-matter knowledge and, in turn, to enhance student achievement. Indeed, in-service training is a critical mediator in the effectiveness of policy for teachers and students (DESIMONE, 2009; SLAVIN et al., 2009). Generally, state and city school systems can intervene only in PD. For example, in Brazil, the universities and the federal government are in charge of pre-service education, and informal training[3] is out of any Education offices control. Besides, many countries have a considerable stock of teachers, with initial education concluded, and have many years of work ahead. Barretto (2015) discussed teacher education policies and found that Brazilian elementary school teachers were not adequately prepared. She explains that a considerable portion of teachers obtained their undergraduate degree online and in private educational institutions that expanded their vacancies in a fast and improvised way without effective development of scientific knowledge production capacity. The Pedagogy courses are excessively generic and lack focus on instruction methods (BARRETTO, 2015). Even in developed countries, continuing development and learning of teachers is one of the most critical targets of education reforms (DESIMONE, 2009). Hence, identify and evaluate professional development initiatives is crucial to improve the quality of the existing teacher labor force.

Usually, the literature that evaluated the impact of PD found no effect on student learning. This is the case of Jacob e Lefgren (2004), Harris e Sass (2011), Garet et al. (2008), Garet et al. (2010) and, Garet et al. (2011), studies that utilize convincing empirical strategies. Jacob e Lefgren (2004) found no effect of a PD program for low-performing schools on student achievement using a regression discontinuity design. This program consisted of an academic probation that provided elementary schools from Chicago with special funding for staff development, technical assistance, and enhanced monitoring. Harris e Sass (2011) found that formal in-service training has little or no effect on teacher quality in elementary schools. These authors were able to overcome three methodological challenges in estimating the effects of training on teacher value added - non-random as-

[2] We are using professional development (PD) as a synonym to formal in-service training provided by employers. PD does not include formal education acquired through undergraduate and graduate courses.

[3] Harris e Sass (2011) define informal training as the knowledge acquired through on-the-job experience. Desimone (2009) gives as an example the informal "hallway" discussions with other colleagues about instruction techniques.

signment of students and teachers to classrooms, self-selection of schools and teachers to training and lack of detailed data about the various types of training teachers receive and that link the training of teachers to the achievement of the students they teach - by using an extensive panel data set of school administrative records from Florida. The panel data allowed them to include school, teacher, and student fixed effects into their models and test for dynamic student-teacher sorting.

Garet et al. (2008), Garet et al. (2010) and Garet et al. (2011) conducted randomized controlled trials to evaluate PD programs for early reading teachers and for middle school math teachers after one and two years of the training implementation, respectively. Although Garet et al. (2008) detected positive impacts on teacher's knowledge of scientifically based reading instruction and on one of the three instructional practices promoted by the training, the intervention did not improve the students' performance in Reading. Garet et al. (2010) and Garet et al. (2011) found no impact of a training designed to improve teacher knowledge of rational number topics on student achievement in Mathematics. These papers illustrate the difficulties involved in studying PD direct impacts and suggest that spillovers can be even harder to detect. Our empirical strategy consist of estimating a educational production function based on Harris e Sass (2011).

Our results indicate the absence of spillovers effects on student not target by LPRA, suggesting that the training indeed focused in literacy instructional methods and the use of the program's materials. Possibly, it not included teaching behaviors that are expected to apply generically to all school subjects. The impact's signs suggest that the knowledge acquired would depreciate over time or become less relevant as the pupils get older, that face-to-face could be better than distance training and that contemporaneous PD might help to improve the target subject, probably because of the assistance provided to teachers and the in-person interaction between trainers and teachers from priority cities, which could have helped to solve doubts about how to instruct children target or not by the literacy intervention.

The remainder of this paper is organized as follows. Nest section provides background on the LPRA's training. Section 3 describes our data and Section 4 explains our empirical strategy. Section 5 presents findings on spillovers of trained teachers and examines the heterogeneity in effects across students. Section 6 discusses some of the implications of our results and concludes.

3.2 LPRA's training

Before the LPRA's implementation, the Ceara Committee on the Elimination of School Illiteracy diagnosed that Ceara teachers were not able to teach children how to read and write properly and used outdated methods to teach literacy (MARQUES, 2018). The

Committee understood that a structured material was necessary to provide to teachers a methodological structure to instruct correctly (SEDUC-CE, 2012).

The state and municipalities offices developed the teacher training and the structured material aligned with the official curriculum and other elements of LPRA. Based on the Committee diagnosis, the state office prepared the first-grade from scratch and selected the second-grade materials among those already available by publishing companies in 2007 and 2008 (SEDUC-CE, 2012; SEDUC-CE, 2015). According to SEDUC-CE (2012), the materials provide a daily routine of classroom and homework activities, guiding teachers and students. All literacy teachers should be trained to use these materials.

As described by SEDUC-CE (2012), for the first grade, the training occurs as follows: the state trains professionals from city offices and, in turn, these people train and monitor teachers through face-to-face and distance activities. In cities with low average achievement, called priority, there is extra help. The teams from central and regional state departments have gone to these municipalities to get together with the mayor, Education secretary and municipal departments responsible for LPRA, visit the schools, watch the classes, talk to teachers and principals, and, most importantly, train teachers entirely face-to-face. The training for second-grade teachers was done by professionals hired by the publishing companies chosen to provide the educational material and monitored by regional and local LPRA's teams (SEDUC-CE, 2012).

Initially, the activities occurred monthly with a minimum workload of 8 hours (MARQUES, 2018). This workload can be considered high in comparison with other on-the-job training implemented in U.S. (JACOB; LEFGREN, 2004) and it could be associated with more significant improvement of teacher knowledge and, in turn, with the existence of positive spillovers, on the one hand. But, the bigger the workload, the more time PD takes away from classroom instruction or preparation time, on the other side.

The training meetings should occur preferably on Saturdays, but, if it was not possible, the meeting could happen on class days, and the school should or replace the teacher in training temporarily with a substitute or add more school days to the calendar (SEDUC-CE, 2012). If a sufficiently high portion of teachers missed classes and were substituted by colleagues less effective or unable to maintain the continuity of instruction in the permanent teacher's absence, contemporaneous negative spillovers could occur. But, given that the PD focused on literacy and should have provided some assistance for teachers, particularly for those in priority cities, the contemporaneous spillover could be positive for Reading and negative for other subjects. While we can distinguish ongoing from complete training, unfortunately, we could not separate the face-to-face training from PD types for all years investigated.

Marques (2018) analyzed the program's teacher training and verified that the State government used as a training model the effective professional development, even without

having a clear conception of its meaning. Desimone (2009) defines effective PD as the teacher learning opportunities with at least the following characteristics: content focus, active learning, coherence, duration, and collective participation. According to the author, these features are critical to improving teacher knowledge, skills, practice, and ultimately the student learning.

As the most non-treated students were not literate adequately at the right age, according to the Committee's diagnosis, the knowledge obtained from the PD could help teachers to solve learning problems in Reading of children not focused by the LPRA. Regarding Math, if, besides the use of the program's materials and literacy instructional methods, the training covered teaching behaviors that are expected to apply generically to all school subjects, the training could yield positive spillovers on other subjects. This spillover seems unlikely, though, because of the content focus identified by Marques (2018).

3.3 Data and samples

We utilize datasets provided by the Ceara State Education Department that merge information from the State Evaluation System and School Census. The State Evaluation System (EES) is a large-scale external assessment of Ceara student learning in key grades. Created in 1992, the system was implemented gradually, and nowadays it evaluates every year second grade in Reading, fifth and ninth grades of elementary and middle levels and all grades of high school in Reading and Mathematics. All local and state schools participate and teachers, principals and students from fifth-grade on answer a contextual questionnaire.

The student socioeconomic characteristics - gender, color, and economic status - came from the contextual questionnaire. We applied an adapted version of Brazil Criteria to identify the pupils economic status, because of the questionnaire from 2010 missed information about parents education. The original criteria, that estimate the purchasing power of families, is based on ownership of comfort items and head of household education ABEP (2011). We were able to identify priority cities in 2007 and 2008 based on their mean test score in Reading and students who changed schools merging the data from different years.

The School Census is a yearly survey that collects data on each student, on each teacher who is in classroom and on each class of every school in the country (INEP, 2019). These data are inputs for the formulation of public policies and the accountability necessary to the distribution of public resources. It is conducted in cooperation with State and Municipal Boards of Education and is answered by principals of all public and private schools. The Census methodology changed drastically in 2007, when the identification of students and teachers become unique and time-invariant, breaking the historical series

started in 1995.

In order to identify the training's effect, we use data for the cohort of pupils who were in second grade in 2007 and managed to achieve to fifth grade in 2010 (students that did not repeat a grade) or 2011 (students that repeated once third or fourth grade). This cohort did not participate in LPRA, as the program started in 2008 for second graders and -graders in 2009. We do not know which teachers participated in the training exactly and we needed to identify them indirectly. A teacher from this cohort was considered trained by the program if she taught to the program's target grades between 2008 and 2010 (or 2011) and a teacher was considered not treated otherwise. Ideally, we should know who took part in the training as well as the number of hours spent on it.

We identify the students that have direct contact with trained or in training teachers in third, fourth or fifth grades, and also the number of years they interacted with each other. We restricted the student sample to the pupils with only one teacher per year found in the School Census each year investigated. For example, a student for cohort 2007-2010 was included in the sample if she had one teacher in 2008, one in 2009 and one in 2010 e her teachers were found in the database in these years. This restriction helps to guarantee that the student spent most of each year with the same teacher and that we know the grades from all teachers classes.

Table 3.1 – Descriptive statistics of student sample by cohort

Covariate	All sample		2007-2010		2007-2011	
	Mean	SD	Mean	SD	Mean	SD
Contact with one trained teacher	0.25	0.43	0.24	0.42	0.32	0.47
... two trained teachers	0.06	0.23	0.04	0.21	0.14	0.34
... three or four trained teachers	0.01	0.10	0.00	0.07	0.05	0.22
... trained teacher at 3rd-grade	0.06	0.23	0.05	0.21	0.14	0.35
... trained teacher at 4th-grade	0.13	0.34	0.12	0.33	0.22	0.41
... trained teacher at 5th-grade	0.17	0.38	0.17	0.38	0.21	0.40
... teacher in training at 3rd-grade	0.05	0.22	0.05	0.21	0.11	0.31
... teacher in training at 4th-grade	0.09	0.29	0.08	0.28	0.14	0.34
... teacher in training at 5th-grade	0.09	0.29	0.09	0.29	0.11	0.32
Face-to-face training	0.05	0.23	0.05	0.21	0.12	0.33
Priority city	0.19	0.39	0.18	0.38	0.26	0.44
Reading raw score	184.25	42.50	185.77	42.34	173.47	42.07
Prior reading raw score	134.72	69.37	141.33	68.42	87.71	56.69
Mathematics raw score	198.20	46.55	198.94	46.61	192.98	45.79
Girl	0.51	0.50	0.53	0.50	0.38	0.48
Black	0.67	0.47	0.67	0.47	0.70	0.46
Economic status A or B	0.00	0.06	0.00	0.01	0.02	0.15
Economic status C	0.15	0.35	0.11	0.31	0.42	0.49
Economic status D	0.66	0.47	0.69	0.46	0.42	0.49
Economic status E	0.19	0.39	0.20	0.40	0.14	0.35
Morning shift	0.48	0.50	0.48	0.50	0.47	0.50
School mover	0.27	0.44	0.25	0.44	0.35	0.48
Observations	31781		27862		3919	

Notes: Excepting the test scores, all variables are binaries.

Table 3.2 – Descriptive statistics of students sample by contact status

Covariate	No contact		With contact	
	Mean	SD	Mean	SD
Contact with one trained teacher	0.00	0.00	0.79	0.41
... two trained teachers	0.00	0.00	0.18	0.38
... three or four trained teachers	0.00	0.00	0.03	0.18
... trained teacher at 3rd-grade	0.00	0.00	0.18	0.39
... trained teacher at 4th-grade	0.00	0.00	0.43	0.49
... trained teacher at 5th-grade	0.00	0.00	0.56	0.50
... teacher in training at 3rd-grade	0.00	0.00	0.17	0.38
... teacher in training at 4th-grade	0.00	0.00	0.29	0.45
... teacher in training at 5th-grade	0.00	0.00	0.30	0.46
Face-to-face training	0.00	0.00	0.17	0.38
Priority city	0.19	0.40	0.17	0.38
Reading raw score	184.79	42.72	183.07	42.00
Prior reading raw score	137.17	69.55	129.32	68.64
Mathematics raw score	198.82	46.74	196.86	46.11
Girl	0.51	0.50	0.49	0.50
Black	0.67	0.47	0.67	0.47
Economic status A or B	0.00	0.05	0.01	0.07
Economic status C	0.14	0.34	0.17	0.37
Economic status D	0.67	0.47	0.64	0.48
Economic status E	0.19	0.39	0.18	0.39
Morning shift	0.52	0.50	0.39	0.49
School mover	0.26	0.44	0.27	0.44
Observations	21842		9939	

Notes: Excepting the test scores, all variables are binaries.

Student descriptive statistics can be seen in tables 3.1 and 3.2. About 31% of the sample had direct contact with at least one teacher that participated or were participating in the PD, most interactions occurred just once, when the children were at 4th or 5th-grade and while the teachers were receiving the training. When the students were in third grade, almost all interactions occurred while the teachers were taking part of the PD, which means that a 3rd-grader only could have contact with a teacher whose PD occurred at a previous year if he was doing this grade for the second time. The cohort 2007-2011 have higher average test scores than cohort 2007-2011, which is expected since the students that arrived at 5-th grade in 2011 repeated once, and a higher proportion of girls.

In tables 3.3 and 3.4, we present descriptive statistics of teachers. One-third of them taught to first or second-graders of elementary school at least once between 2008 and 2010 (or 2011 for those who teach students from cohort 2007-2011), excepting grade 1 in 2008. We could not merge the teacher questionnaire from EES with the database that links them with their students. Thus, all the information we have came from the School

Table 3.3 – Descriptive statistics of teacher sample

Covariate	All sample		Trained		Not trained	
	Mean	SD	Mean	SD	Mean	SD
Trained	0.34	0.47	1.00	0.00	0.00	0.00
Less than 30 years old	0.17	0.38	0.15	0.36	0.18	0.38
30 to 40 years old	0.40	0.49	0.43	0.49	0.39	0.49
40 to 50 years old	0.35	0.48	0.35	0.48	0.34	0.47
50 to 60 years old	0.08	0.27	0.06	0.24	0.08	0.28
More than 60 years old	0.01	0.08	0.00	0.05	0.01	0.09
Women	0.88	0.32	0.94	0.23	0.85	0.36
Work at state school	0.04	0.20	0.03	0.16	0.05	0.22
High school	0.30	0.46	0.31	0.46	0.30	0.46
High education	0.69	0.46	0.69	0.46	0.70	0.46
Post graduation	0.15	0.36	0.13	0.33	0.17	0.37
Teaching certificate	0.68	0.47	0.67	0.47	0.68	0.47
Observations	12139		4077		8062	

Notes: Excepting the age, all variables are binaries.

Table 3.4 – Proportion of trained teacher by number of years, dates and cohorts

Covariate	All		Only 2007-2010		Both cohorts	
	Mean	SD	Mean	SD	Mean	SD
Trained in 2008	0.30	0.46	0.21	0.41	0.36	0.48
Trained in 2009	0.49	0.50	0.49	0.50	0.49	0.50
Trained in 2010	0.51	0.50	0.53	0.50	0.50	0.50
Trained in 2011	0.49	0.50	0.51	0.50	0.48	0.50
Years of training	1.72	0.82	1.70	0.82	1.74	0.83
One year of training	0.52	0.50	0.53	0.50	0.50	0.50
Two years of training	0.24	0.43	0.24	0.43	0.25	0.43
Three years of training	0.24	0.43	0.23	0.42	0.25	0.43
Observations	4077		1765		2312	

Notes: All trained teachers have students from 2007-2010 cohort and 62% of them also have students from 2007-2011 cohort.

Census (age, gender, and formal education). Age is our proxy for experience, and formal pre and post-service education is captured by binary variables that indicate if the teacher completed university or high school for those without a bachelor degree, have a teaching degree and post graduation.

3.4 Empirical strategy

We can capture the program's spillovers through an education production function inspired by Harris e Sass (2011):

$$y_{i,t+k} = \alpha_0 + \boldsymbol{\gamma} \boldsymbol{D}_i + y_{i,t}\alpha_1 + \boldsymbol{x}'_{it}\boldsymbol{\alpha}_2 + \boldsymbol{z}'_{jt}\boldsymbol{\alpha}_3 + \mu_s + \epsilon_{it} \tag{3.1}$$

$y_{i,t+k}$ refers to fifth grade standardized test score[4] of student i in year $t+k$, $k = 3, 4$ depending if the pupil repeated or not the third or fourth grade, $y_{i,t}$ is the student achievement in second grade, $t = 2007$, $\boldsymbol{x}_{it}$ is a vector of student characteristics (economic status in period $t+k$, gender, school move between t and $t+k$ etc.), $\boldsymbol{z}_{it}$ contains teacher characteristics (age and education, for example), and μ_s is a school fixed effect.

$\boldsymbol{D}_i$ is a set of dummies that says if the student i had direct contact with in-training or trained teachers between t and $t+k$ in grades 3, 4 and 5. We interact these dummies with an indicator of contemporaneous PD to verify how the fact that the interaction occurred simultaneously to the PD would influence our outcome variables. We also check if there is a difference between students who had direct contact with only one trained teacher and those who had two or more trained teachers.

Jacob e Lefgren (2004) alert to the potential presence of bias in $\boldsymbol{\gamma}$ due to selection of schools and teachers into training based on unobserved characteristics. While all schools that offer first and second grades should participate, teachers can change the grades of their classes to receive or to avoid the PD, and this self-selection can be correlated with their motivation, which will tend to bias $\boldsymbol{\gamma}$ upward. Additionally, Harris e Sass (2011) call attention to another source of bias - the possible correlations between observed teacher attributes and unobserved student characteristics due to the non-random assignment of students and teachers to classrooms.

To alleviate both kinds of bias, the mentioned authors suggest the inclusion of teacher and student fixed effects, which, together with the school fixed effect, would rule out selection based on time-invariant characteristics. Unfortunately, there are no data available for another cohort of students that were not exposed to the program, because the LPRA test and the new School Census began in 2007, and we ended up with a cross-section.

[4] For each year and grade, we subtracted the mean from the raw score and then divided it by its standard deviation.

3.5 Results

The estimates of the equation 3.1 for Reading are in table 3.5 and for Mathematics in 3.6. The first four columns show the spillovers impacts controlling for student covariates and school fixed effects and the next four columns present the estimates that add teacher variables as controls. The results indicate the absence of spillovers effects of direct contact between student from grades 3 to 5 and trained teachers. This can be due to the training content focus - literacy instruction methods and use of the program's materials - that probably apply better for young children, aged 6 to 7, and did not cover teaching behaviors that could apply generically to other school subjects.

Keeping in mind the insignificant estimates, we can note that the contact with just one trained teacher would marginally reduce the average test scores of both subjects (columns 1 and 5 of the tables 3.5 and 3.6), but the contact with two or more teachers would slightly improve student achievement (column 2 and 6), suggesting that the learning problems caused by not being literate at the right age can not be solved during just one year of the upper elementary reading stage.

The sign of spillovers seems to be related to the timing of the interaction. The contact at 3th-grade is associated with slightly worse learning (columns 3 and 7), but when we separate the pupils by type of PD participation (columns 4 and 8) - contemporaneous or not - we see that the students who repeated the grade drive this result and that the effect of the teacher in-training is actually positive. The scenario is inverted for grade 4, where the PD acquisition in the current year decreases the learning. The coefficients for 5th-grade in columns 3 and 7 differ by subject, being positive for Reading and negative for Mathematics, but in columns 4 and 8 we can note that the contemporaneous training increased achievement for both subjects, but not enough to revert the negative effect for Mathematics.

The Reading positive spillover of contemporaneous PD is intuitive because training included monitoring activities that could have solved teacher doubts and allowed corrections of mistakes rapidly. In addition, the content included the use of the Reading materials, what could help saving time of class preparation of grades 1 and 2, leaving more time for preparing classes of other grades. The pattern for Reading in 4th-grade is unexpected. The positive effects of PD happening at the same time of the interaction in grades 3 and 5 for Math was not expected, since the time spent with the training activities would take away time from classroom instruction or preparation time.

Table 3.5 – Effect of direct contact with trained teachers - Reading

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Student								
Contact with one trained teacher	-0.00617	-0.00681			-0.00700	-0.00811		
	(0.0190)	(0.0186)			(0.0189)	(0.0185)		
... two trained teachers		0.0230				0.0225		
		(0.0349)				(0.0346)		
... three or four trained teachers		0.0856				0.0847		
		(0.0627)				(0.0624)		
... trained teacher at 3rd-grade			-0.0286	-0.181*			-0.0304	-0.177
			(0.0344)	(0.0905)			(0.0342)	(0.0907)
... trained teacher at 4th-grade			0.00472	0.0203			0.00421	0.0215
			(0.0229)	(0.0407)			(0.0228)	(0.0405)
... trained teacher at 5th-grade			0.00866	-0.0281			0.00850	-0.0271
			(0.0271)	(0.0340)			(0.0266)	(0.0335)
... teacher in training at 3rd-grade				0.164				0.158
				(0.0972)				(0.0973)
... teacher in training at 4th-grade				-0.0318				-0.0342
				(0.0502)				(0.0500)
... teacher in training at 5th-grade				0.0704				0.0684
				(0.0405)				(0.0404)
Face-to-face training	0.0191	0.0205	0.0168	0.0254	0.0155	0.0183	0.0140	0.0226
	(0.0610)	(0.0549)	(0.0552)	(0.0546)	(0.0606)	(0.0547)	(0.0550)	(0.0543)
Prior reading score	0.381***	0.376***	0.376***	0.376***	0.381***	0.376***	0.376***	0.376***
	(0.00759)	(0.00737)	(0.00736)	(0.00736)	(0.00760)	(0.00737)	(0.00736)	(0.00736)
Girl	0.113***	0.116***	0.116***	0.117***	0.113***	0.117***	0.117***	0.117***
	(0.0104)	(0.0101)	(0.0100)	(0.0101)	(0.0104)	(0.0101)	(0.0100)	(0.0101)
Black	0.0451***	0.0496***	0.0497***	0.0497***	0.0447***	0.0492***	0.0493***	0.0493***
	(0.0108)	(0.0104)	(0.0104)	(0.0104)	(0.0108)	(0.0104)	(0.0104)	(0.0104)
Economic status D	0.0261	0.0271	0.0278	0.0271	0.0263	0.0276	0.0285	0.0278
	(0.0170)	(0.0163)	(0.0163)	(0.0163)	(0.0171)	(0.0164)	(0.0164)	(0.0163)
Economic status E	-0.239***	-0.245***	-0.244***	-0.244***	-0.238***	-0.243***	-0.242***	-0.243***
	(0.0194)	(0.0187)	(0.0187)	(0.0187)	(0.0195)	(0.0187)	(0.0187)	(0.0187)
Morning shift	0.0401*	0.0361*	0.0360*	0.0371*	0.0378*	0.0344*	0.0343*	0.0354*
	(0.0166)	(0.0162)	(0.0162)	(0.0162)	(0.0165)	(0.0161)	(0.0161)	(0.0161)
School mover	0.0154	0.0168	0.0167	0.0171	0.0158	0.0172	0.0170	0.0174
	(0.0162)	(0.0152)	(0.0152)	(0.0152)	(0.0161)	(0.0151)	(0.0152)	(0.0152)
2007-2010 cohort	0.281***	0.271***	0.263***	0.260***	0.279***	0.269***	0.261***	0.258***
	(0.0228)	(0.0221)	(0.0223)	(0.0225)	(0.0229)	(0.0223)	(0.0224)	(0.0226)
Teacher								
Less than 30 years old					-0.00784	-0.00132	-0.00325	-0.00674
					(0.0534)	(0.0506)	(0.0508)	(0.0506)
30 to 40 years old					-0.0359	-0.0290	-0.0310	-0.0341
					(0.0482)	(0.0457)	(0.0458)	(0.0459)
40 to 50 years old					-0.0436	-0.0364	-0.0374	-0.0402
					(0.0464)	(0.0442)	(0.0442)	(0.0443)
Women					0.0871**	0.0922**	0.0921**	0.0906**
					(0.0330)	(0.0328)	(0.0328)	(0.0327)
Work at state school					0.0318	0.0247	0.0256	0.0285
					(0.0591)	(0.0567)	(0.0570)	(0.0572)
High school					0.263	0.290	0.293	0.294
					(0.188)	(0.164)	(0.164)	(0.164)
High education					0.269	0.286	0.287	0.291
					(0.197)	(0.173)	(0.173)	(0.173)
Post graduation					0.0163	0.0200	0.0199	0.0197
					(0.0296)	(0.0285)	(0.0286)	(0.0285)
Teaching certificate					0.0500	0.0426	0.0426	0.0389
					(0.0664)	(0.0635)	(0.0637)	(0.0634)
Observations	29,681	31,781	31,781	31,781	29,681	31,781	31,781	31,781
R-squared	0.186	0.185	0.185	0.185	0.186	0.185	0.185	0.186
Number of cd_escola	3,268	3,391	3,391	3,391	3,268	3,391	3,391	3,391

Notes: Robust standard errors clustered by schools in parentheses.
*** p<0.001, ** p<0.01, * p<0.05.
The omitted categories of economic status are A, B and C.
All models controls for school fixed effects.
Columns (1) and (5) compares students that had contact with just one trained teacher with those who haven't.
The remaining columns compares students that had any contact with trained teacher, either once or more times, with those who haven't.

Table 3.6 – Effect of direct contact with trained teachers - Mathematics

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Student								
Contact with one trained teacher	-0.0178	-0.0162			-0.0158	-0.0149		
	(0.0211)	(0.0206)			(0.0212)	(0.0207)		
... two trained teachers		0.00344				0.00709		
		(0.0352)				(0.0351)		
... three or four trained teachers		0.0257				0.0331		
		(0.0647)				(0.0648)		
... trained teacher at 3rd-grade			-0.00994	-0.151			-0.0107	-0.148
			(0.0346)	(0.0994)			(0.0346)	(0.0983)
... trained teacher at 4th-grade			0.00794	0.00897			0.00741	0.0104
			(0.0240)	(0.0384)			(0.0240)	(0.0388)
... trained teacher at 5th-grade			-0.0296	-0.0311			-0.0236	-0.0234
			(0.0273)	(0.0360)			(0.0271)	(0.0357)
... teacher in training at 3rd-grade				0.157				0.153
				(0.106)				(0.105)
... teacher in training at 4th-grade				-0.00166				-0.00427
				(0.0468)				(0.0469)
... teacher in training at 5th-grade				0.000863				-0.00237
				(0.0442)				(0.0440)
Face-to-face training	0.0230	0.0400	0.0359	0.0378	0.0175	0.0378	0.0333	0.0348
	(0.0561)	(0.0531)	(0.0518)	(0.0518)	(0.0568)	(0.0536)	(0.0523)	(0.0523)
Prior reading score	0.353***	0.350***	0.350***	0.350***	0.353***	0.349***	0.349***	0.349***
	(0.00734)	(0.00710)	(0.00709)	(0.00710)	(0.00732)	(0.00708)	(0.00708)	(0.00708)
Girl	-0.154***	-0.150***	-0.150***	-0.150***	-0.153***	-0.149***	-0.149***	-0.149***
	(0.0101)	(0.00981)	(0.00979)	(0.00980)	(0.0101)	(0.00980)	(0.00978)	(0.00979)
Black	0.0487***	0.0498***	0.0499***	0.0499***	0.0483***	0.0494***	0.0495***	0.0494***
	(0.0110)	(0.0108)	(0.0108)	(0.0108)	(0.0110)	(0.0108)	(0.0108)	(0.0108)
Economic status D	-0.00821	-0.00873	-0.00855	-0.00891	-0.00829	-0.00829	-0.00803	-0.00838
	(0.0167)	(0.0160)	(0.0161)	(0.0160)	(0.0168)	(0.0160)	(0.0161)	(0.0160)
Economic status E	-0.283***	-0.283***	-0.282***	-0.282***	-0.283***	-0.282***	-0.281***	-0.281***
	(0.0197)	(0.0192)	(0.0192)	(0.0192)	(0.0197)	(0.0192)	(0.0192)	(0.0192)
Morning shift	0.0118	0.0159	0.0154	0.0153	0.00703	0.0118	0.0114	0.0113
	(0.0181)	(0.0178)	(0.0178)	(0.0178)	(0.0182)	(0.0179)	(0.0179)	(0.0179)
School mover	-0.0275	-0.0177	-0.0178	-0.0177	-0.0266	-0.0169	-0.0171	-0.0170
	(0.0161)	(0.0152)	(0.0152)	(0.0152)	(0.0160)	(0.0151)	(0.0151)	(0.0151)
2007-2010 cohort	0.253***	0.236***	0.234***	0.231***	0.257***	0.239***	0.236***	0.233***
	(0.0234)	(0.0235)	(0.0235)	(0.0235)	(0.0236)	(0.0237)	(0.0237)	(0.0237)
Teacher								
Less than 30 years old					0.0422	0.0432	0.0430	0.0434
					(0.0552)	(0.0521)	(0.0521)	(0.0523)
30 to 40 years old					-0.0265	-0.0221	-0.0211	-0.0210
					(0.0485)	(0.0461)	(0.0461)	(0.0461)
40 to 50 years old					-0.0216	-0.0239	-0.0232	-0.0229
					(0.0472)	(0.0451)	(0.0451)	(0.0452)
Women					0.0394	0.0497	0.0506	0.0505
					(0.0358)	(0.0346)	(0.0346)	(0.0346)
Work at state school					0.00808	0.0115	0.0104	0.0110
					(0.0596)	(0.0588)	(0.0587)	(0.0590)
High school					-0.0619	-0.0594	-0.0534	-0.0567
					(0.180)	(0.163)	(0.164)	(0.163)
High education					-0.0187	-0.0263	-0.0205	-0.0226
					(0.193)	(0.175)	(0.176)	(0.176)
Post graduation					0.0789**	0.0826**	0.0816**	0.0812**
					(0.0286)	(0.0282)	(0.0282)	(0.0281)
Teaching certificate					0.0182	0.00130	-1.00e-05	-0.00143
					(0.0647)	(0.0617)	(0.0620)	(0.0619)
Observations	29,682	31,782	31,782	31,782	29,682	31,782	31,782	31,782
R-squared	0.160	0.157	0.157	0.157	0.161	0.158	0.158	0.158
Number of cd_escola	3,268	3,391	3,391	3,391	3,268	3,391	3,391	3,391

Notes: Robust standard errors clustered by schools in parentheses.

*** p<0.001, ** p<0.01, * p<0.05.

The omitted categories of economic status are A, B and C.

All models controls for school fixed effects.

Columns (1) and (5) compares students that had contact with just one trained teacher with those who haven't.

The remaining columns compares students that had any contact with trained teacher, either once or more times, with those who haven't.

The sign pattern of the training completed and put to use in the classroom over the grades (columns 4 and 8) suggests that the knowledge acquired depreciated or became less relevant as the students got older. The 3th-grade sign reflects the fact that just students that repeated the grade could have contact with teachers whose training had finished and tell us nothing about the non-contemporaneous effect of the PD. The positive impact in grade 4 and negative in 5 seems to indicate that spillover vanish.

The face-to-face training, a dummy that indicates if the students were in priority cities in 2007 or 2008 and simultaneously had contact with teacher exposed to PD, is associated with a small increase of student learning in general. This finding point that in-person interactions are better than distance ones. The rest of pupil variables present the expected or commonly reported effects - prior achievement is positively related to learning, girls do better in Reading and worse in Math, wealthier, and 2007-2010 cohort students do better. The only exception is the effect of color. Usually, it would be negatively associated with learning due to its correlation with family background.

Regarding the teacher covariates, their inclusion did not change the results already discussed, and practically all of them are not related to achievement, which is aligned with the literature. The exceptions are the effect of gender - female teachers seem to improve Reading learning - and post-graduation - teachers with mostly specialization seems to increase Math performance. This last relation cited could be explained by the content of the post graduation courses attended by this sample, though we do not have information.

Tables 3.7 and 3.8 presents heterogeneous effects of the contact with trained teachers. The only statistically and almost pedagogically relevant effect is the negative spillover on the 20% of students with the highest test scores in both disciplines (columns 1). It is worth mention that the first-quintile students would have a positive spillover, suggesting that the PD content is more suitable to the pupils not fully literate at the right time and that progressed to the next grades, which is aligned with the LPRA. These two results together indicate that the attempt to help the low-achievers warm the higher-ones (those who probably were literate at second grade), at least in the manner that was done. We can affirm that the general negative spillovers detected above could be explained by the in-service training effects on high-achievement, female, not black and poorer students. All controls variables that were omitted from the tables 3.7 and 3.8 presented coefficients similar to the ones in tables 3.5 and 3.6.

Table 3.7 – Heterogeneous effects of direct contact with trained teachers - Reading

VARIABLES	(1)	(2)	(3)	(4)
Contact with trained teacher	0.0353	0.000639	-0.00870	0.0192
	(0.0240)	(0.0209)	(0.0239)	(0.0309)
CTT x 1st-q prior read. score	0.0131			
	(0.0328)			
CTT x 2nd-q prior read. score	-0.0459			
	(0.0270)			
CTT x 4th-q prior read. score	-0.00967			
	(0.0262)			
CTT x 5th-q prior read. score	-0.117***			
	(0.0318)			
CTT x girl		-0.00806		
		(0.0212)		
CTT x black			0.00765	
			(0.0229)	
CTT x ES D				-0.0227
				(0.0328)
CTT x ES E				-0.0462
				(0.0383)
Observations	31,781	31,781	31,781	31,781
R-squared	0.186	0.185	0.185	0.185
Number of cd_escola	3,391	3,391	3,391	3,391

Notes: Robust standard errors clustered by schools in parentheses.
*** p<0.001, ** p<0.01, * p<0.05.
All models controls for school fixed effects, student and teachers covariates.
CTT means contact with trained teacher.
All models compares students that had any contact with trained teacher,
either once or more times, with those who haven't.

Table 3.8 – Heterogeneous effects of direct contact with trained teachers - Mathematics

VARIABLES	(1)	(2)	(3)	(4)
Contact with trained teacher	0.0329	-0.00383	-0.0172	-0.00469
	(0.0272)	(0.0233)	(0.0261)	(0.0314)
CTT x 1st-q prior read. score	-0.00893			
	(0.0322)			
CTT x 2nd-q prior read. score	-0.0530			
	(0.0283)			
CTT x 4th-q prior read. score	-0.0452			
	(0.0264)			
CTT x 5th-q prior read. score	-0.0921**			
	(0.0322)			
CTT x girl		-0.0160		
		(0.0213)		
CTT x black			0.00772	
			(0.0228)	
CTT x ES D				-0.00649
				(0.0318)
CTT x ES E				-0.0173
				(0.0375)
Observations	31,782	31,782	31,782	31,782
R-squared	0.159	0.158	0.158	0.158
Number of cd_escola	3,391	3,391	3,391	3,391

Notes: Robust standard errors clustered by schools in parentheses.
*** p<0.001, ** p<0.01, * p<0.05.
All models controls for school fixed effects, student and teachers covariates.
CTT means contact with trained teacher.
All models compares students that had any contact with trained teacher,
either once or more times, with those who haven't.

3.6 Final remarks

To eliminate child illiteracy, the Literacy Program at the Right Age (LPRA) rely on professional development (PD) as a way to improve teacher pedagogical and subject-matter knowledge. This strategy seemed to boost the learning of children exposed to the program and, based on its model, its relatively high workload, and the extra help provided to teachers from low-achievement cities, we conjecture that it would spillover to children from grades not targeted by the program.

Our empirical strategy consisted of estimating educational production functions that include all relevant variables that explain student performance that were available to us. Unfortunately, we could not deal appropriately with the potential endogeneity of teacher allocation to PD or the student-teacher non-random matching, and we expect an upward-bias on the effect of the training. The data sources are relatively new and lack information about previous non-treated student cohorts in Ceara, precluding the existence of a panel. In the future, we must search the laws about teacher allocation to grades - if teachers are free to choose the grade or if there is some objective criteria.

We utilize data sets provided by the Ceara State Education Department that merge information from the State Evaluation System and School Census. The identification of trained teachers was indirect - a teacher was considered in-training or trained by the LPRA if she taught classes of grades 1 and 2 from 2008 to 2011. Given the existence of federal literacy interventions previous to LPRA, such as PROFA and Pró-Letramento[5], we would like to check if the hours spent in PD raised after 2008 and if there is an association between them and the spillover effects.

Our results indicate the absence of spillovers effects on students not target by LPRA, suggesting that the training indeed focused on literacy instructional methods and the use of the program's materials, and probably not included teaching behaviors that are expected to apply generically to all school subjects. Finally, our findings corroborate the commonly reported result by the literature that observable teacher characteristics do not explain teacher quality.

[5] Programa de Formação de Professores Alfabetizadores and Programa de Formação Continuada de Professores dos Anos/Séries Iniciais do Ensino Fundamental, respectively

Bibliography

AARONSON, D.; BARROW, L.; SANDER, W. Teachers and student achievement in the chicago public high schools. *Journal of Labor Economics*, v. 25, n. 1, p. 95–135, 2007. 15, 24

ABEP, A. B. de Empresas de P. *Critério de Classificação Econômica Brasil - versão 2009*. 2011. Disponível em: <http://www.abep.org/Servicos/Download.aspx?id=04>. Acesso em: 20.03.2019. 59, 99

BARRETTO, E. S. de S. Políticas de formação docente para a educação básica no brasil: embates contemporâneos. *Revista Brasileira de Educação*, v. 20, n. 62, p. 679–701, 2015. 96

BLUNDELL, R.; DIAS, M. C. Alternative approaches to evaluation in empirical microeconomics. *The Journal of Human Resources*, v. 44, n. 3, p. 565–640, 2009. 62

BLUNDELL, R. et al. Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association*, v. 2, n. 4, p. 569–606, 2004. 62

BRAMOULLE, Y.; DJEBBARI, H.; FORTIN, B. Identification of peer effects through social networks. *Journal of Econometrics*, v. 150, n. 1, p. 41–55, 2009. 16, 18, 25, 26, 27, 28, 30

CAIN, K.; OAKHILL, J. Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities*, v. 44, n. 5, p. 431–443, 2011. 53

CEARA, G. do Estado do C. Decreto 29881 de 31 de agosto de 2009. *Diário Oficial do Estado do Ceará*, 2009. 56

COSTA, L. O.; CARNOY, M. The effectiveness of an early-grade literacy intervention on the cognitive achievement of brazilian students. *Educational Evaluation and Policy Analysis*, v. 37, n. 4, p. 567–590, 2015. 54, 95

DESIMONE, L. M. Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, v. 38, n. 3, p. 181–199, 2009. 96, 99

EPPLE, D.; ROMANO, R. E. Peer effects in education: A survey of the theory and evidence. In: ______. *Handbook of Social Economics*. [S.l.]: Elsevier, 2011. v. 1B. 15

GARET, M. S. et al. The impact of two professional development interventions on early reading instruction and achievement. *U.S. Department of Education, NCEE, Washington, DC.*, 2008. 96, 97

GARET, M. S. et al. Middle school mathematics professional development impact study: findings after the first year of implementation. *U.S. Department of Education, NCEE, Washington, DC.*, 2010. 96, 97

GARET, M. S. et al. Middle school mathematics professional development impact study: findings after the second year of implementation. *U.S. Department of Education, NCEE, Washington, DC.*, 2011. 96, 97

GRIMM, K. J. Longitudinal associations between reading and mathematics achievement. *Developmental Neuropsychology*, v. 33, n. 3, p. 410–426, 2008. 55

HANUSHEK, E. A. Conceptual and empirical issues in the estimation of educacional production functions. *The Journal of Human Resources*, v. 14, n. 3, p. 351–388, 1979. 14

HANUSHEK, E. A.; RIVKIN, S. G. Teacher quality. In: ______. *Handbook of the Economics of Education*. [S.l.]: Elsevier, 2006. v. 2. 15

HANUSHEK, E. A.; RIVKIN, S. G. Generalizations about using value-added measures of teacher quality. *American Economic Review: Papers & Proceedings*, v. 100, n. 2, p. 267–271, 2010. 15, 16

HARRIS, D. N.; SASS, T. R. Teacher training, teacher quality and student achievement. *Journal of Public Economics*, v. 95, n. 7-8, p. 798–812, 2011. 16, 96, 97, 104

HECKMAN, J. J.; ICHIMURA, H.; TODD, P. E. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, v. 64, n. 4, p. 605–654, 1997. 63

IBGE. Atlas do censo demográfico 2010. Instituto Brasileiro de Geografia e Estatística (IBGE), Rio de Janeiro - RJ - Brasil, 2013. Disponível em: <https://censo2010.ibge.gov-.br/apps/atlas/pdf/209_213_Glossario_ATLASDEMO\%202010.pdf>. 27

IMBENS, G. W. Matching methods in practice: three examples. *The Journal of Human Resources*, v. 50, n. 2, p. 373–419, 2015. 59, 63

INEP, I. N. de Estudos e P. E. A. T. *Saeb 2005: primeiros resultados: médias de desempenho do Saeb 2005 em perspectiva comparada*. 2007. Disponível em: <http://download.inep.gov.br/educacao_basica/prova_brasil_saeb/menu_do_professor-/resultados/Saeb_resultados95_05_UF.pdf>. Acesso em: 16.01.2015. 79, 80

INEP, I. N. de Estudos e P. E. A. T. *Saeb*. 2019. Disponível em: <http://portal.inep-.gov.br/>. Acesso em: 28.02.2019. 58, 99

JACKSON, K.; BRUEGMANN, E. Teaching students and teaching each other: the importance of peer learning for teachers. *American Economic Journal: Applied Economics*, v. 1, n. 4, p. 85–108, 2009. 15, 16, 17, 18, 28, 44

JACOB, B. A.; LEFGREN, L. The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in chicago. *The Journal of Human Resources*, v. 39, n. 1, p. 50–79, 2004. 96, 98, 104

KASMIRSKI, P. R.; GUSMAO, J.; RIBEIRO, V. M. O paic e a equidade nas escolas de ensino fundamental cearenses. *Estudos em Avaliação Educacional*, v. 28, n. 69, p. 848–872, 2015. 54, 95

KOEDEL, C.; MIHALY, K.; ROCKOFF, J. E. Value-added modeling: A review. *Economics of Education Review*, v. 47, n. ?, p. 180–195, 2015. 15, 23, 24, 25

LAVOR, D. C.; ARRAES, R. de Albuquerque e. *Qualidade da educação básica e uma avaliação de política educacional para o Ceará*. 2014. Disponível em: <http://www2.ipece.ce.gov.br/encontro/2014/trabalhos-/QUALIDADE_DA_EDUCACAO_BASICA_E_UMA_AVALIACAO_DE_POLITICA_EDUCACIONAL_PARA_O_CEARA.pdf>. Acesso em: 02.03.2015. 54, 95

LEE, D. S. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, v. 76, n. 3, p. 1071–1102, 2009. 83

MANSKI, C. F. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, v. 60, n. 3, p. 531–542, 1993. 26, 27

MARKS, G. N. Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. *Oxford Review of Education*, v. 34, n. 1, p. 89–109, 2008. 74

MARQUES, F. C. *Formação continuada de professores no "Programa de Alfabetização na Idade Certa" (PAIC) : peça-chave para o sucesso da política educacional cearense?* Dissertação (Dissertação de Mestrado) — Fundação Getulio Vargas - Escola de Administração de Empresas de São Paulo, 2018. 97, 98, 99

MAS, A.; MORETTI, E. Peers at work. *The American Economic Review*, v. 99, n. 1, p. 112–145, 2009. 14, 15

MCEWAN, P. J. Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, v. 85, n. 3, p. 353–394, 2015. 64

MEC, M. da E. *Pacto Nacional pela Alfabetização na Idade Certa. Interdisciplinaridade no ciclo de alfabetização. Caderno de Apresentação*. 2015. Disponível em: <http://pacto.mec-.gov.br/materiais-listagem/item/download/21_9945a2941359afb9a5bc726869f697c5>. Acesso em: 10.03.2019. 57

NRC, N. R. C. *Preventing Reading Difficulties in Young Children*. [S.l.]: The National Academies Press, 1998. 53

SACERDOTE, B. Peer effects in education: How might they work, how big are they and how much do we know thus far? In: ______. *Handbook of the Economics of Education*. [S.l.]: Elsevier, 2011. v. 3. 15, 16, 27, 44

SAO PAULO. Decreto nº 58154, de 22 de março de 2018. *Diário Oficial da Cidade*, v. 63, n. 54, p. 8–32, 2018. 27

SEDUC-CE, S. de Estado da Educação do C. *Regime de colaboração para a garantia do direito à aprendizagem: o Programa Alfabetização na Idade Certa (PAIC) no Ceará*. [S.l.]: Secretaria Estadual de Educação, 2012. 55, 56, 57, 96, 98

SEDUC-CE, S. de Estado da Educação do C. *Ações desenvolvidas (2007/2010)*. 2015. Disponível em: <http://www.paic.seduc.ce.gov.br/images/PASTA_TATIANA-/acoes_paic_2007_a_2011.pdf>. Acesso em: 02.03.2015. 56, 57, 98

SEDUC-CE, S. de Estado da Educação do C. *História do MAIS PAIC*. 2019. Disponível em: <http://www.paic.seduc.ce.gov.br/index.php/o-paic/objetivos-e-competencia>. Acesso em: 23.02.2019. 55

SEGATTO, C. I. Análise da implementação de políticas públicas: o programa de alfabetização na idade certa em dois municípios cearenses. *Temas de Administração Pública*, v. 4, n. 7, p. 1–16, 2012. 54

SLAVIN, R. E. et al. Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*, v. 79, n. 4, p. 1391–1466, 2009. 52, 53, 54, 56, 64, 96

SOARES, M. Letramento e alfabetização: as muitas facetas. *Revista Brasileira de Educação*, s/v, n. 25, p. 5–17, 2004. 52

STANOVICH, K. E. Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education*, v. 189, n. 1/2, p. 23–55, 2009. 53

TODD, P. E.; WOLPIN, K. I. On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, v. 113, n. 485, p. F3–F33, 2003. 23

TPE, T. P. E. *Nota técnica preliminar. Metodologia para a obtenção das metas finais e parciais.* 2007. 66, 69

UNESCO. *Literacy for life. EFA Global Monitoring Report 2006.* [S.l.]: UNESCO, 2005. 52

WINTERS, M. A.; DIXON, B. L.; GREENE, J. P. Observed characteristics and teacher quality: Impacts of sample selection on a value added model. *Economics of Education Review*, v. 31, n. 1, p. 19–32, 2012. 15