Educational Evaluation and Policy Analysis
Month 201X, Vol. XX, No. X, pp. 1–24
DOI: 10.3102/0162373715571437

The Effectiveness of an Early-Grade Literacy Intervention on the Cognitive Achievement of Brazilian Students

Leandro Oliveira Costa
World Bank
Martin Carnoy
Stanford University

Beginning in 2007, the Literacy Program at the Right Age (Pacto pela Alfabetização na Idade Certa [PAIC]) in Brazil's Ceará state required municipal schools to implement a tiered, whole-school early-grade literacy intervention. This intervention was complemented by other policies to help municipalities improve student achievement. The present study identifies PAIC's impact using a triple-difference analysis (Difference-in-Difference-in-Difference [DDD] approach) that relies on comparisons of the test-score changes in Ceará and bordering states and for grades and schools treated and not treated by the literacy program since 2007. The results show that PAIC had a positive effect on student achievement in both Portuguese and mathematics, but that it did not help reduce the gap between students who had or had not participated in early childhood education.

Keywords: *literacy, early childhood, achievement gains*

In 2012, the Brazilian Ministry of Education launched the National Pact of Literacy at the Right Age (Pacto Nacional pela Alfabetização na Idade Certa [PNAIC]). The PNAIC's goal is ambitious. It seeks to improve the literacy skills of 6- to 8-year-olds in public schools throughout Brazil. The government's decision to implement this literacy intervention nationally in the first three grades of primary school was influenced by the apparent success of a similar 3-year program—Literacy Program at the Right Age (Pacto pela Alfabetização na Idade Certa [PAIC])—first implemented in 2007 in one (low-income) northeastern Brazilian state, Ceará. Ceará's illiteracy rate among 7- to 14-year-olds declined from 18% in 2001 to 14% in 2007, similarly to the rest of the Northeast, in a period of increasing educational attainment for this age group. However, after introducing the PAIC, the literacy rate in Ceará fell to 6% by 2011, a significantly larger drop than in other northeastern states (Instituto Brasileiro de Geografía e Estadística, 2001, 2007, 2011).

In addition to more rapid declines in illiteracy rates after 2007, primary school–age students in Ceará made significant gains in reading and mathematics achievement compared with other states, as measured by Brazil's national test, the *Prova Brasil*. Yet, no rigorous impact evaluation has been done to assess whether students' test-score gains (and, by implication, increases in literacy rates) in Ceará were directly due to the PAIC or rather due to other educational conditions specific to Ceará in the period, 2007 to 2011.

PAIC resembles a widely implemented U.S. early reading intervention, Reading First. Based on the U.S. national evaluation of Reading First, we would not expect PAIC effects on student outcomes to be significant. Reading First had

a considerable impact on how teachers teach reading and the implementation of more effective literacy curricula, small positive effects in first grade, but insignificant effects on overall students' achievement scores (Gamse et al., 2008). Nevertheless, initial achievement levels are much lower in Brazil than in the United States. In a developing country context, early literacy interventions such as PAIC could have had much larger effects and positive implications for educational improvement.

In this article, we test PAIC's causal effect on student achievement in Ceará. We use the extensive databases from *Prova Brasil* (Instituto Nacional de Estudos e Pesquisas [INEP], 2008, 2010, 2012) plus the fact that the same-age students in bordering states' schools and older students in the same schools were not exposed to the PAIC program. We analyze whether the improvement in reading and mathematics for students who attended Ceará's PAIC program was mainly the result of PAIC or of other, unobserved factors affecting all early-grade students in the region (and bordering states) or all students in Ceará's municipal schools. Finally, we assess whether the PAIC has reduced the gap between students with and without early childhood education and whether the program affects student groups with lower and higher levels of achievement equally.[1] Our research has implications for the potential success of the national program (PNAIC) and represents an important contribution to the empirical literature on large-scale early-grade literacy interventions.

The PAIC was based on cooperation between the office of the state's Secretary of Education (Secretaria da Educação [SEDUC]) and the state's municipal governments, the administrative entities most responsible for pre-school and early primary education. It was designed to help achieve five interrelated program goals in the early grades of Ceará's municipal schools: (a) reading promotion, (b) supporting municipal literacy strategies, (c) strengthening municipal management, (d) supporting early childhood education, and (e) providing external learning evaluation. In terms of implementation, the program was divided into three broad activities related to three program areas: (a) teacher and administrators training, (b) municipality network mobilization, and (c) provision of books to students.

PAIC resembles Reading First in its focus on new literacy teaching materials and on teacher training to use the materials. However, there are important differences between the two programs. In Reading First, coupling between the designers of the intervention and its implementers is loose. The federal government provides funds to states, which in turn distribute those funds to districts to implement district-developed literacy improvement plans that include materials and teacher training of districts' choosing. In contrast, PAIC is administered directly by Ceará's state education ministry, which both developed the literacy materials and organized the professional development workshops for municipal teachers to assist them in applying the materials in their classrooms.

Thus, in PAIC, the state's SEDUC distributes specific teaching materials and a teaching guide directly to municipalities to be distributed to teachers in schools and provides face-to-face training to early-grade teachers. PAIC training takes place 3 to 5 times per year in 20 regional centers throughout the state and 3 more centers in the state capital, Fortaleza, run by external consultants working with a team from the state's SEDUC. Municipalities send teachers in groups to these centers to be trained to apply the model. The training is a good example of a well-designed, scripted literacy and numeracy training program initially developed in a single Ceará municipality (Sobral) and successfully scaled up to hundreds of municipalities in Ceará state.

This training is combined with *in loco* visits to evaluate the progress of the implementation of the scripted material. The team from the state's SEDUC also plays an active role in promoting the importance of the PAIC program in each municipality among key municipal actors and creating an environment in which schools share best practices for improving measured literacy outcomes ("network mobilization").[2]

In addition, unlike Reading First or other early reading interventions in other countries, PAIC adds financial incentives for schools and municipalities whose third-grader students achieve the largest literacy gains in external literacy assessments. The state rewards schools whose third graders rank highest in the percent achieving an adequate level of literacy skills (as defined by the state) on a jointly run state and municipal annual

literacy assessment given all students in the first, second, and third grades of public schools. Implemented along with the PAIC in 2007, the assessment is called the Sistema Permanente de Avaliação da Educação Basicá do Ceará–Alfa (SPAECE-Alfa). The monetary reward is defined in per student terms, but the total reward to the school is based on the number of third graders participating in the SPAECE-Alfa.[3] The state also rewards municipalities whose students achieve the highest levels and gains on a state literacy index (for details on these incentives, see Gusmão & Ribeiro, 2011).[4]

The materials distributed by the state's SEDUC were designed for improving the teaching of literacy, but the teacher training sessions focus not only on using the specific materials but also on better instructional practice more generally, which likely spills over into the teaching of mathematics. Furthermore, the incentive system rewards higher test scores in both literacy and mathematics.

The outcome measures we use in our analysis are students' mathematics and Portuguese scores on the *Prova Brasil*. From 2005 onward, the *Prova Brasil* has been administered every other year to all students in the last year of public primary and middle schools.[5] The test assesses mathematics and Portuguese language and also surveys students, teachers, and principals with extensive questionnaires. As the *Prova Brasil* is only administered to students at the end of primary and middle school, the 2011 *Prova Brasil* was the first to assess the cohort starting primary school at the age of 6 in 2007. Furthermore, since the PAIC began to be implemented in 2007 and was fully in force by 2008 in the first three grades of Ceará's municipal public schools, the students in the fifth grade who took the *Prova Brasil* in 2011 represent *the first cohort* participating in at least 2 years of the PAIC program, and for many, the full 3 years. This cohort is our treatment group.

We include three "control" groups in our study: fifth graders in Ceará who took the *Prova Brasil* in 2007 (they were not exposed to the PAIC), fifth graders in neighboring states who took the *Prova Brasil* in both 2007 and 2011 (neither group participated in the PAIC), and ninth graders who took the *Prova Brasil* in 2007 and 2011, who were not exposed to the PAIC, but whose achievement may have been affected by other Ceará reforms also influencing fifth graders' achievement.[6] We use the presence/absence of PAIC implementation between years, states, and grades to compare the 2011 fifth-grade student cohort in Ceará state with the untreated fifth graders in 2007 and in states bordering Ceará as well as the untreated ninth graders in Ceará state (Figure 1). We argue that this quasi-experimental research design allows us to estimate the effect of the literacy program on gains in students' cognitive achievement, accounting for the influence of state-level constant unobservable factors and within-state programs that could have affected the students during the same period.

Our results indicate that PAIC produced statistically significant gains in the average Portuguese performance of (treated) fifth graders in 2007 to 2011 (compared with our untreated control groups) and greater gains in their mathematics performance. Increases were estimated to be larger for students with higher test scores, but differed little for those students who had had or not had early childhood education.

This article is structured as follows: In the following section, we briefly review the relevant empirical literature on literacy interventions and compare them with PAIC. Section "Data" describes the data. Section "Causal Inference Strategy" outlines our empirical strategy and Section "Results" shows the results. Section "Robustness Checks" performs some robustness checks. Finally, Section "Conclusion" summarizes the results and concludes.

The Impact of Literacy Interventions

The argument that early educational interventions have high payoffs in reducing large achievement gaps at school entry between socially advantaged and disadvantaged children has led to increasing emphasis on early intervention programs (see, for example, Barnett, 2002; Cunha & Heckman, 2010; Currie & Thomas, 1993; Heckman & Masterov, 2007; Jencks & Phillips, 1998).

Much of the policy focus of early interventions has been on early childhood educational development (ECD) because of studies showing cognitive and non-cognitive gains by disadvantaged children exposed to pre-school programs, producing

					Prova Brasil Wave	***Prova Brasil Wave***				***Prova Brasil Wave***
Year	*Ceará 1st grade*	*Ceará 2nd grade*	*Ceará 3rd grade*	*Ceará 4th grade*	*Ceará 5th grade*	*Bordering States 5th grade*	*Ceará 6th grade*	*Ceará 7th grade*	*Ceará 8th grade*	*Ceará 9th grade*
2006										
2007					**Control 1**	**Control 2**				**Control 3**
2008										
2009					One Year					
2010										
2011					**Treated**	**Control 2**				**Control 3**
2012										

FIGURE 1. *Schematic of DDD design for testing PAIC effect on Ceará students' achievement scores, treatment group, and control groups.*
Source. Authors.
Note. Orange: 3 years of participation in PAIC; Red: 2 years of participation; Pink: 1 year of participation; Blue: Control groups. Fifth graders in *Prova Brasil* 2007 were not exposed to the PAIC program; fifth graders in *Prova Brasil* 2009 were exposed up to 1 year to PAIC; and fifth graders in *Prova Brasil* 2009 were exposed to 2 to 3 years of PAIC (full treatment). DDD = Difference-in-Difference-in-Difference. PAIC = Pacto pela Alfabetização na Idade Certa.

high economic and social payoffs in adulthood (Cunha & Heckman, 2010). In the same spirit of trying to reduce the achievement gap existing at school entry, there have also been large-scale literacy interventions in kindergarten and the primary grades. Programs with distinct designs have been implemented in various countries. The most important of these are Reading First (written in the U.S. No Child Left Behind [NCLB] Act) and Reading Recovery (New Zealand, the United Kingdom, and the United States).

Reading First is the intervention most closely related to the PAIC. It is a multi-layer intervention in kindergarten through third grade applied in school districts in a number of U.S. states as part of the 2001 NCLB, first implemented in 2003 and funded by the U.S. federal government. Funding can be used for special literacy curricula and materials, for professional development and coaching for teachers on how to use these materials and how to work with struggling readers, and for monitoring reading difficulties of struggling readers. Its main goal was to have all children read at or above grade level by the end of third grade (Gamse et al., 2008).

Early tutoring programs are an alternative to early in-school interventions such as Reading First and PAIC. Reading Recovery is the most well known of these programs. It was developed in the 1970s as a short-term one-to-one supplemental tutoring intervention designed for children aged 5 or 6, who are struggling to be literate after their first year of school (see Ashdown & Simic, 2000; Lyons, 1998; Schwartz, 2005). It purports to increase literacy instruction time as a method to improve pupils' reading in the early grades.

There have also been a number of small-scale literacy teaching interventions in Africa (see, for example, Ralaingita & Wetterberg, 2011) and a large-scale effort to promote such early reading interventions through the introduction in developing countries of a cheap, simple-to-use literacy assessment instrument—the Early-Grade Reading Assessment (EGRA). Although it is a policy intervention directed at improving early literacy, EGRA differs qualitatively from curricular or teacher training interventions. It is based on the premise that reducing the complexity and cost of accurately assessing (low) reading skills in early grades will induce governments to introduce reading improvement programs and conduct impact evaluations (Gove & Wetterberg, 2011).

Table 1 provides a brief comparison of such other early-grade literacy intervention programs with the PAIC.

Evaluations of Reading First and Reading Recovery show mixed results. The mandated national evaluation of Reading First (Gamse et al., 2008) used a regression discontinuity design and found that at national level, the program had a significant positive impact on instructional time spent on the five key elements of the program and significantly increased the multiple practices promoted by the program (including professional development), but it did not have a significant impact on average student reading comprehension scores in Grades 1, 2, or 3.[7] However, Baker et al. (2011) argued that national studies of Reading First could not be generalized to individual states, and found that in Oregon, student's reading outcomes improve over time with high-quality reforms and strong implementation of the program. Their identification strategy was an outcome comparison of two implementation cohorts through a test for difference of aggregated student data at the school level.

The evaluations of Reading Recovery (Ashdown & Simic, 2000; Lyons, 1998; Schwartz, 2005) and one-to-one tutoring programs more generally (Elbaum, Vaughn, Hughes, & Moody, 2000) suggest mixed, although generally positive results. Slavin, Lake, Davis, and Madden's (2011) more recent survey of 97 studies that used randomized or well-matched groups concludes that one-to-one tutoring, including Reading Recovery, is very effective in improving reading performance, and that teachers are more effective as tutors than paraprofessionals and volunteers. Furthermore, another recent evaluation using random assignment to estimate the impact of a major national scaling-up (in 2010) of Reading Recovery to almost 90,000 students also showed large statistically significant improvements in reading scores (May et al., 2013).

Overall, then, early reading interventions in the United States show some positive effects on students' literacy, but these improvements are hardly consistent, and the intervention program most directly similar to PAIC—Reading First—shows insignificant average reading gains at a national level, but more positive results in a

TABLE 1
Comparison of Early Literacy Program Designs, by Goals, Target Group, Areas for Intervention, Educational Activities, and Complementary Policies

Variable	PAIC in Ceará	PNAIC in Brazil	Reading First in the United States	Reading Recovery in New Zealand, the United Kingdom, and the United States	EGRA in Africa, LAC, and Asia
Main goals/strategies	All students to read at 7 years old	All students to read at 8 years old	All students to read at grade level by the end of third grade	Literacy of at-risk children	Educational policymakers/assess students' reading skills
Areas of intervention					
Promotion of reading in early grades	X	X	X	X	X
Support of local strategies to implement literacy program	X		X	X	X
Strengthening of municipal/district management	X	X			
ECD support	X	X			
External evaluation	X		X	X	X
Educational activities					
Teacher and administrator training	X	X	X	X	X
Teacher coaching			X		
Network mobilization among municipalities/ districts	X	X			X
Use of literacy materials developed for program	X	X	X	X	
One-to-one supplemental tutoring				X	
Complementary policies					
Financial reward program	X				
Financial support policy	X				

Source. PAIC: Gusmão and Ribeiro (2011); PNAIC: http://pacto.mec.gov.br/o-pacto; Reading First: Gamse et al. (2008); Reading Recovery: Lyons (1998); EGRA: Gove and Wetterberg (2011).
Note. PAIC = Pacto pela Alfabetização na Idade Certa; PNAIC = Pacto Nacional pela Alfabetização na Idade Certa; EGRA = Early-Grade Reading Assessment; LAC = Latin America and Caribbean; ECD = early childhood educational development.

single state study. Against this backdrop, we turn to estimating the effects of the PAIC, a program implemented in one Brazilian state.

Data

The data for this study are drawn from Brazil's extensive educational evaluation system. The National Evaluation System of Basic Education (Sistema de Avaliação da Educação Básica [SAEB]) is composed of two complementary surveys that aim to evaluate the quality of Brazil's educational system. SAEB's main survey from 1995 to 2005 involved a biannual state-level student *sample*, in both the public and private school systems, of schools located in rural and urban

areas, covering the fourth and eighth grades of elementary school (as noted, in 2011, the test was applied to fifth and eighth graders, and in 2015, will test fifth and ninth graders) and the third year of high school. The SAEB has continued to be administered biannually after 2005, although only the 2011 sample became publicly available.

The second survey is the Avaliação Nacional do Rendimento Escolar (ANRESC)—the National Assessment of Educational Achievement—which is publicly known as *Prova Brasil*. The *Prova Brasil* is also a biannual school-level assessment of fifth and ninth graders. It differs from SAEB in that it is administered to *all* students (not a sample of schools) in *public* schools with a minimum of 20 students per class. It has been administered since 2005, but it is available as student-level microdata only for 2007, 2009, and 2011. For these three waves, we are therefore able to conduct our analysis at the individual student level within each test year. This rich data set provides information about students' mathematics and reading (Portuguese) exam scores and the socio-economic background of students, teachers, and principals. The exam scores in each grade are comparable over time and permit comparison of results between grades. That is, we can follow the performance of grade cohorts, schools, networks, and the system as a whole (but not of individual students, as we do not have matchable student identifiers across grades and years). In 2011, this data set provided information on 5,201,730 students enrolled in 55,924 schools in 27 states, and in 2007, for 4,109,265 students and 48,704 schools.

In the 2011 *Prova Brasil*, 61.8% of the schools surveyed were administered by municipalities, 38.2% by states, and 0.1% by the federal government. Students attended schools in three shifts—55.5% in the morning shift, 42.4% in the afternoon, and 2.1% at night. To be able to include most of the schools participating in Ceará's PAIC program, and to compare fifth and ninth graders in these schools within Ceará and fifth graders in similar schools between Ceará and bordering states, we selected only urban municipal schools in the morning and afternoon shifts. Also, we restricted our sample to schools with fifth and ninth grades that participated in the *Prova Brasil* both in 2007 and in 2011, in which we could identify the teacher's classroom subject. This resulted in a school-level panel of student and teacher background variables with a sample size of 1,002 schools and 275,072 individual students in five states (Ceará, Piaui, Pernambuco, Paraiba, and Rio Grande do Norte).[8]

Table 2 presents the means of our outcome variables and important covariates in this sample. The Portuguese and math scores on the *Prova Brasil* exam are our outcome variables. The scale of this variable ranges from 0 to 500 in both the fifth and ninth grades. The concepts covered in the *Prova Brasil* test items are based on the 2001 revised version of a reference matrix of the Brazilian curriculum. This matrix was created through a broad national analysis of state-level curricula and textbooks used by primary and secondary teachers in Brazilian schools. Based on these curricula and textbooks, test designers constructed a set of competencies and skills that students should be able to execute at the end of the two basic education cycles (fifth and ninth grades) and used item response theory (IRT) to design the test at each level. With IRT, we can compare the results of fifth graders across years and ninth graders across years.[9]

We use a series of student covariates to control for characteristics that could affect student achievement. We also control for teacher and school characteristics that may be related to the quality of education, such as teachers' education and wages, and class size. Student attendance at an ECD center, day care, or kindergarten reflects early childhood education that could also influence a student's later school achievement. We created an ECD variable based on a positive answer to either of two questions in *Prova Brasil*'s student questionnaire that ask whether the student had attended a day care center and whether the student had attended a pre-school.[10] In Table 2, we report the average percentage of students in the fifth and ninth grades with ECD.

We chose to compare students in Ceará with students in the bordering northeastern neighbors not only because of their geographic proximity but also because they are socio-economically similar and students there and in Ceará have had similar average scores on the SAEB and *Prova Brasil* since 1999. In contrast, students in the south and southeast regions averaged scores more than 10% higher than students in the Northeast.

TABLE 2
Ceará and Bordering States: Summary Statistics for Student, Teacher, and School Variables, by Grade, 2007 and 2011 Waves of Prova Brasil

	2007 fifth grade		2011 fifth grade		2007 ninth grade		2011 ninth grade	
	Ceará	Bordering	Ceará	Bordering	Ceará	Bordering	Ceará	Bordering
Variable	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Portuguese score	157.30 (37.22)	156.77 (36.49)	181.54 (43.63)	168.96 (40.25)	216.90 (40.88)	216.31 (41.33)	233.42 (43.95)	226.65 (43.26)
Math score	172.69 (37.33)	174.37 (37.43)	197.58 (43.95)	185.83 (40.35)	223.59 (38.83)	224.86 (39.43)	236.15 (46.18)	232.35 (44.37)
Black	0.09 (0.29)	0.11 (0.31)	0.07 (0.26)	0.09 (0.28)	0.08 (0.28)	0.09 (0.28)	0.07 (0.26)	0.08 (0.27)
Mother no primary	0.31 (0.46)	0.36 (0.48)	0.16 (0.37)	0.19 (0.39)	0.48 (0.50)	0.50 (0.50)	0.23 (0.42)	0.26 (0.44)
Mother degree	0.07 (0.25)	0.06 (0.24)	0.09 (0.28)	0.08 (0.27)	0.03 (0.18)	0.04 (0.21)	0.14 (0.34)	0.15 (0.36)
Father no primary	0.24 (0.43)	0.29 (0.46)	0.17 (0.37)	0.20 (0.40)	0.43 (0.49)	0.46 (0.50)	0.25 (0.43)	0.28 (0.45)
Father degree	0.06 (0.24)	0.06 (0.24)	0.07 (0.25)	0.06 (0.23)	0.02 (0.14)	0.02 (0.15)	0.09 (0.29)	0.10 (0.30)
Teacher no degree	0.12 (0.33)	0.20 (0.40)	0.08 (0.27)	0.14 (0.34)	0.04 (0.20)	0.03 (0.16)	0.03 (0.16)	0.03 (0.16)
Teacher low wage	0.39 (0.49)	0.38 (0.48)	0.16 (0.37)	0.12 (0.32)	0.35 (0.48)	0.19 (0.40)	0.20 (0.40)	0.11 (0.32)
Teacher high wage	0.07 (0.26)	0.03 (0.17)	0.09 (0.29)	0.03 (0.17)	0.05 (0.23)	0.03 (0.18)	0.07 (0.25)	0.05 (0.23)
Class size	30.95 (6.96)	31.94 (8.73)	29.25 (5.58)	29.91 (6.61)	31.36 (8.51)	33.63 (9.46)	31.34 (6.93)	32.50 (8.59)
ECD (% in school)	0.64 (0.48)	0.66 (0.48)	0.77 (0.42)	0.72 (0.45)	0.80 (0.40)	0.76 (0.43)	0.84 (0.36)	0.76 (0.42)
n	27,321	30,306	25,473	28,422	15,673	19,838	18,801	22,101

Source. Prova Brasil, 2007 and 2011, microdata.
Note. ECD = early childhood educational development.

The means presented in Table 2 show this similarity between Ceará and the comparison border states of pretreatment fifth-grade test scores, student characteristics, class size, percentage of students who attended ECD, and most teacher characteristics. The percentage of Black and female students, parents' education, class size, and ECD attendance do not differ statistically between Ceará and its border states. The one exception is that teachers' wages in Ceará's ninth grade appear to be significantly lower than teachers' wages in neighboring states.

Causal Inference Strategy

PAIC was implemented in many of Ceará's municipal public schools in 2007 and reached all schools by 2008. As the *Prova Brasil* is only given to fifth and ninth graders every 2 years (odd years), the 2011 fifth-grade cohort is the first to be both exposed to the literacy program (either 2 or 3 years) and assessed by the *Prova Brasil* exam. The exam results for this cohort allow us to evaluate the program's impact (achievement gains from 2007 to 2011) based on a comparison with the 2007 to 2011 gains for the

same cohort in bordering states and the 2007 to 2011 gains for the ninth-grade cohort in Ceará (see Figure 1).

To estimate the effect of PAIC on students' achievement, we develop a model using state-level and school-level panel data from *Prova Brasil*. The identification strategy we use is the Difference-in-Difference-in-Difference (DDD) estimation. Good examples of the use of this methodology are Tyler, Taylor, Kane, and Wooten (2010); Ravallion, Galasso, Lazo, and Philipp (2005); Chaudhury and Parajuli (2010); and Chiapa, Garrido, and Prina (2012). Angrist and Pischke (2008) pointed out that this model is a modification of the Difference-in-Difference (DD) model with possible higher order control groups that contribute to the identification strategy of programs with more than one dimension of external variation.

To identify the impact of the PAIC on test scores, consider the DD estimates for the fifth grade of Ceará and bordering states (Ceará fifth graders vs. bordering states' fifth graders), and the DD estimates for fifth- and ninth-grade students in Ceará state (Ceará fifth-grade students' gains vs. ninth-grade students' gains). The first DD model (DD_1) seeks to estimate the effect of the PAIC program for the fifth-grade students in Ceará compared with students in bordering states that have not implemented any early-grade literacy intervention. The second DD model (DD_2) aims to control for any other program in Ceará state that could have affected fifth-grade students' achievement since 2007. The DDD model is the difference of these two models. It controls for across state school differences and for between-grade differences. We need state-level panel data of *Prova Brasil* for fifth and ninth grades in Ceará state and border states. To set up this model, consider the following conditional expectation function (CEF):

$$E\left(y_{istg}|s,t,g,D,X_{it}\right)=\alpha_s+\beta_t+\delta_g+\mu_{st}+\gamma_{gt}+\theta_{sg}+\rho_{stg}+X_{istg}'\beta, \quad (1)$$

where *s* represents states, *t* represents year, and *g* represents the grade of the student in a municipal public school. Angrist and Pischke (2008) argued that this model provides full non-parametric control for state-specific time effects that are common across grade groups (μ_{st}), time-varying grade effects (γ_{gt}), and state-specific grade effects (θ_{sg}). Also, we control for state-level fixed effects (α_s), a time trend (β_t), and a dummy for the grade effect (δ_g).

To estimate the interaction of the PAIC program impact and whether a student participated in ECD, we stratify our data and results into students who attended ECD centers or not. We would like to fit Equation 1 separately for each of these two different literacy background groups. However, we cannot fit the hypothesized DDD model in Equation 1 because we do not have student identifiers who allow us to construct longitudinal student-level data on students' test scores. Fortunately, based on Equation 1, the parameter of interest, ρ_{sgt}, representing the effect of the PAIC on the achievement of students who attended fifth grade in 2011 as affected by state of residence is as follows:

$$\begin{aligned}\mathrm{DDD}=\mathrm{DD}_1-\mathrm{DD}_2&=\left(\Delta y_{\mathrm{CE},5g}-\Delta y_{\mathrm{BO},5g}\right)\\&\quad-\left(\Delta y_{\mathrm{CE},9g}-\Delta y_{\mathrm{BO},9g}\right)\\&=\rho_{stg}.\end{aligned} \quad (2)$$

In essence, Equation 2 permits us to identify the impact of the PAIC program on the achievement of fifth graders in Ceará ($y_{\mathrm{CE},5g}$) based on variation across states ($y_{\mathrm{BO},5g}$) and between grades ($y_{\mathrm{CE},9g}$) in two periods. The Δ represents the difference in time, so each component with Δ differences out any time fixed effect of each cohort in Ceará and its bordering states. Where the DD_1 differences net out any constant effect for fifth graders' achievement gains between Ceará and bordering states, the DD_2 differences net out ninth graders' constant effects. The DDD differences net out the fixed effect between the fifth- and ninth-grade cohorts. The triple differencing strategy allows us to account for the effect on achievement gains for fifth graders in Ceará of being exposed to the PAIC compared with untreated students in bordering states, but to net possible effects on test-score gains of unobserved, concurrent, efforts to improve education across all grades.

This can be estimated by using the following regression framework:

$$\begin{aligned}y_{it}&=\mu_s+\beta\,2011_{it}+\delta\,5\text{th grade}_{it}+\alpha\,\text{Ceará}_{it}\cdot 2011_{it}\\&\quad+\gamma\,5\text{th grade}_{it}\cdot 2011_{it}+\theta\text{Ceará}_{it}\cdot 5\text{th grade}_{it}\\&\quad+\rho\text{Ceará}_{it}\cdot 5\text{th grade}_{it}\cdot 2011_{it}+\varepsilon_{it},\end{aligned} \quad (3)$$

where 2011 is a dummy variable for the 2011 year of the *Prova Brasil*; fifth grade is a dummy variables for the fifth-grade students in 2007 and 2011 (within-state comparison); and Ceará is a dummy for Ceará state students differentiated from students in border states (cross-state comparison). The DDD effect is measured by estimating the ρ coefficient, which is obtained by the interaction of the fifth-grade dummy, the 2011 year dummy, and the Ceará state dummy. Finally, μ_s is the state-level fixed effect (the detailed mathematical explanation on how differences identify this effect is available from the authors).

A usual challenge raised to estimating the effect of a program in a classical regression using multi-level data is that the estimate of the standard errors is biased. For example, the likely heteroskedasticity yielded by the school clusters of students' achievement affects the classical standard error estimate. Furthermore, when we work with information in a panel format we can expect a significant serial correlation. Following the strategy suggested in Angrist and Pischke's (2008) study, we considered the school clusters to estimate the standard errors due to presumed heteroskedasticity and serial correlation.

In addition to estimating the overall PAIC effect on student achievement in Ceará, we estimate the DDD model using quantile regressions to provide information about the relationship between students' achievement y_{it} and covariates at different points in the conditional distribution of y_{it}. We use the sample median as an estimator of the population median. If $F(y) = \Pr(Y \leq y)$ defines the cumulative distribution function (CDF), then $F(y_{\text{med}}) = 1/2$ is the equation whose solution defines the median $y_{\text{med}} = F^{-1}(1/2)$. The quantile q, $q \in (0,1)$, is defined as the value of y that splits the data into the proportions q below and $1 - q$ above, that is, $F(y_q) = q$ and $y_q = F^{-1}(q)$. These concepts extend to the conditional quantile regression function, denoted as $Q_q(y \mid x)$, where the conditional quantile will be taken to be linear in x.

Furthermore, the PAIC effect may differ for schools with a smaller or larger achievement gap between students in a school who attended ECD or not. We developed a model that estimates the interaction between the literacy program (PAIC) and the ECD achievement gap of students in the same school and grade. This school ECD achievement gap can be defined as the expected difference in achievement of students at the same school who have attended or not attended some form of ECD.

We can represent the school ECD gap as $E(y_{j,ECD} - y_{j,noECD} | \mathbf{s}, t, g, \mathbf{e}, D, X_{it})$, where **s** is a vector of unobservable but fixed school confounders, such as parents' motivation and ability to enroll their children in ECD centers, and **e** is the vector of observed previous ECD enrollment. We assume that the school ECD achievement gap is a characteristic correlated with parent's school choice, based on the assumption that the unobserved variable related to attending ECD centers is likely the same as attending school, and that ECD attendance rates differ due to the supply of ECD centers in the regions. Also, because both the demand for and the supply of ECD have changed, we expect that the ECD gap varies across grades.

Thus, the main assumption to identify the interaction of the program impact and the school ECD achievement gap is the inclusion of school fixed effects that can control for unobserved variables related to the selection bias of ECD gap in each school; that is, $E(y_{j,ECD} - y_{j,noECD} | s, t, g, e, D, X_{it}) = \tau_j$. This assumption combined with the external variation across grades and schools yields the interaction between the individual student's ECD and participation in the PAIC program, controlling for school fixed effects. This means a modification of DDD that includes an additional difference, this one at the school level. We call this the "four difference" method (4D). The following equation represents this model:

$$\begin{aligned} E\left(y_{ijstge} | t, s, g, e, D, X_{it}\right) = {} & \tau_j + \mu_s + \beta_t + \delta_g \\ & + \alpha_{ste} + \gamma_{gte} + \theta_{sge} \\ & + \rho_{tsg} + \pi_{tsge} + X_{itsge}'\beta, \end{aligned} \tag{4}$$

where τ_j is the school fixed effect of school j and the π_{tsge} is the interaction of attending ECD and the fifth grade in Ceará state in 2011 (ECD and treatment variable).

We model the second difference of students' ECD status and time difference as $\Delta^2_{j\text{ECD}} = (\Delta_{\text{ECD}} y_{jt} - \Delta_{\text{ECD}} y_{jt-1}) = (y_{j\text{ECD}(t)} - y_{j\text{noECD}(t)}) - (y_{j\text{ECD}(t-1)} - y_{j\text{noECD}(t-1)})$. ECD($t$) represents the students with ECD status in time t and no ECD(t) represents the students with no ECD status in time t in the same school j. Note that the subscript j is in the school-level fixed effect, τ_j, and the subscript i

is in the dependent variable and the error variable. This means that we intend to examine the achievement gap of ECD inside schools (the difference in the achievement gain of students with and without ECD in each school) and how it interacts with the PAIC effect (interested readers can contact the authors for the derivation of the model). The new difference method to obtain the interaction between the school ECD gap and participation in the PAIC program at the state level (difference-in difference-in-difference-in difference [DDDD]) is as follows:

$$\begin{aligned} 4\mathrm{D} = \mathrm{DD}_1 - \mathrm{DD}_2 &= \left(\Delta^2_{je} y_{\mathrm{CE},5g} - \Delta^2_{je} y_{\mathrm{BO},5g}\right) \\ &\quad - \left(\Delta^2_{je} y_{\mathrm{CE},9g} - \Delta^2_{je} y_{\mathrm{BO},9g}\right) \\ &= \rho_{stge}. \end{aligned} \tag{5}$$

This can be estimated by using another convenient regression framework:

$$\begin{aligned} y_{it} &= \tau_j + \mu_s + \beta\, 2011_{it} + \delta\, 5\mathrm{thgrade}_{it} \\ &\quad + \alpha\ \mathrm{Ceará}_{it} \cdot 2011_{it} \cdot \mathrm{ecd}_{it} + \gamma\, 5\mathrm{thgrade}_{it} \cdot \\ &\quad 2011_{it} \cdot \mathrm{ecd}_{it} + \theta\, \mathrm{Ceará}_{it} \cdot 5\mathrm{thgrade}_{it} \cdot \mathrm{ecd}_{it} \\ &\quad + \rho\ 2011_{it} \cdot \mathrm{Ceará}_{it} \cdot 5\mathrm{thgrade}_{it} + \pi\, 2011_{it} \cdot \\ &\quad \mathrm{Ceará}_{it} \cdot 5\mathrm{thgrade}_{it} \cdot \mathrm{ecd}_{it} + \varepsilon_{it}. \end{aligned} \tag{6}$$

An important assumption of this estimation strategy is that the proportion of ECD students is constant across years and the ECD variable is homogeneous among all students. This assumption may not hold, and, if it does not, selection bias in access to ECD and the amount and quality of ECD result in an overestimate of PAIC's effect on student achievement in Ceará. Nevertheless, this estimation strategy provides us with an upper bound estimate of the program impact controlling for the proportion of ECD students.

Finally, we run a series of robustness checks on the DDD model that test whether the PAIC effect is affected by (a) enlarging the sample by not restricting it to schools whose fifth and ninth grades were both tested by *Prova Brasil* in 2007 and 2011 (Table 3), (b) enlarging the comparison group of reference states, and (c) estimating a "1-year" PAIC effect on the cohort of Ceará fifth graders who were in the third grade in 2007 and were tested as fifth graders in the 2009 wave of the *Prova Brasil* (difference of 2009 fifth graders with 2007 fifth graders).[11]

There may also be further threats to our identification strategy (Bertrand, Duflo, & Mullainathan, 2004) that we cannot correct for because we lack available data. For example, if there were significant numbers of Ceará students in the fifth grade in 2011 who had not attended 2 to 3 years of the PAIC program implementation because they had migrated or had failed to pass into fifth grade in 2010, this could bias our results.

Results

This section presents the results of regression estimates based on the various empirical strategies discussed above. However, before turning to those results, it is worth examining the overall pattern of SAEB and *Prova Brasil* public school achievement scores that persuaded policymakers in the Brazilian government to take PAIC nationwide.

Mean Test Scores in Ceará and Neighboring States, 1995 to 2011

The estimated Portuguese mean scores in the SAEB surveys for fifth and ninth graders for Ceará (CE), border states (Borders), and the northeast states without Ceará (NEsCE) are shown in Figure 2. The mean test scores in all these state groupings decreased in 1995 to 2001. This decrease is associated with the universalization of basic education in this period, in which the proportion of lower socio-economic background students attending school rose significantly (Vieira, Vidal, & Costa, 2010). Scores increased after 2005 and increased more rapidly in Ceará schools from 2007 to 2011. The increase is 10 points for the ninth grade and 15 points for the fifth grade. For policymakers, this appeared to be evidence of PAIC's positive impact on student literacy skills.

Figure 3 presents the *Prova Brasil* mean mathematics scores for these same state groupings. The decline in scores after 1995 was smaller, particularly for ninth graders, and the difference in trends between Ceará and the other states is also smaller than for Portuguese. However, the overall gain for math was also large after 2005. The Ceará "advantage" after 2007 was much smaller than for Portuguese, but even so, there was a 3-point spread in the ninth grade and 6-point spread in the fifth grade by 2011.

Figure 4 shows the mean test scores on the *Prova Brasil* from 2007 to 2011 for fifth and ninth

TABLE 3

Ceará: Estimates of PAIC Effect on Portuguese and Mathematics Achievement Scores, DDD State-Level Fixed-Effects Model

	Portuguese					Mathematics				
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ceará × Fifth grade × 2011	0.096 (0.025)***	0.072 (0.024)**	0.079 (0.024)***	0.074 (0.026)***	0.080 (0.0351)***	0.179 (0.0315)***	0.146 (0.0324)***	0.143 (0.0312)***	0.143 (0.0327)***	0.120 (0.0465)**
Ceará × Fifth grade	0.013 (0.019)***	0.020 (0.0170)	0.004 (0.017)***	0.005 (0.018)	0.035 (0.023)	−0.013 (0.023)	0.003 (0.023)	0.008 (0.022)	0.000 (0.024)**	0.076 (0.03)**
Ceará × 2011	0.101 (0.021)	0.124 (0.019)***	0.118 (0.019)***	0.100 (0.020)***	0.137 (0.03)***	0.100 (0.026)***	0.123 (0.027)***	0.139 (0.026)***	0.125 (0.027)***	0.159 (0.04)***
Fifth grade × 2011	0.068 (0.016)***	−0.091 (0.015)***	−0.101 (0.016)***	−0.108 (0.017)***	−0.108 (0.022)***	0.111 (0.019)***	0.043 (0.02)**	0.051 (0.02)***	0.042 (0.021)	0.038 (0.029)
Ceará	−0.089 (0.041)**	−0.111 (0.035)**	−0.099 (0.033)***	−0.091 (0.035)**	−0.120 (0.035)*	−0.171 (0.043)***	−0.206 (0.04)*	−0.212 (0.036)***	−0.202 (0.038)***	−0.251 (0.037)***
Fifth grade	−1.206 (0.012)***	−1.478 (0.013)***	−1.458 (0.013)***	−1.486 (0.014)***	−1.333 (0.017)***	−1.111 (0.016)***	−1.547 (0.019)***	−1.519 (0.018)***	−1.542 (0.019)***	−1.361 (0.024)***
2011	0.099 (0.013)***	0.249 (0.013)***	0.246 (0.013)***	0.255 (0.014)***	0.217 (0.02)***	0.136 (0.016)***	0.193 (0.017)***	0.177 (0.016)***	0.190 (0.017)***	0.144 (0.026)***
Student characteristics	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Teacher characteristics	No	No	Yes	Yes	Yes	No	No	Yes	Yes	No
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Complete	Complete	Complete	ECD students	No ECD students	Complete	Complete	Complete	ECD students	No ECD students
R^2	.349	.396	.398	.394	.372	.285	.322	.323	.315	.302
Observations	274,845	274,845	274,845	192,113	74,234	181,806	181,806	181,806	134,810	46,954

Source. Prova Brasil, 2007 and 2011, microdata.

Note. PAIC = Pacto Nacional pela Alfabetização na Idade Certa; DDD = Difference-in-Difference-in-Difference; ECD = early childhood educational development. Values in parentheses are standard errors of the estimated coefficients.

*Statistically significant at 10% significance level. **Statistically significant at 5% significance level. ***Statistically significant at 1% significance level.

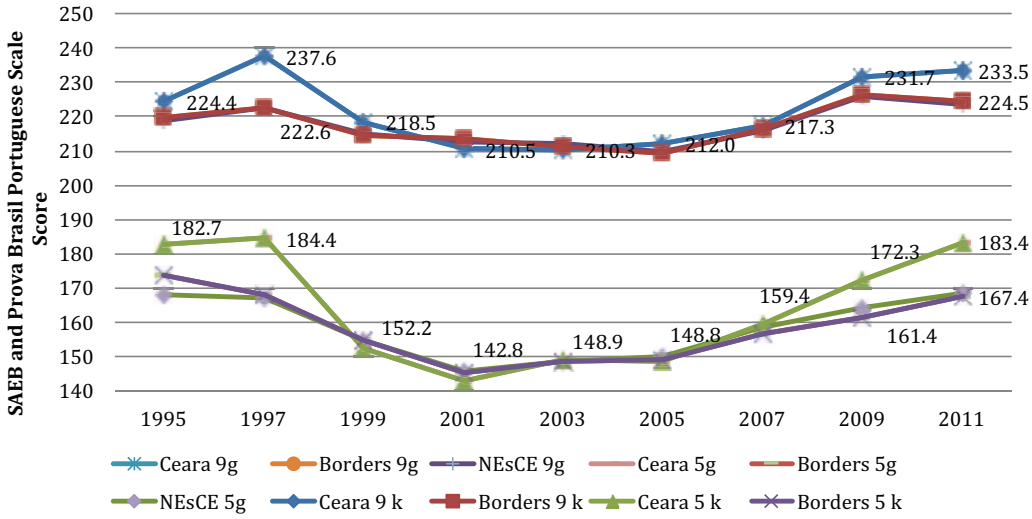


FIGURE 2. *Brazil: SAEB (1995–2005) and* Prova Brasil *(2007–2011) mean Portuguese scale scores, Ceará, bordering states, and other northeastern states, fifth and ninth graders.*
Source. Instituto Nacional de Estudos e Pesquisas (INEP), SAEB, 1995 to 2005; *Prova Brasil* 2007, 2009, 2011. Authors' estimates from publicly available microdata.
Note. SAEB = Sistema de Avaliação da Educação Básica.

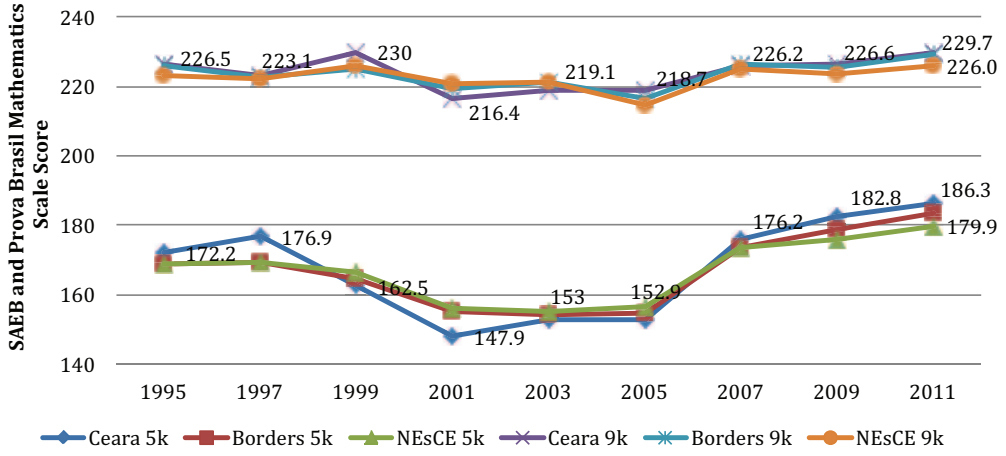


FIGURE 3. *Brazil: SAEB (1995–2005) and* Prova Brasil *(2007–2011) mean Mathematic scores, Ceará, bordering states, and other northeastern states.*
Source. INEP, SAEB, 1995 to 2005; *Prova Brasil* 2007, 2009, 2011. Authors' estimates from publicly available microdata.
Note. SAEB = Sistema de Avaliação da Educação Básica.

graders in Portuguese and mathematics in Ceará (CE) and its border states (Bordering) for those students who attended ECD and those without ECD. The significant difference of 10 points in the *Prova Brasil* scale between the ECD and non-ECD students remains constant after 2007, which supports claims for a positive relation between students' cognitive gains in pre-school and kindergarten and later success in conventional literacy tasks (Neuman & Dickinson, 2006). It also suggests that the PAIC program did not reduce the long-term disadvantage of students who did not attend ECD. Nevertheless, the slope for the Ceará fifth graders for both ECD and non-ECD students is steeper than the slope of test scores for bordering states' fifth and for ninth graders.

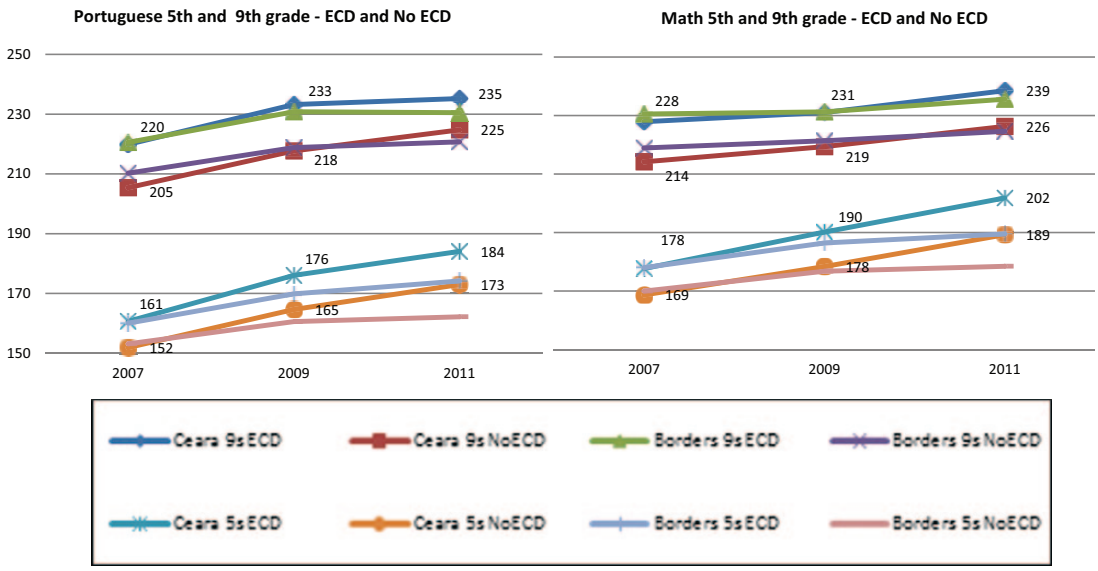


FIGURE 4. *Brazil:* Prova Brasil *(2007–2011) mean Portuguese and mathematics scores, Ceará and bordering states, by students with and without Early Childhood Education (ECD).*
Source. INEP, *Prova Brasil* 2007, 2009, 2011. Authors' estimates from publicly available microdata.
Note. ECD = early childhood educational development.

Estimating the Effect of PAIC on Student Achievement in Ceará

We now turn to our estimates testing whether this observed difference is the result of the PAIC intervention in Ceará. Table 3 shows the results of applying the DDD model (including state fixed effects) to estimate the impact of the PAIC program in Ceará on (treated) fifth-grade students' achievement compared with the achievement of (untreated) fifth graders in bordering states and (untreated) ninth graders in the same schools in Ceará (see Figure 1 for the schematic of the estimation model). The model also includes students' and teachers' covariates. In Table 3, we also present results for those students who did and did not attend ECD before entering primary schools. Quantile regression estimates are presented in Table 4 and in Figure 5 to estimate PAIC's effects across different groups of (treated) Ceará fifth graders.

The PAIC program effect was estimated to increase Ceará fifth graders' test scores by approximately 0.1 standard deviations (*SD*s) in Portuguese between 2007 and 2011 when using the DDD approach. The estimated effect of PAIC on mathematics is much higher, 0.18 *SD*s (column 6). The effect is shown by the coefficient of the interaction between the Ceará, fifth grade, and 2011 variables in columns 1 and 6 in Table 3. These estimates of PAIC effects are sensitive to controls for student and teacher covariates. The specification in columns 2 and 3 controlling for student and teacher covariates results in a smaller effect size for the PAIC on Portuguese scores (0.07–0.08 *SD*s). There was also a significant decrease in the estimated effect of PAIC on math scores in columns 7 and 8 (0.14–0.15 *SD*s), but the math effect remains higher than the effect on Portuguese scores. The average program impacts in all these controlled models are statistically significant at the 99% confidence level. We can also argue that these estimates represent a lower bound of the PAIC effect. As not all the 2011 Ceará fifth graders (our treatment group) were in schools with PAIC in 2007, some were not exposed to the full 3 years of treatment. Our data do not allow us to identify schools by year of PAIC implementation.

For comparison purposes, the appendix shows the estimates of the DD model for Ceará and bordering states in 2011 (gains of Ceará fifth graders in 2007–2011 compared with the gains of bordering states' fifth graders in 2007–2011) and estimates of the DD model for ninth graders in Ceará

TABLE 4

Ceará: Estimates of PAIC Effect on Portuguese and Mathematics Achievement Scores, Quantile Regressions, DDD State-Level Fixed-Effects Model

	Quantile 1	Quantile 2	Quantile 3	Quantile 4	Quantile 5	OLS
Portuguese						
Ceará × Fifth grade × 2011	0.067 (0.022)***	0.129 (0.020)***	0.178 (0.019)***	0.166 (0.022)***	0.148 (0.032)***	0.096 (0.026)***
Ceará × Fifth grade	−0.040 (0.013)***	−0.027 (0.012)**	0.000 (0.012)	0.043 (0.013)***	0.108 (0.019)	0.014 (0.019)***
Ceará × 2011	0.084 (0.017)***	0.115 (0.015)***	0.125 (0.015)***	0.141 (0.017)***	0.117 (0.024)***	0.100 (0.021)***
Fifth grade × 2011	0.030 (0.015)*	0.010 (0.014)	0.044 (0.013)***	0.126 (0.015)***	0.228 (0.022)*	0.071 (0.016)***
Ceará	−0.082 (0.013)*	−0.097 (0.012)***	−0.099 (0.012)***	−0.109 (0.013)***	−0.147 (0.019)*	−0.089 (0.041)***
Fifth grade	0.112 (0.009)***	−0.019 (0.008)**	−0.106 (0.008)***	−0.153 (0.009)***	−0.123 (0.013)***	−1.211 (0.013)***
2011	0.111 (0.012)***	0.121 (0.010)***	0.129 (0.010)***	0.119 (0.011)***	0.129 (0.016)***	0.100 (0.013)***
Mathematics						
Ceará × Fifth grade × 2011	0.177 (0.024)***	0.187 (0.022)***	0.211 (0.021)***	0.234 (0.026)***	0.259 (0.038)***	0.179 (0.032)***
Ceará × Fifth grade	−0.040 (0.017)	−0.020 (0.016)	−0.013 (0.015)	−0.002 (0.019)	0.015 (0.027)	−0.013 (0.023)
Ceará × 2011	0.020 (0.018)***	0.077 (0.017)***	0.120 (0.017)***	0.169 (0.02)***	0.190 (0.029)***	0.100 (0.026)***
Fifth grade × 2011	0.229 (0.016)***	0.123 (0.015)***	0.075 (0.015)***	0.042 (0.018)	0.129 (0.026)***	0.111 (0.019)***
Ceará	−0.129 (0.016)***	−0.171 (0.016)***	−0.184 (0.015)***	−0.206 (0.018)***	−0.270 (0.026)***	−0.171 (0.043)***
Fifth grade	−0.088 (0.012)***	−0.144 (0.011)**	−0.175 (0.010)***	−0.170 (0.013)***	−0.173 (0.018)***	−1.111 (0.016)***
2011	0.015 (0.012)***	0.117 (0.012)***	0.166 (0.011)***	0.236 (0.014)***	0.264 (0.020)***	0.136 (0.016)***

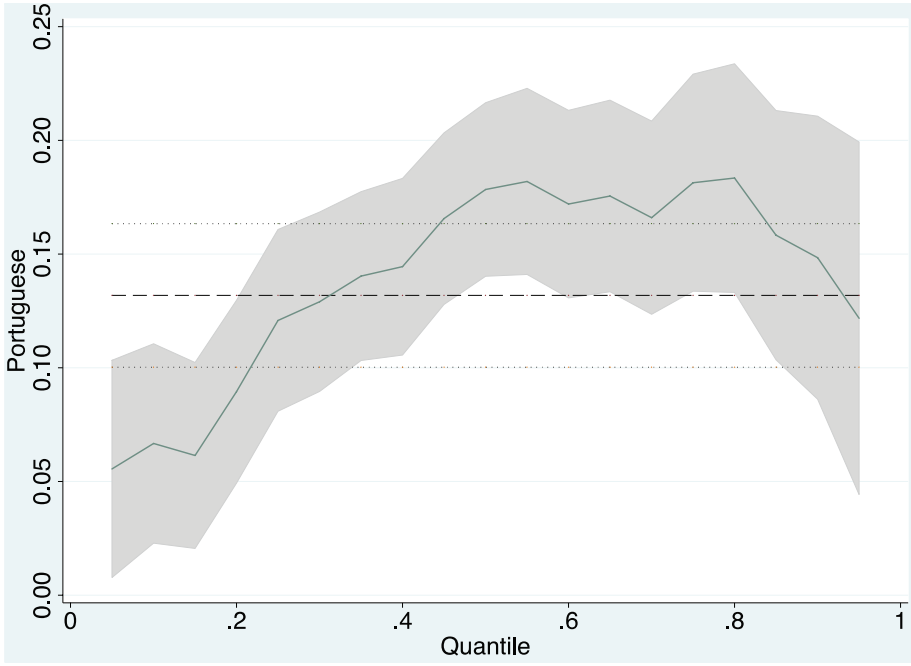
Source. Prova Brasil, 2007 and 2011, microdata.

Note. PAIC = Pacto Nacional pela Alfabetização na Idade Certa; DDD = Difference-in-Difference-in-Difference; OLS = ordinary least squares. Values in parentheses are standard errors of the estimated coefficients.

*Statistically significant at 10% significance level. **Statistically significant at 5% significance level. ***Statistically significant at 1% significance level.

state in 2011. The results in Table A1 show that the effects of the PAIC program are greater than the Table 3 estimates. The effect on Portuguese scores increases to 0.20 *SD*s and on mathematics scores to 0.28 *SD*s. However, the justification for including the third difference (compared with the gains of untreated ninth graders) is that, for whatever reason, the scores of ninth graders in Ceará also rose relative to the scores of ninth graders in the border states (Table A2), about 0.1 to 0.14 *SD*s (see also Figures 2 and 3). This is a type of differential "trend" effect in Ceará and bordering states.

There could be many reasons why the efforts in Ceará to raise early achievement may have affected scores of students not directly exposed to early intervention programs. The efforts by the SEDUC to place more emphasis on coherent



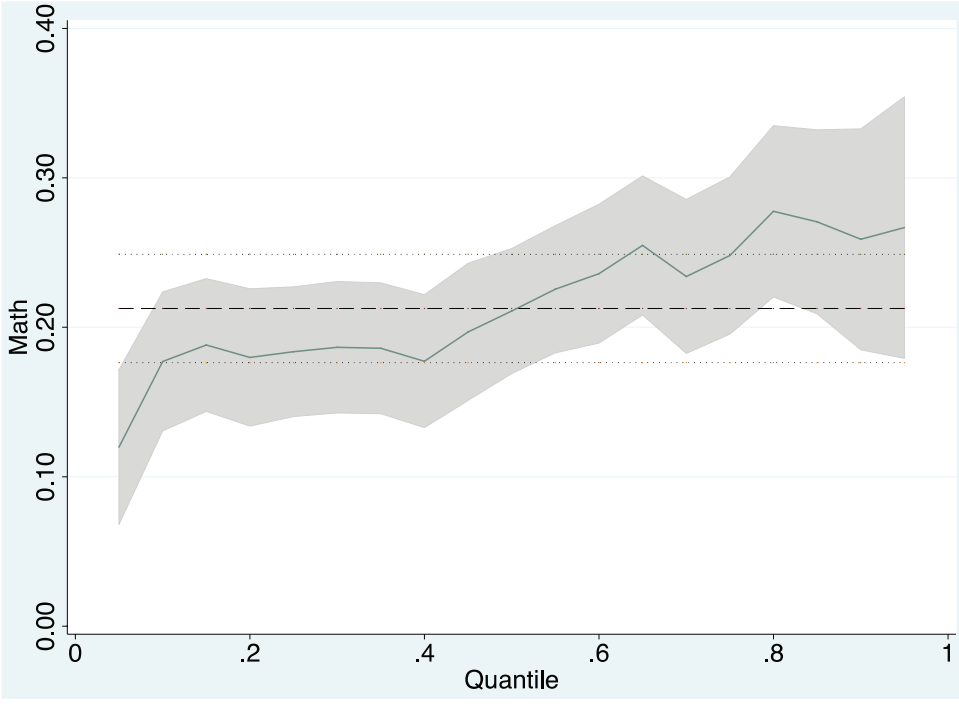


FIGURE 5. *Brazil, Ceará, and bordering states: Quantile regression results for PAIC effect,* Prova Brasil *Portuguese and mathematic score gains, 2007 to 2011.*
Source. Authors' estimates from *Prova Brasil*, 2007 and 2011, publicly available microdata.

instruction in municipal schools could have spilled over into the sixth to ninth grades, most of which are in the same schools as early grades. Should the effect of this effort be included in estimates of the impact of the literacy intervention? The argument for it being included is that the effort to improve school instruction in municipal schools overall is part and parcel of the PAIC program. The argument against is that we are trying to estimate the impact of a literacy intervention, and that only our fifth-grade cohort in Ceará was subject to that treatment. If we assume that unobserved "Ceará education effects" in our sample of schools that increased ninth graders' achievement in Ceará relative to achievement in bordering states also influenced fifth graders' gains in Ceará, we would want to net out those effects to estimate the PAIC effect.

The estimated PAIC effect using the DDD method may be biased even when controlling for important student, teacher, and classroom variables. Many students in Ceará and bordering states attended early childhood programs (ECD), and the PAIC effect could be confounded with the effect of ECD. However, when we make separate estimates of the PAIC effect on students with and without ECD, the results are similar to those for the total set of students. The 0.07 *SD* estimate for the PAIC effect on Portuguese achievement in Table 3, column 4, for students who attended ECD is about the same as the PAIC effect for no ECD students (column 5). Similarly, for mathematics achievement, the PAIC effect size for students who attended ECD is about the same as for no ECD students (0.14 vs. 0.12). This means that although the PAIC program had a statistically significant positive effect on students' achievement, it did not reduce the achievement gap between students who attended and did not attend ECD.

We test for possible heterogeneous effects of the PAIC intervention across the distribution of initial 2007 *Prova Brasil* scores of Ceará's treated fifth graders with quantile DDD state-level fixed-effects regressions. The results are presented in Table 4 and in Figure 5. They show smaller PAIC effects (0.06 and 0.12 *SD*s) on Portuguese achievement gains for fifth-grade students in the first and second quantiles (initial Portuguese scores in the lowest 20th percentile and 20th–40th percentiles than for students in the third and fourth quintiles—0.18 and 0.17 *SD*s). The PAIC effect fell slightly to 0.15 *SD*s for students in the fifth quintile. The pattern of PAIC effects is similar for mathematics gains. The effect size rises steadily from the lowest to the highest quintile, from 0.18 to 0.26. Thus, even though the PAIC intervention increased achievement gains across the entire range of student initial Portuguese and mathematics achievement levels, initially lower achievers benefit less from the PAIC intervention than initially median and higher achieving students in both Portuguese and mathematics. It is therefore likely that the PAIC failed to reduce the test-score gap between initially lower and higher scoring students, and may have contributed to spreading the gap between them.

Robustness Checks

As a check on the robustness of our estimates of the PAIC effect, we re-estimate the effects based on strategic changes in the specifications of the model. First, we estimate the impact of the PAIC program conditioning on the interaction between the literacy program and the ECD achievement gap of students in the same school and grade, as proposed in Equations 5 and 6 in the methodology section. The results are shown in Table 5. Table 6 shows the DDD estimates for three new model specifications: (a) removing the "fifth and ninth grades tested in both test years" restriction on schools included in the sample, (b) adding four other states to the control group (Sergipe, Alagoas, Bahia, and Maranhão), and (c) comparing fifth graders in the 2009 *Prova Brasil* wave with fifth graders in the 2007 wave to test the PAIC effect on the cohort that was only exposed to the program for 1 year. Again, not all students in the cohort were exposed even for a year because PAIC was not implemented in all Ceará municipal schools in 2007.

Table 5 shows the estimated results of the interaction within a school (controlling for school fixed effects) of the PAIC program effect with the achievement difference between students with and without ECD. The larger the ECD achievement gap in a school, the smaller the PAIC effect. Although the size of ECD interaction with the PAIC effect does change when we control for student, teacher, and classroom characteristics, the PAIC effect remains significantly

TABLE 5
Ceará: Estimates of PAIC Effect on Portuguese and Mathematics Achievement Scores, 4D (With ECD Interaction) School-Level Fixed-Effects Model

	Portuguese		Mathematics	
Dependent variable	(1)	(2)	(3)	(4)
Ceará × Fifth grade × 2011	0.212	0.113	0.303	0.275
	(0.02)***	(0.020)***	(0.025)***	(0.024)***
ECD × Ceará × Fifth grade	0.197	0.204	0.204	0.198
	(0.015)***	(0.014)***	(0.019)***	(0.018)***
ECD × Ceará × 2011	0.140	0.109	0.126	0.100
	(0.01)***	(0.009)***	(0.013)***	(0.013)***
ECD × Fifth grade × 2011	0.179	0.015	0.215	0.123
	(0.011)***	(0.010)	(0.007)***	(0.013)***
ECD × Ceará × Fifth grade × 2011	−0.299	−0.187	−0.315	−0.253
	(0.026)***	(0.025)***	(0.029)***	(0.028)***
Ceará	−0.958	−0.926	−0.536	−0.057
	(0.007)***	(0.009)***	(0.007)***	(0.013)***
Fifth grade	−1.256	−1.505	−1.142	−1.531
	(0.008)***	(0.009)***	(0.01)***	(0.012)***
2011	0.057	0.187	0.098	0.159
	(0.009)***	(0.009)***	(0.011)***	(0.012)***
Student characteristics	No	Yes	No	Yes
Teacher characteristics	No	Yes	No	Yes
School fixed effects	Yes	Yes	Yes	Yes
R^2	.399	.437	.349	.378
Observations	275,072	275,072	181,806	181,806

Source. Prova Brasil, 2007 and 2011, microdata.
Note. PAIC = Pacto Nacional pela Alfabetização na Idade Certa; 4D = four difference; ECD = early childhood educational development. Values in parentheses are standard errors of the estimated coefficients.
*Statistically significant at 10% significance level. **Statistically significant at 5% significance level. ***Statistically significant at 1% significance level.

negative and large, 0.19 to 0.25 *SD*s smaller for each *SD* of ECD achievement gap. If we assume that the ECD gap represents an unobserved quantity/quality measure of the ECD taken, this suggests that the higher the quantity/quality of students' ECD in a school, the smaller the impact of the PAIC on students with ECD. The result makes some sense. It implies that although the PAIC effect does not reduce average differences between students with and without ECD (Table 3), it may reduce differences between students with higher and lower quality ECD. At the same time, controlling for the interaction of the ECD gap, the average PAIC effect within a school is larger than the average PAIC effect across schools in Table 3 (as measured by the DDD coefficient—Ceará × Fifth grade × 2011), suggesting that the PAIC effect on students' gains is greater in schools where students made smaller relative gains from "fixed" unobserved factors.

Table 6 presents the results of estimating the PAIC effect with alternative specifications of the sample and treatment group. Columns 1, 2, 3, and 6, 7, 8 in Table 6 present estimates of model specifications similar to those in Table 3, columns 1, 2, 3 and 6, 7, 8. However, the student sample size is much larger in Table 6 because unlike in Table 3 we included schools in which either fifth graders or ninth graders (not necessarily both) were tested in a *Prova Brasil* test wave. With this expanded sample, the PAIC effect is somewhat larger than in Table 3 for

TABLE 6

Ceará: Estimates of PAIC Effect on Portuguese and Mathematics Achievement Scores, DDD State-Level Fixed-Effects Model, Robustness Tests for Less Restrictive School Sample, Additional Northeastern States, and 2009 Treatment Cohort

	Portuguese					Mathematics				
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ceará × Fifth grade × 2011	0.127	0.098	0.099	0.138		0.175	0.152	0.152	0.182	
	(0.042)**	(0.043)*	(0.044)*	(0.02)***		(0.034)***	(0.037)**	(0.038)**	(0.017)***	
Ceará × Fifth grade × 2009					0.039					0.058
					(0.037)					(0.039)
Ceará × Fifth grade	0.006	0.032	0.021	−0.008	0.005	0.008	0.043	0.034	−0.003	0.008
	(0.054)	(0.044)	(0.047)	(0.028)	(0.053)	(0.055)	(0.047)	(0.049)	(0.03)	(0.054)
Ceará × 2011 or 2009	0.013	−0.021	−0.029	0.001	−0.073	0.111	0.072	0.064	0.104	0.129
	(0.042)	(0.042)	(0.044)	(0.02)	(0.037)	(0.034)**	(0.036)	(0.038)	(0.017)***	(0.039)**
Fifth grade × 2011 or 2009	0.128	0.161	0.159	0.157	0.072	0.123	0.156	0.155	0.149	0.038
	(0.011)***	(0.01)***	(0.011)***	(0.022)***	(0.017)**	(0.011)***	(0.012)***	(0.012)***	(0.025)***	(0.016)*
Ceará	−0.100	−0.139	−0.130	0.026	0.086	−0.140	−0.196	−0.188	−0.034	0.037
	(0.026)**	(0.021)***	(0.023)***	(0.019)	(0.033)*	(0.032)**	(0.029)***	(0.029)***	(0.024)	(0.035)
Fifth grade	−1.186	−1.618	−1.604	−1.173	−1.186	−1.120	−1.578	−1.567	−1.108	−1.119
	(0.054)***	(0.057)***	(0.059)***	(0.028)***	(0.053)***	(0.055)***	(0.06)***	(0.062)***	(0.03)***	(0.054)***
2011 or 2009	0.261	0.280	0.275	0.232	0.239	0.167	0.190	0.183	0.141	0.062
	(0.011)***	(0.01)***	(0.011)***	(0.022)***	(0.017)***	(0.011)***	(0.012)***	(0.012)***	(0.025)***	(0.016)**
Bordering states	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes
Student characteristics	No	Yes	Yes	No	No	No	Yes	Yes	No	No
Teacher characteristics	No	No	Yes	No	No	No	No	Yes	No	No
Northeast states	No	No	No	Yes	No	No	No	No	Yes	No
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2009 treatment	No	No	No	No	Yes	No	No	No	No	Yes
R^2	.288	.342	.343	.268	.298	.231	.269	.270	.216	.225
Observations	489,491	473,659	473,659	793,520	433,562	489,491	473,659	473,659	793,520	433,562

Source. Prova Brasil, 2007, 2009, and 2011, microdata.

Note. PAIC = Pacto Nacional pela Alfabetização na Idade Certa; DDD = Difference-in-Difference-in-Difference. Values in parentheses are standard errors of the estimated coefficients.

*Statistically significant at 10% significance level. **Statistically significant at 5% significance level. ***Statistically significant at 1% significance level.

Portuguese (about 0.02–0.03 larger) but not for mathematics. The most likely reason for this somewhat larger PAIC effect on Portuguese is that the larger sample includes a higher fraction of schools where the PAIC program was implemented in 2007, and this longer average exposure had an effect on reading but not mathematics gains.

The coefficients of the model adding the other northeastern states in the control group are reported in columns 4 and 9. The PAIC effect is slightly larger than in columns 1 and 6. The results using the 2009 test results to estimate the PAIC effect on the cohort that was only exposed a maximum of 1 year to the PAIC program—just in the third grade (see Figure 1)—are shown in columns 5 and 10. They are positive for both Portuguese and mathematics but not significantly different from zero. This suggests very little effect on students' achievement with an average of less than a year of the PAIC program. This is an expected result.

Conclusion

Early-grade literacy interventions have been implemented in both developed and developing countries to mitigate the disadvantage faced by children entering primary school with poor or no pre-school education and home environments that prepare children poorly for school. The Program at the Right Age (PAIC) that began in Ceará state in Brazil in 2007 is a tiered, whole-school early-grade literacy intervention combined with financial and bonus reward mechanisms meant to focus schools on teaching children to learn to read by second grade.

We find that this complex and innovative policy resulted in an average impact (DDD effect) on students' achievement of 0.07 to 0.10 *SD*s in Portuguese and 0.14 to 0.18 *SD*s in mathematics. In terms of effect sizes of "successful" educational interventions, these are not considered large. However, they are larger than interventions that just focus on improving teaching (Carrasco, 2014; Hill, Rowan, & Ball, 2005). For example, as the *SD* of the *Prova Brasil* fifth-grade Portuguese and mathematics scores in Ceará in 2011 was about 40 points, the PAIC effect represents about 40% of the Portuguese score gap and 50% of the mathematics score gap between ECD and no ECD students (see Figure 4). Furthermore, because not all students in our treatment group were exposed to the full 3 years of the program, our estimates may be downward biased.

Initially lower achieving students appear to benefit less from the intervention than median and higher achieving students. This suggests that the intervention may tend to exacerbate differences in students' achievement as they move into higher grades. Neither does the program close the gap between students who attended early childhood programs before entering Ceará's primary schools and those who did not. In addition, the interaction of the PAIC program and the school ECD gap seems not to influence the program's effectiveness. Thus, this early-grade intervention has a moderate impact on improving the average results of students' achievement, but it does not succeed in closing the gap between at-risk and more advantaged students. This gap is correlated with being an at-risk student and not attending an ECD center.

Robustness checks on the DDD model show that alternative specifications generally make the estimated PAIC effect larger. Including a school fixed effect and the achievement gap for ECD increases the estimated PAIC program effect, as does expanding the sample to include more schools in Ceará and bordering states, as well as including additional Northeastern states.

As the PAIC program was designed mainly to improve early-grade literacy skills, its greater estimated effect size on mathematics achievement raises questions about whether our identification strategy correctly estimated the impact of PAIC. As we noted earlier, the PAIC program is not only a literacy intervention but also a set of incentives and supports designed to promote better instruction more generally and to improve students' achievement across subjects.

We attempted to separate these more general aspects of the intervention from the literacy component by differencing out the relative achievement gains of Ceará ninth graders, who could have been affected by the broader emphasis on improving instruction. The larger effect size of the intervention on mathematics achievement may indeed suggest that the most important components of this early literacy intervention are the greater emphasis placed on instruction, the

monitoring of teacher implementation of improved literacy teaching (which could reasonably be expected to also affect the teaching of mathematics in early grades), and the incentives for teachers to meet accountability goals for student achievement scores, which could also be expected to influence mathematics gains.

It may be that the materials and teacher training conducted to improve reading are not the most important part of the PAIC intervention, but rather primarily a vehicle to improve instruction overall. Furthermore, the pressure on and incentives for teachers and schools to improve test scores could work to improve student achievement across subjects regardless of a literacy intervention, although providing better materials and teacher training would have likely complemented increased accountability. Increasing teacher incentives to teach to Ceará's evaluation tests, such as the SPAECE-Alfa, could have spilled over to improve scores on the *Prova Brasil*. This possibly may also help explain why the Reading First evaluation showed no significant effect of that intervention on student achievement scores, but that our analysis of the PAIC effect does. The PAIC intervention's focus on organizational reforms, changing overall instruction in a context of initially quite low student achievement and poor teaching, and its closer coupling between those initiating the intervention and the teachers implementing the reform's materials and pedagogy make PAIC somewhat different programmatically and contextually from Reading First. This may have contributed to PAIC's greater impact on test scores.

Will the national scale-up of PAIC into PNAIC similarly benefit primary school students in the rest of Brazil? Perhaps. However, Ceará had a particularly favorable political context for state–municipal cooperation in this period (2005–2014), and the PAIC includes financial incentives for schools that are not part of the national program (see Table 1). If these elements are key to PAIC's success, early-grade literacy interventions in other Brazilian states may not produce similarly positive student achievement gains. We could learn a lot from studying the implementation of PNAIC with rigorous impact evaluations in each state. These could identify the effectiveness of the PNAIC program in different political and social contexts, and with and without financial incentives as part of the intervention.

Appendix

TABLE A1

Ceará: Estimates of PAIC Effect on Portuguese and Mathematics Achievement Scores, DD State-Level Fixed-Effects Model, Fifth Graders

	Portuguese		Math	
Dependent variable	(1)	(2)	(3)	(4)
Ceará × 2011	0.198	0.198	0.280	0.279
	(0.018)***	(0.017)***	(0.026)***	(0.026)***
Ceará	−0.054	−0.065	−0.145	−0.156
	(0.042)	(0.034)*	(0.047)***	(0.040)***
2011	0.167	0.145	0.238	0.225
	(0.011)***	(0.011)***	(0.016)***	(0.016)***
Student characteristics	No	Yes	No	Yes
Teacher characteristics	No	Yes	No	Yes
State fixed effects	Yes	Yes	Yes	Yes
R^2	.040	.112	.064	.105
Observations	164,390	157,511	1,113,361	106,815

Source. Prova Brasil, 2007 and 2011, microdata.

Note. PAIC = Pacto pela Alfabetização na Idade Certa; DD = Difference-in-Difference. Values in parentheses are standard errors of the estimated coefficients.

*Statistically significant at 10% significance level. **Statistically significant at 5% significance level. ***Statistically significant at 1% significance level.

TABLE A2

Ceará: Estimates of PAIC Effect on Portuguese and Mathematics Achievement Scores, DD State-Level Fixed-Effects Model, Ninth Graders

	Portuguese		Mathematics	
Dependent variable	(1)	(2)	(3)	(4)
Ceará × 2011	0.100	0.122	0.102	0.147
	(0.018)***	(0.020)***	(0.026)	(0.027)***
Ceará	−0.119	−0.140	−0.223	−0.280
	(0.042)***	(0.036)***	(0.042)***	(0.035)***
2011	0.100	0.266	0.133	0.196
	(0.011)***	(0.014)***	(0.016)***	(0.017)***
Student characteristics	No	Yes	No	Yes
Teacher characteristics	No	Yes	No	Yes
State fixed effects	Yes	Yes	Yes	Yes
R^2	.016	.101	.018	.097
Observations	110,455	108,836	75,813	74,898

Source. Prova Brasil, 2007 and 2011, microdata.

Note. PAIC = Pacto pela Alfabetização na Idade Certa; DD = Difference-in-Difference. Values in parentheses are standard errors of the estimated coefficients.

*Statistically significant at 10% significance level. **Statistically significant at 5% significance level. ***Statistically significant at 1% significance level.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors wish to thank the Lemann Center for Entrepreneurship and Educational Innovation in Brazil at the Graduate School of Education, Stanford University, for its financial support of this study.

Notes

1. Because the Literacy Program at the Right Age (Pacto pela Alfabetização na Idade Certa [PAIC]) is applied as a package of interventions, including financial incentives, to all first to third graders, it is not possible to disentangle the effects on student achievement of PAIC's various components.

2. For example, the Secretary of Education (Secretaria da Educação [SEDUC]) organizes regular meetings between all municipal education secretariats and SEDUC's regional managers to discuss the results of annual literacy assessment and to share strategies to improve the results.

3. The reward program is called School Grade 10 (Escola Nota 10; (http://www.paic.seduc.ce.gov.br/index.php/o-paic/premio-escola-nota-10). Large schools (those with large numbers of third graders) might be more motivated to improve their students' scores on the test because their total reward would be larger; but as Kane and Staiger (2002) have shown, small schools are much more likely to fluctuate in test results than the large ones, so they are also much more likely to occasionally get rewards in incentive systems based on high test scores. Thus, large schools may end up not being as motivated as small schools because it is more difficult for large schools to improve their students' average scores.

4. The reward is an increased share of total amount of tax collected by the state that is returned to the municipalities based on a meritocratic formula. The name of this policy is ICMS (Imposto sobre Circulação de Mercadorias e Serviços) Share (Cota Parte do ICMS). It reallocates tax revenue among municipalities based on a composite index with a 0.75 weight for the achievement of students in literacy exams (http://www.ipece.ce.gov.br/icms/icms).

5. In 2007, Brazil implemented a major reform that changed the entrance age into primary school from 7 years old to 6 years old. The reform simultaneously increased the length of primary school beginning with that first-grade cohort from 4 to 5 years, and the total length of basic education (primary plus middle school)

from 8 to 9 years. Thus, the 2005, 2007, and 2009 *Prova* tested students who had only been in primary school 4 years and 8 years, but the 2011 *Prova* tested cohorts of students who had been in school 5 years and 8 years. The 2015 *Prova* will be the first to test students completing basic education with the full 9 years of schooling. As a matter of nomenclature, the Brazilian government renamed the grades after 2007 to 2008, so that *Prova Brasil* tested students in the "fifth" and "ninth" grades even though the first cohort to actually be in school 5 years when they were tested was the 2011 fifth graders, and the first cohort to actually be in school 9 years in the ninth grade will be the 2015 ninth-grade cohort.

6. As mentioned in the previous note, the ninth graders in 2011 had been in school only 8 years, but in terms of the new definition, they were in the ninth grade. We shall refer everywhere in this article to students as being in the fifth and ninth grades, even when we discuss earlier years data.

7. In Grade 1, the effect size was 0.1, statistically significant at a 10% level.

8. In the *Prova Brasil* 2011, there is no subject taught variable in the data set of teacher questionnaires, because many of the teachers in the fifth grade teach both subjects. But there were some questions specific to math and to Portuguese teachers that allowed us to identify the teacher subject taught.

9. For an explanation of the *Prova Brasil* scale and the content of test items, see http://download.inep.gov.br/educacao_basica/prova_brasil_saeb/downloads/livretos/livreto_prova_brasil_2009.pdf

10. We use these two questions to categorize each student as having attended early childhood educational development (ECD) or not, and we use this individual student ECD dummy variable for all our regression estimates, except for our robustness check on the influence of school-level ECD interaction with PAIC, where we use the percentage of fifth graders with ECD in each school.

11. We also use panel data to test whether the PAIC effect holds up when the achievement of fifth- and ninth-grade Ceará students in 2011 attending schools that participated in all the 2001–2011 SAEB/*Prova Brasil* tests is compared with the fifth- and ninth-grade achievement of students in bordering states in 2007 and all previous years (trend effects). This is a greatly reduced sample of schools, but should be reasonably representative of all schools. These results are available from the authors on request.

References

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Ashdown, J., & Simic, O. (2000). Is early literacy intervention effective for English language learners? Evidence from reading recovery. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing*, *5*, 27–42.

Baker, S. K., Smolkowski, K., Smith, J. M., Fien, H., Kame'enui, E. J., & Thomas Beck, C. (2011). The impact of Oregon Reading First on student reading outcomes. *Elementary School Journal*, *112*, 307–331.

Barnett, W. S. (2002). Early childhood education. In A. Molnar (Ed.), *School reform proposals: The research evidence* (pp. 1-26). Charlotte, NC: Information Age Publishing.

Bertrand, M., Duflo, E., & Mullainathan, E. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, *119*, 249–275.

Carrasco, R. (2014). *Leveling the playing field: How can we address educational inequalities?* (Unpublished doctoral dissertation). Stanford Graduate School of Education, Stanford University, Stanford, CA.

Chaudhury, N., & Parajuli, D. (2010). Conditional cash transfers and female schooling: The impact of the female school stipend program on public school enrolments in Punjab, Pakistan. *Applied Economics*, *42*, 3565–3583.

Chiapa, C., Garrido, J. L., & Prina, S. (2012). The effect of social programs and exposure to professionals on the educational aspirations of the poor. *Economics of Education Review*, *31*, 778–798.

Cunha, F., & Heckman, J. (2010). Investing in our young people. In A. J. Reynolds, A. J. Rollnick, M. M. Englund, & J. A. Temple (Eds.), *Childhood programs and practices in the first decade of life* (pp. 381-414). New York, NY: Cambridge University Press.

Currie, J., & Thomas, D. (1993). *Does Head Start make a difference?* (NBER Working Paper No. 4406). Cambridge, MA: National Bureau of Economic Research.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, *92*, 605–619.

Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., Unlu, F., Bozzi, L., . . . Rosenblum, S. (2008). *Reading First impact study: Final report* (NCEE 2008-4038). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Gove, A., & Wetterberg, A. (Eds.). (2011). *The Early Grade Reading Assessment: Applications and*

interventions to improve basic literacy. Research Triangle Park, NC: RTI Press.

Gusmão, J. B., & Ribeiro, V. M. (2011). Colaboração entre estado e municípios para a alfabetização de crianças na idade certa no Ceará [Collaboration between the state and municipalities to make children in Ceará literate at the right age]. *Cadernos Cenpec*, *1*(1): 9–34.

Heckman, J., & Masterov, D. (2007). The productivity argument for investing in young children. *Review of Agricultural Economics*, *29*, 446–493.

Hill, H., Rowan, B., & Ball, D. (2005). Effects of teachers' mathematics knowledge for teaching on student achievement. *American Educational Research Journal*, *42*, 371–406.

Instituto Brasileiro de Geografía e Estadística. (2001). *Pesquisa nacional de amostra de domicilios (PNAD)* [Microdata files]. Retrieved from http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2011/microdados.shtm

Instituto Brasileiro de Geografía e Estadística. (2007). *Pesquisa nacional de amostra de domicilios (PNAD)* [Microdata files]. Retrieved from http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2011/microdados.shtm

Instituto Brasileiro de Geografía e Estadística. (2011). *Pesquisa nacional de amostra de domicilios (PNAD)* [Microdata files]. Retrieved from http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2011/microdados.shtm

Instituto Nacional de Estudos e Pesquisas. (2008). *Prova Brasil, 2007*. Brasilia, Brazil: INEP. Available from http://portal.inep.gov.br/

Instituto Nacional de Estudos e Pesquisas. (2010). *Prova Brasil, 2009*. Brasilia, Brazil: INEP. Available from http://portal.inep.gov.br/

Instituto Nacional de Estudos e Pesquisas. (2012). *Prova Brasil, 2011*. Brasilia, Brazil: INEP. Available from http://portal.inep.gov.br/

Jencks, C., & Phillips, M. (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.

Kane, T., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, *16*, 91–114.

Lyons, C. A. (1998). Reading recovery in the United States: More than a decade of data. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing*, *3*, 77–92.

May, H., Gray, A., Gillespie, J. N., Sirinides, P., Sam, C., Goldsworthy, H., Armijo, M., & Tognatta, N. (2013). *Evaluation of the i3 scale-up of reading recovery: Year One Report, 2011-2012* (Report RR-76). Philadelphia, PA: Consortium for Policy Research in Education.

Neuman, S. B., & Dickinson, D. K. (2006). *Handbook of early literacy research*. New York, NY: Guilford Press.

Ralaingita, W., & Wetterberg, A. (2011). Program effectiveness with EGRA: Impact evaluations in South Africa and Mali. In A. Gove, & A. Wetterberg (Eds.), *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy* (pp. 83-112). Research Triangle Park, NC: RTI Press.

Ravallion, M., Galasso, E., Lazo, T., & Philipp, E. (2005). What can ex-participants reveal about a program's impact? *Journal of Human Resources*, *40*, 208–230.

Schwartz, R. M. (2005). Literacy learning of at-risk first-grade students in the Reading Recovery early intervention. *Journal of Educational Psychology*, *97*, 257–267.

Slavin, R., Lake, C., Davis, S., & Madden, N. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, *6*, 1–26.

Tyler, J. H., Taylor, E. S., Kane, T. J., & Wooten, A. L. (2010). Using student performance data to identify effective classroom practices. *The American Economic Review*, *100*, 256–260.

Vieira, S. L., Vidal, E., & Costa, L. O. (2010). *Ensino fundamental: Fim de um ciclo expansionista?* [Basic Education: The end of an expansionist cycle?]. Brasilia, Brazil: Study commissioned by the Center for Strategic Studies and Management in Science, Technology and Innovation (CGEE).

Authors

LEANDRO COSTA is an economist at the World Bank in Brasilia. He received his PhD from the Federal University of Ceará and specializes in the economics of education and public policy evaluation.

MARTIN CARNOY is Vida Jacks Professor of Education at Stanford University. He has published extensively in the economics of education, political economy, public policy, and comparative education.

Manuscript received December 28, 2013
First revision received September 19, 2014
Second revision received November 23, 2014
Accepted December 30, 2014