



**“READING FIRST LITERACY PROGRAM: EVIDENCE FROM A
QUASI-EXPERIMENTAL EVALUATION IN CHILE”**

**TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN ANÁLISIS ECONÓMICO**

Alumno: Daniela Paz

Profesor Guía: Dante Contreras

Santiago, Junio 2017

Reading First Literacy Program: Evidence from a Quasi-Experimental Evaluation in Chile

Daniela Paz*

This Version: June, 2017

Abstract

This study evaluates Reading First literacy program that aims to significantly reduce the reading comprehension gap of students from vulnerable contexts in Chile. The program has in its DNA to be scalable at a sustainable cost. The results from a quasi-experimental evaluation (propensity score matching and difference in differences) finds that the program has a positive impact of 0.4 standard deviations (S.D.) on reading comprehension and text production results after 1 year and of 0.37 S.D. after 2 years in management of language results. Students with the program were more likely to have favorable feelings towards reading (pleasure for reading questionnaire). These results show that an expansion of the program is cost-effective and would improve the reading comprehension results in the short term.

Key words: quasi-experimental evaluation, comprehensive reading, pleasure for reading questionnaire.

JEL codes: I21, I26, I28, O15.

*Economics Department, University of Chile. E-mail: dpaz@fen.uchile.cl. Acknowledgments: I am grateful to Growth for All Educational Foundation (Fundacion Crecer con Todos) for the unconditional support and provision of data. I am also grateful to my thesis director Dante Contreras for his comments and support to carry out this evaluation. I also thank the Ministry of Education of Chile and Fondecyt (Grant No. 1140918) for funding support.

I. Introduction

How to improve the education of children from vulnerable backgrounds is one of the most relevant public policy questions in Chile. Since language results for the year 2015¹ show that 60% of students do not have basic reading skills and that this percentage increases to 78% among children of low socioeconomic level. Even though policy makers understand that reading comprehensively at an early age is critical for the development of cognitive abilities such as language (Coates et al. 2017) and could have positive impact on various daily life outcomes (from positive externalities in various areas such as math to better social and soft skills), the progress in this area has been null.²

There are also alarming numbers of functional illiteracy among adults in Chile. Recent data (PIAAC 2016; David Bravo and Microdatos Center, 2013) shows that 53% of the population between 16 and 65 years does not understand what they read, while only 15% can make complex inferences. There has not been any significant change compared to the measurements made in 1998, reflecting it has not been possible to generate relevant changes in the last 15 years. The lack of these basic skills is also more pronounced among vulnerable and less educated layers of society.

Results indicate that the vast majority of the adult population of Chile is not able to compare and integrate information, nor to make inferences. The repercussions of this are big. This would affect their performance in terms of productivity and income (Cunha et al., 2006) and other non-cognitive and daily life skills (Heckman et al., 2012; Grantham et al., 2007). Early educational problems persist over time indicating that the education received by students from vulnerable contexts at an early age is a fundamental of inequality. A segregated educational system such as the Chilean one generates an uneven distribution of reading comprehension skills that lead to inequalities in working conditions, income and access to opportunities that become more pronounced over time.

This paper suggests that early interventions focused on the development of comprehensive literacy are especially attractive to improve the country's educational performance. The purpose of this study is to present the evaluation design and the results of the impact evaluation of an educational program called Reading First. Reading First is a

contextualization of an "Effective Core Reading Program"³ and is focused on teaching literacy. The program uses a tutor methodology in which a trained tutor works alongside with the teacher, favoring the installation of participatory and innovative methodologies and modeling pedagogical practices. The final objective of the program is to leave installed capacities in the teacher after 3 years of mentoring inside the classroom⁴. The program is able to scale up rapidly because it works with tutors that are local residents of the region where the program is being implemented.

The results measured using a quasi-experimental methodology show that Reading First has a positive impact of 0.4 and 0.37 standard deviations (S.D.) on the results of reading comprehension and text production (RC-TP) after one and two years respectively. A propensity score matching technique was used to match treatment schools of the municipalities of Arica and Pedro Aguirre Cerda (PAC) to San Joaquin (SJ). The total sample consisted of a total of 40 schools or 1,569 students. All school considered were public schools and the measurements begun on school year, at the beginning of 2015 and persisted until December of 2016, allowing to evaluate the impact of the program with a difference in differences technique.

We argue that the results are an approach to obtain causal empirical evidence. The results give insights of the implementation process of the program in two different municipalities and the mechanisms behind it. The cost-effectiveness analysis presented is useful to improve public policy decision making where Reading First appears as a cost-effective program that has great potential to be scaled up.

The relevance of the study is that it helps to guide the decision making process of educational investment. On the other hand, it differs from others in that it seeks to understand in detail the mechanisms by which the program is having effect. This is valuable both for understanding the final impact of the program and for an internal evaluation of Growth for All Foundation (foundation in charge of implementing the program). This is especially relevant in the context of present-day Chile where major educational reforms are under way and the resources that can be directed to these types of impact initiatives are rarely evaluated and very limited.

The rest of this paper is organized as follows. The next section II explains the national and international literature in which the program is framed. Section III gives a brief description of the

program, which details its main axes as well as how it works. Section IV shows descriptive statistics, such as sample size and balance between groups. Section V explains the identification strategy and then Section VI presents the results. Section VII presents some robustness checks and section VIII a cost-effectiveness comparison of the program. The paper ends in section IX with some discussion and conclusions.

II. Literature Review

Much of the literature of the effects of early childhood interventions is based on international data. In Chile, there is a lack of evaluations that study the effects of early literacy programs (Bedregal et al. 2007). Bedregal and Villalon (2008) argue that early literacy is critical for lifelong learning and the full development of people, a task that exceeds the limits of formal education and that available evidence for the country comes from scattered and non-impact based studies. The early development literature is guided by studies such as Contreras and Thievos (2014), Contreras and González (2015) or Urzúa(2011) that focus for example on the effect of the enrollment in child care centers. Both studies find enrollment in day care or preschool during the first three years of the child's life have a significant role in the child's psycho-motor development.

Our study is framed in the next educational level (children are 6 years old) and therefore contributes to analyze the effectiveness of interventions that occur later in time, at the limit of the so called "window of opportunity". Along this line, an evaluation that sought to improve reading skills of fourth grade students (Country Service in Education Program, Cabezas et al., 2011) that worked through volunteer tutors who attended language classes to form reading support groups. After three months with the program, positive and significant results were observed on the results of cognitive and non-cognitive tests (reading taste questionnaire, same instrument that the one used here), but only in one of the evaluation regions (Metropolitan Region).

International studies also emphasize how highly cost-effective these interventions are, as well as their power to reduce inequalities of origin. Literature such as Cunha et al. (2006) or Carneiro and Heckman (2003) highlight its productive role, especially in terms of cognitive and non-cognitive results. The link between cognitive skills and income is summarized in that the increase of a standard deviation in adult test scores corresponds to an increase in income that goes between 17 – 22%⁵ on average. The central point is that early interventions feed future interventions.

The closest to our type of interventions, researchers from the World Bank as Patrinos (2016) (Blog Document) highlight the importance of literacy to access information and increase productivity, especially for the poorest. It also analyzes their possible impact on the diffusion of new technologies, and thus on their high economic returns. Their evaluation is based on programs such as "READ PNG" in Papua New Guinea where reading scores increased by 0.51 S.D., and costs were \$60 per pupil. This program assumes continuous support to the teacher, with the delivery of institutional material and tools to identify students that are being left behind (same lines of the Reading First program studied here and detailed in the next section).

With respect to the effectiveness of programs such as Reading First work through mentoring, the empirical study by Burley et al.(2007) investigates the effect of mentoring on achievement in reading basic education students with cognitive deficits and specific students who are part of MORE (Mentoring in Ohio for Reading Excellence) project. This study highlights the importance of tutoring on early reading, reading interventions, and one-on-one instruction. Students were tested both before and after interventions with standard measures of reading performance and those students who received treatment were matched by students in the control group at similar schools in the district. The results show significant positive effects of the intervention. Generally, the treated students showed improvements each month compared to the students in the control group during the six months of the intervention.

Following the same research line, Borman et al.(2007) used a cluster-level randomization design, where schools were randomly assigned to implement a program called "Success for All". The program focuses on schools and consists of early prevention and intensive intervention designed to detect and solve reading problems as early as possible. Part of the solutions established by the program are: improving teachers' classroom management and increasing parental involvement both in the learning of children and in the school in general. Each school was provided with a tutor or facilitator of the program, who is responsible for supervising the daily operation of the program, providing assistance and coordinating. The duration of the program was three years and was set from kindergarten to the second grade. The results reveal statistically significant school-level effects in treated versus control children, with improved reading comprehension of one-third of the standard deviation.

In summary, there is a complete set of literature that proves that literacy program can have big short time effects. These studies are internationally based and so there is a need to discover what happens in Chile. This study pretends to fill that gap and also extend the evaluation period to one in which the students follow up is extended to more than one year.

III. Reading First Program

Reading First is a contextualization of Effective Core Reading Programs (ERP), and is focused on teaching literacy. The objective of the program is to ensure that all students attending public or private-voucher schools, read and write comprehensively from 1st grade. In particular, the main axes of the program are:

1. Innovative and interactive methodology: based on most effective international evidence (ERP) with special focus on reading comprehension and vocabulary. Oral communication is also an important focus of the program, since a good development of oral language is key for the child to be motivated to develop written language.
2. Professional development model: importance of teachers methodological appropriation of the literacy model and strategies.
3. Support to the management team.
4. Evaluation and monitoring learning: performance evaluations for each program unit (set of quizzes and standardized tests as part of the materials handled).
5. Work with lagged students: this is critical for the program, it is the "Reading First More" program, that establishes a systematic work with Differential Educators for students who require additional support.
6. Parents involvement.

The program is implemented within the language class (10 hours per week) throughout the academic year (March to December), and is articulated and approved with the Plans and Programs for Language and Communication from the National Curriculum. The program is planned to be implemented from pre-kinder to 3rd grade.

The work methodology is based on continuous support to the teacher through tutors or mentors. Tutors must be teachers and have at least 3 years of work experience. They receive

120 hours of training that allows them to be better qualified in the Reading First Methodology, so that they can guide teachers, help them anticipate problems, make decisions and understand the meaning of each specific activity. One of the most important aspects is that the program is able to scale up rapidly because it works with tutors that are local residents from the region/municipality where the program is being implemented.

The role of the tutor is to work alongside and give feed-back to the teacher with the purpose of favoring the installation of participatory and innovative methodologies and modeling pedagogical practices. There are also a set of materials that are given in order to facilitate the work: daily schedules of classes, students work guides, students evaluations and a classroom library. The final objective is to leave installed capacities in the teacher after 2 years of company inside the classroom.

At present, Reading First is implemented in 13 municipalities throughout the country (more than 100 schools). The objective of this paper is to evaluate the impact of the program in only two of these municipalities, Arica, in the north part of the country and Pedro Aguirre Cerda (PAC) in the center, where the program was implemented as a pilot because of an evaluation alliance with the Ministry of Education of Chile (MINEDUC). MINEDUC wanted to evaluate the programs impact in two differently located municipalities with different pedagogical management capacities, to verify the programs ability to be scaled. There were no other selection considerations, but to definitely reject alternative hypothesis we make further explorations of treated and control schools previous evolution in the section VII.B Falsification Exercise (page 15).

When comparing both municipalities, the most recent CASEN (2015 household survey) indicates there is a 8,9% of poverty rate in Arica (mean household income of \$991, 364 CHL pesos) and 11% in PAC (mean household income of \$941, 276 CHL pesos). 76% of the population reads and write in Arica, while 78% in PAC. Arica has an enrollment of 8,812 students in primary education, while PAC has 2, 434 students. The average municipal school attendance in 88% for Arica and 80% for PAC. One of the biggest difference between both is the presence of ethnic groups were Arica has a 26% of ethnic population (mainly Aimara indigenous group) and PAC only has 6,3%.

IV. Descriptive Statistics

A. Implementation Data

The treatment group consists of a total of 1,297 students, who come from 33 public schools in 2 municipalities of Chile: Arica, in the north part of the country and Pedro Aguirre Cerda (PAC) in the center. The following Table 1 shows the detail of the sample size:

Table 1: Sample Size of Treatment and Control Groups

Variable/ Municipality	Treatment			Control				
	Arica	PAC	Total	San Joaquin	Quilicura	Cerro Navia	Puente Alto	Total
N Schools	20	13	33	7	4	13	16	40
N Classes	31	13	44	10	9	17	30	66
N Students	937	360	1297	272	343	568	1183	2366

Source: Own elaboration based on implementation data.

The control group used in the main identification strategy consists of 272 students from 7 public schools in the municipality of San Joaquin (SJ).⁶ For both Control (Arica and PAC) and control (SJ) groups, in March 2015 (early in the school year) a baseline survey and diagnostic test were implemented. At the end of the year, in November and December two other instruments were applied to measure the impact after 1 year of program: a Taste for Reading Questionnaire and a Reading Comprehension and Text Production (RC-TP) standardized test for 1st graders. To measure the impact after two years of implementation a second RC-TP test was carried out.

It is important to highlight that the inclusion of other municipalities, in addition to SJ, as part of the control sample in Table 1 (Quilicura, Cerro Navia and Puente Alto) is due to the late inclusion of these to the experiment. At first, the evaluation design considered all the public schools of Copiapo municipality, also on the north of the country, as part of the control group. However, due to the occurrence of a flood disaster while carrying out field activities, this option was discarded. It is estimated that there were 28,000 victims in the region (the president called the State of Constitutional Disaster Emergency). For public education, the implications were that schools were used as shelters and classes were canceled for more than 2 months. This made it impossible to comply with the original evaluation design.

Given this, a matching procedure was performed for the post-tests between the Arica treated and a control composed of a set of municipalities of Santiago. The summary of the data gathering process can be seen in Appendix 1 and the results of the Matching procedure can be seen in Appendix 2.⁷ This allowed to compare the results with all the municipal establishments of Cerro Navia, Quilicura and Puente Alto in the robustness checks section. The total sample consists of 3,278 students from 73 municipal establishments in the country.⁸

B. Balance Between Groups

The following Table 2 shows the results of initial characteristics balance between the available treatment group (Arica and PAC) and control (San Joaquin), after the implementation of the propensity score matching (PSM). The details of this strategy will be explained on the next section of Identification Strategy.

Table 2: Statistical Balance Between Treatment and Control Groups

Variable	Treatment	Control	(p-value)	N
Diagnostic Tests	16.4	17.03	0.023***	700
Income (CHL pesos)	362,974	367,954	0.591	700
Education mother (years)	12.2	12.6	0.145	700
Education father (years)	11.54	11.73	0.492	700
Books	23.5	23.8	0.87	700
Gender (1=male)	0.49	0.52	0.461	700
Home stimulus	1.87	1.93	0.334	700
Home rules	0.91	0.88	0.93	700
Ethnic (any parent)	0.25	0.07	0.00***	700
Class size	34.13	34.33	0.674	700

Source: Own elaboration based on base-line survey information.

Notes: To estimate de PS index we estimate a probit regression using cluster standard errors.

Both groups appear to be balanced on most characteristics⁹. However, the control group scores higher on the baseline test, although it is less than one point of difference (less than a 5%

difference on the total test score, that does not mean a different in terms of performance level) . On the other hand, the treatment group has a higher percentage of students with indigenous ancestry which is generated especially by the greater proportion of indigenous people in the northern municipality of Arica. When the same comparison is done between the municipalities of PAC and San Joaquin the difference is not significant.

When controlling for the ethnics ancestry dummy in the PSM, the number of observations that are on the common support is reduced and other important variables become unbalanced. There is a trade-off between using more characteristics on the PSM index and the efficiency of the results. We discuss this further in Appendix 4.

V. Identification Strategy

A. Reading Comprehension and Text Production Test (RC-TP)

A propensity score matching and then differences in differences technique was carried out to compare the evolution of a treatment and control group before and after (1 and 2 years) the program¹⁰. Following Rosenbaum and Rubin (1983), the Average Treatment Effect (ATE) of the program can be defined as follows:

$$\Delta_i = Y^A - Y^N \quad (1)$$

Where Y^A defines the test result (RC-TP)¹¹ for student i with the program and Y^N for the control group. When estimating the impact of equation 1, the classic problem of impact evaluation arises because of the impossibility to observe the same student with and without the program. What we observe is:

$$Y_i = D_i Y^A + (1 - D_i) Y^N \text{ with } D = 0, 1. \quad (2)$$

Denoting P as the probability of observing the student with $D = 1$, the ATE is:

$$ATE = P[E(Y^A/D = 1) - E(Y^N/D = 1)] + (1 - P)[E(Y^N/D = 0) - E(Y^N/D = 1)] \quad (3)$$

The main problem of causal inference is that it is not possible to observe the counterfactuals $E(Y^A/D = 0)$ and $E(Y^N/D = 1)$. This paper addresses this problem by using the PSM method that summarizes the pre-treatment characteristics of each subject into an index

variable and then uses this to match similar individuals. The PSM can be represented by:

$$p(X) = Pr[D = 1/X] = E[D/X]; p(X) = F\{h(X)\} \quad (4)$$

Where F is represented by a normal cumulative distribution in this case and X is a set of pre-treatment characteristics: baseline test, gender, parents education, income and books at home.¹²

Once the propensity score is estimated, the ATT can be estimated:

$$ATT = E\{Y_{iA} - Y_{iN/D = 1}\} \quad (5)$$

$$ATT = E\{E\{Y_{iA} - Y_{iN/D = 1}, p(X)\}\} \quad (6)$$

$$ATT = E\{E\{Y_{iA}\}/D=1, p(X)\} - E\{E\{Y_{iN/D = 0}, p(X)\}/D = 1\} \quad (7)$$

To match the treated with the untreated the nearest neighbor matching (NNM)¹³ procedure was used in which treated students are matched with the control that has the closest propensity score. It is applied with replacement of the control units. We go one step further and after using the PSM, we use the Difference in Differences technique, in which the ATT is obtained as the difference of all the average results:

$$ATT = (after - before)_A - (after - before)_N \quad (8)$$

The simplified version of the main equation to be estimated would be the following¹⁴:

$$Y_{ijt} = \alpha + \beta_1 Treat_{ij} + \beta_2 Time_t + \beta_3 Treat * Time_{ijt} + \varphi X_i + s_{ijt} \quad (9)$$

Where Y_{ijt} is the test result for student i , of group j in time t . $Treat_{ij}$ is a dummy variable that takes the value of 1 if it is a treated student. $Time_t$ is a time variable that takes the value of 1 if its at the end of the year (December) or 0 if its at the beginning (March). Finally, X_{ijt} are control variables.

B. Taste for Reading Questionnaire

The taste for reading questionnaire contains 10 questions that can be categorized according to 4 sections (the 4 sections are: interest in reading, self-perception as a reader, enjoyment of reading and perception of reading in school. For more details on the questionnaire see Appendix 11). Within the items it is possible to create indexes that move between 0 and 3.

A value of zero indicates little affinity with reading and a three means the opposite. As these are ordered variables, the correct methodology in this case is to use an ordered logit regression model¹⁵. This model is used to see the impact of the program on 4 dimensions and for each equation a binary treatment variable is included that represents the ITT.

VI. Results

A. Reading Comprehension and Text Production (RC-TP)

The following Graphs 1, 2 and 3 summarize the results in terms of total RC-TP score. Each graph uses Arica and PAC as treatment (in red) and SJ (in gray) as control to show differences between groups distribution and evolution.

Graphs 1, 2 and 3: Reading Test (RC-TP) Results in Treatment and Control Groups

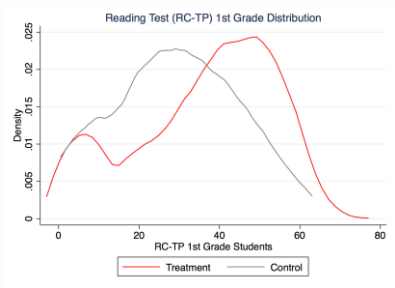


Figure 1. RC-TP Distribution 1st Grade

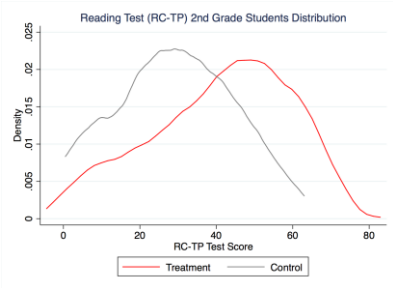


Figure 2. RC-TP Distribution 2nd Grade

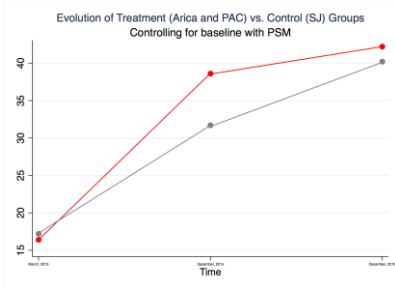


Figure 3. Difference in Difference Summary

Graphs 1 and 2 show that the distribution of scores of treated students after 1 and 2 years is on the right, that is, associated with higher scores, than that of the control group. Graph 3 shows the evolution of both groups before and after the program. The results of December 2015 reflect that students with the program obtain, on average, between 7-10 additional points in the RC-TP test. Previous literature (Cabezas et al. 2011) show that for 4th grade students 1,5 points in the RC-TP test are equivalent to 6 Simce points. Assuming this as certain fact for our case, the effect of the program would be to increase Simce scores by 28-40 points. In practical terms, this means evolving from an emerging performance level, where needs predominate over strengths, to a developed one, where strengths overcome weaknesses (or more than 20% difference in the total score). Considering historical average scores of the RC-TP test (only available before the year 2012), the additional points allows treated public schools to catch up with the average total score private schools of the country. This difference is substantially reduced during the second year (2 points of difference and not statistically significant for the RC-TP total score) reflecting a fade out effect of the program after 1 year of implementation.

The results of the main regressions for the RC-TP test and each of its sections (Reading Com- prehension (RC), Text Production(TP) and Management of Language (ML)) can be seen in the following Table 3:

Variable	Sample	Treated	Control	Diff.	S.E.	T-stat	Obs.
RC-TP							
1st grade	Unmatched	0,35	-0,2	0,55	0,12	4,65	537
	ATT	0,35	-0,05	0,4	0,15	2,71	537
RC	1st						
grade	Unmatched	0,27	-0,3	0,57	0,11	5,23	537
	ATT	0,27	-0,19	0,46	0,14	3,2	537
TP	1st						
grade	Unmatched	0,39	-0,22	0,61	0,12	4,96	537
	ATT	0,39	-0,14	0,53	0,16	3,22	537
ML	1st						
grade	Unmatched	0,29	-0,03	0,33	0,11	3,05	537
	ATT	0,29	0,05	0,24	0,14	1,67	537
RCTP	2nd						
grade	Unmatched	0,03	-0,22	0,25	0,12	2,05	438

		ATT	0,03	-0,22	0,25	0,16	1,55	438
RC	2nd	Unmatched	0,05	-0,24	0,29	0,13	2,3	438
		ATT	0,05	-0,18	0,23	0,17	1,33	438
TP	2nd	Unmatched	-0,13	-0,19	0,06	0,14	0,43	438
		ATT	-0,13	-0,25	0,12	0,17	0,7	438
ML	2nd	Unmatched	0,1	-0,3	0,4	0,12	3,37	438
		ATT	0,1	-0,27	0,37	0,15	2,51	438

Source: Own elaboration based on RC-TP test results.

Notes: All results are standardized and show the difference in differences coefficients. The Stata command `psmatch2` is used and propensity score and treatment effect are estimated simultaneously, so there is no need for further correction of the standard errors.

The results show there is a significant effect at 1% level of confidence of the program over the first year, measured by the total RC-TP test score, the Reading Comprehension (RC) score and the Text Production (TP) sections result. The magnitude of the effect goes from 0, 4 S.D. in the total score after one year of implementation to 0, 46 S.D. in the RC section and 0, 53 S.D. in the TP section. Nevertheless, this effect then fades out after the second year and it is only significant for the Language Management (LM) section in which the magnitude is 0, 37 S.D.. When dividing our sample by treated municipality we see the program has an heterogeneous impact: there is a significant effect in Arica (0.27 S.D. in 2nd grade RC-TP total score and significant at the 10% level), while there is no effect in PAC.

One possible explanation for this fading out effect is that there were clear implementation differences between treated municipalities. On one hand, Arica is considered a well-managed municipality with well-aligned stakeholders (school principals, pedagogical-technical unit (UTP) managers and regional educational directors). On the other hand, PAC has severe pedagogical management issues. We explore this differences using two data sources: first, we use administrative data to analyze rotation and education of teachers and the directive team (see Appendix 6 for more detail); then we use the Work Task Motivational Scale for Teacher (WTMST)

to study a different measures of teachers motivation in several daily activities (Class preparation, Teaching, Student Evaluations, Classroom Management, Administrative Tasks and Complementary Tasks). The unit of analysis used in both strategies is the classroom, were we have only 44 classes (one response for each teacher) and therefore we do not have enough variation to include this factor as an interactive variable in our main regressions.

Following our first strategy (administrative data) we see that only 60% of the teachers were hired at the beginning of the school year (Mineduc "Docentes" Data-base). This made it impossible to train some teachers with the program before classes started. The same problem happened with school directors from 3 schools in PAC. Another management performance indicator was teachers rotation and short medical licenses without replacement, were 30% of the lost classes were due to this factor (Foundations internal implementation information). Also in PAC, the "Reading First More" program addressed to help lagged students was not implemented because the necessary meetings with the management teams were not carried out. Finally, while educational background were similar, UTP managers from Arica had less years of work inside the school (12 years on average) than PACs (17 years on average) (see Appendix 10 for more differences).

Table 4: Teachers Motivation in Activities (WTMST)

Motivation Type	Activity	Arica	PAC
Amotivation	Class Preparation	1.8	1.3
	Teaching	1.4	1.2
	Evaluation of Students	1.4	1.4
	Classroom Management	1.3	1.4
	Administrative Tasks	1.8	2.2
	Complementary Tasks	2	1.7
Identified Regulation	Class Preparation	6.7	6.4
	Teaching	6.6	6.5
	Evaluation of Students	6.2	6.3
	Classroom Management	6.4	6.5
	Administrative Tasks	5.7	6.3

	Complementary Tasks	5.6	6.2
--	---------------------	-----	-----

In terms of the WTMST, the above table shows that PAC teachers feel more "Amotivation" (individuals are amotivated when they have no intention of engaging in a particular behavior and do not really know why they are doing it) in doing Administrative Tasks, which refers specifically to meetings with directive team to solve complicate student cases, meeting with administrative team and other teachers, while there not seems to be a significant difference in other activities. Other interesting result is found analyzing the the identified regulation type of motivation: behavior that individuals choose to perform because it is congruent with their own values, although the activity is not intrinsically interesting, that the average is also higher for PAC. This scale reflects that teachers also feel that there are important pedagogical management issues in PAC¹⁶.

Finally, even though educational outcomes in Chile are usually measured using other indicators (Simnce test) we prefer to standardize the result to illustrate relative the performance of programs. By standardizing the scores, the Reading First has an effect ranging between 0.4 - 0.37 S.D., depending on the year considered. Within the literature this is considered a big effect especially when compared with other educational programs of the country (National System of Evaluation of Performance or SNED), Full School Day (JEC), Service Country Education, among others) and the all of the educational literature (0.05-0.2 S.D.). The cost-effectiveness analysis in section VIII studies these comparisons with more depth.

B. Taste for Reading Questionnaire

The marginal effects of ordered logit of the section of Perception of Reading at School are found in the following Tables 5 and 6. As mentioned before, the taste for reading questionnaire has 10 questions. For each question, the student must choose between 3 categories: the zero category (CAT0 in the next page table) indicates little affinity with reading and a the third category (CAT3 in the next page table) means the opposite.

Table 5: Ordered Logit for Perception of Reading at School

VARIABLES	(1) CAT0	(2) CAT1	(3) CAT2	(4) CAT3
1. Control	0.116*** (0.0214)	0.141*** (0.0213)	0.332*** (0.0226)	0.410*** (0.0427)
2. Treatment	0.0804*** (0.0118)	0.107*** (0.0133)	0.301*** (0.0199)	0.512*** (0.0239)
Observations	550	550	550	550

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Difference between Marginal Effects

VARIABLES	(1) CAT0	(2) CAT1	(3) CAT2	(4) CAT3
Treat.vs.Control	-0.0359* (0.0191)	-0.0343** (0.0173)	-0.0310** (0.0129)	0.101** (0.0477)

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The results show that a student in the control group is more likely to say that they like "Nothing" (see CAT 0) to read, and on the other hand, a student in the treated group is more likely to say that he/she likes "a lot" to read (see CAT3). This difference, in this section, is statistically significant between groups. Other sections results can be found in Appendix 7.

VII. Robustness Checks

A. Difference in Differences, All Data

A second approach to evaluate the impact of the program using the RC-TP test results is to use the set of data for which there is no base-line information but both follow-up RC-TP test

exists. As mentioned in the Data section IV.A, this adds three municipalities (Quilicura, Cerro Navia and Puente Alto) or 2,094 students to the sample¹⁷.

To use the difference in differences technique we estimate a result for the base-line or diagnostic test in time zero (march 2015). To predict the test score we use the RC-TP result as the only independent variable¹⁸. The results of this strategy can be seen in the following Table 7:

Table 7: Difference in Differences, All Data

VARIABLES	1	2	3	4
	RCTP 1stgrade	RCTP 2ndgrade	Difdif 1stgrade	Difdif 2ndgrade
Treat Time			0.642** (0.237)	0.324* (0.165)
Treatment	0.459** (0.180)	0.129 (0.174)	-0.187 (0.179)	-0.187 (0.179)
Time			-0.235 (0.217)	-0.149 (0.148)
Gender(1=male)	-0.240*** (0.0692)	-0.304*** (0.0787)		
Constant	0.0541 (0.157)	0.176 (0.131)	0.151 (0.133)	0.151 (0.133)
Observations	979	776	2,264	2,011
R-squared	0.045	0.028	0.036	0.008

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The first to columns are just exploratory analysis and show the impact of the program assuming perfect randomization of treatment and control (i.e. same base line characteristics and test scores). If this were true, the program would have a significant effect of 0,46 S.D. over the first year and no significant effect on the second year. Also, the only observable variable we have for all the sample is the gender and it shows that being a boy has negative effect.

The difference in difference results of columns 3 and 4 also show a big and significant effect of the program after one year of being implemented (0,6 S.D.) and a smaller but also significant (at 10%) effect of the program after 2 years (0, 32 S.D.). The results of every sub-section are in the same line of the ones we found before and can be seen in Appendix 9.

B. Correlation between Cognitive (RC-TP) and Taste for Reading(TFR)

One possible mechanism by which the program may be having a positive effect is by improving children motivation towards reading in various subjects. For instance, the motivation score obtained in the TFR questionnaire could positively affect the reading comprehension results in the first and second year. It may be that the program has a huge cognitive effect after the first year and control students catch-up with treated during the second year, but motivation towards reading (anything) is still higher in treated students.

We first test this hypothesis by measuring the correlation between both scores and obtain that it is 0.2, that is lower than expected. We then try to measure the impact of the program using an interactive variable of the treatment and the TFR score which has a small positive effect but insignificant (see columns 3 and 4 from Table 9). Finally, we see that the motivational effect is higher during the second year which goes in line with our main hypothesis: the motivation effect is 0.3 S.D. during the first year and increases to 0.46 S.D. in the second year. This results cannot prove that motivation is the main mechanism by which the program is having an effect.

Table 9: OLS including RC-TP and TFR

	1	2	3	4
VARIABLES	RCTP	RCTP	RCTP	RCTP
	1stgrade	2ndgrade	1stgrade	2ndgrade
Treatment	0.593***	0.229	0.489**	0.258
	(0.105)	(0.141)	(0.218)	(0.439)
Treatment*TF			0.0486	0.00290

R			(0.0723)	(0.138)
Gender (1=male)	-0.192*** (0.0416)	-0.251*** (0.0513)	-0.158*** (0.0451)	-0.196*** (0.0504)
TFR (total score)			0.304*** (0.0446)	0.462*** (0.0815)
Constant			-0.875***	-1.153***
	-0.107* (0.0598)	0.0478 (0.0955)	(0.123)	(0.262)
Observations	2,273	1,692	1,915	1,425
R-squared	0.089	0.029	0.120	0.080

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

C. Falsification Exercise

To address issues regarding the school selection bias that could exist between treated and control schools a falsification exercise using PSM and then an OLS regression is done. Both tests use 2nd grade Simce test of 2015 to see if there is any difference between students from the same schools but one grade above the ones part of the treatment and control groups.

The following Table 10 shows the result of using nearest neighbour PSM methodology with Simce score as the dependent variable and income, parents education, books at home, class size and parents expectations as variables to estimate the PSM index. As one can see, there is no significant effect of being part of the treated group. Table 10: PSM Falsification Exercise, using 2nd grade Simce

Variable	Sample	Treated	Control	Difference	S.E.	T-stat
Simce 2nd grade	Unmatched	238.071	238.089	-0,18	2.5	-0.01
	ATT	238.071	236.893	1,17	4.3	0.27

Source: Own elaboration based on Simce test results.

Notes: Results are not are not standardized. The Stata command psmatch2 is used and propensity score and treatment effect are estimated simultaneously, so there is no need for further correction of the standard errors.

Table 11: OLS Falsification Exercise, using 2nd grade Simce

	-1	-2	-3
VARIABLES	No controls	Controls 1	Controls 2
Treatment	1.255	-1.253	-0.198
	-1.757	-2.191	-3.086
Income		1.63e-05***	7.13e-06
		(4.36e-06)	(5.19e-06)
Education Father			0.864*
			(0.458)
Education Mother		1.976***	1.592***
		(0.348)	(0.474)
Age			0.219
			(0.140)
Books		-0.0347	-0.0262
		(0.0333)	(0.0387)
Parents exp		-0.172	-0.514
		(0.577)	(0.777)
Parents satisfaction			1.537*
			(0.926)
Size			0.198
			(0.178)

Constant	232.6***	210.3***	185.0***
	-1.045	-4.180	(11.01)
Observations	3,311	1,973	1,434
R-squared	0.000	0.037	0.033

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The three columns above use an OLS regression where the only difference is the use of control variables, all specified above. Results show no significant effect of treated students over Simce test scores, therefore rejecting the selection hypothesis.

VIII. Cost-Effectiveness Analysis

To see what the cost-effectiveness ratio (C.E.) of the program is we follow Cabezas et al. (2011) and ask the following question: what is the necessary expenditure, per student, to obtain an impact of 0.1 S.D. after 1 year of the program?

The following assumptions are made:

1. It is assumed that the annual cost of Reading First per student is \$107,000 CHL pesos or 160 USD, based on cost information provided by the Growth with All Foundation (this is funded with SEP subsidy, which is approximately \$70,000 CHL pesos per student and month).
2. As an impact measure we will use the RC-TP, since cognitive measures are the most evaluated in the literature.

The results indicate that to improve in 0.1 S.D., Reading First Program would require an annual expenditure of \$27,000 CHL or \$40 USD per student after one year of being implemented and \$29,000 CHL or \$43 USD after two years (considering the Management of Language results). When evaluating this in terms of the SEP subsidy, it means only an approximate 4% of the annual grant.

Even though the cost-effectiveness calculations might not be comparable across program, it helps to illustrate their relative performance. Compared with other Chilean programs, this would place the program above other not-so effective programs such as the JEC (impact of 0.06 S.D. and cost-effectiveness ratio (C.E.) of \$ 635.7). Reading First would be between programs like P-900 (impact of 0.2 S.D. and C.E. \$91.2 USD) or SNED (impact of 0.16 S.D. and C.E. \$2 USD).

IX. Conclusion

The results of this paper demonstrate that low-cost reading programs can have a big impact after one year of implementation. It is important that any program that uses public funds, especially pre- school and primary school education, proves to be effective in terms of results. As mentioned before, reading comprehension has potential effects on quality of life, income, and overall productivity in the country.

The findings reveal that Reading First is being effective after 1 year and even though the effect is reduced during the second year it still has positive effects: the program has a positive impact of 0.4 S.D. on reading comprehension and text production (RC-TP) outcomes after one year and 0.37 S.D. in management of language (one section of the RC-TP test) after two years, and increases the likelihood that students will have favorable reading feelings.

There is a heterogeneous impact between treated municipalities that is due to different pedagogical management abilities between Arica and PAC: the program has a positive effect only in Arica after two years of implementation. This is a crucial factor for future decisions of the Foundation when considering to expand the program to any municipality. Also, it was not possible to prove that the impact of the program is strongly correlated to motivational factors.

Finally, it is important to understand other factors that could affect the performance of the program (school environment, innovation within the establishment, relationship with managers, salary structure, etc.). It would be particularly interesting to understand what kind of matching between tutor and teacher produces better results both in student learning and in motivational types and levels. To do this, we could extend the previous analysis to the 500 teachers who currently participate with the Reading First program (and not just the "treated" groups).

The Foundation aims to continue evaluations of the initial control group (San Joaquin) to see what happens after a year of program implementation (is there any remaining installed capacity?). It is planned measure the results of the Simce in fourth grade and so on to see the long-term effect of the program.

Bibliography

1. Adeyinka A. and O. Adedotun (2013), Teachers motivation on students performance in mathematics in government secondary schools, *International Journal of Humanities and Social Science Invention*, Vol.2, pp.35-41.
2. Akerlof G., and M.E. Kranton (2002), Identity and Schooling: Some Lessons for the Economics of Education, *Journal of Economic Literature*, Vol. XL , pp. 1167-12.
3. Bedregal P., V. Figueroa and M.Villalon (2008), "Alfabetización Inicial: Claves de acceso a la lectura y escritura desde los primeros meses de vida, Ediciones UC.
4. Bedregal P., P. González, M.Kotliarenko and D.Raczinski, (2007), "Chile: Lineamientos de Políticas para la Primera Infancia", UNICEF y BID.
5. Borman G.D., A.M. Chamberlain, B.Chambers, A. Cheung, Madden N. and R.E. Slavin, (2007), Final Reading Outcomes of the National Randomized Field Trial of Success for All, *American Educational Research Journal*, 44 (3), pp. 701-731.
6. Bravo D., "Competencias Básicas de Población Adulta en Chile", (2013), Centro de Microdatos, Universidad de Chile
7. Burley M., A. Freeman, E. Jones, J. Osborn, S. Rychener and R. Wilson (2007), Effect of Tutoring on Reading Achievement for Students with Cognitive Disabilities, Specific Learning Disabilities, and Students Receiving Title I Services, *Education and Training in Developmental Disabilities*, Vol (42), pp. 467-474.
8. Cabezas V., J. Cuesta and F. Gallego (2011), Effects of Short-Term Tutoring on Cognitive and Non-Cognitive Skills: Evidence from a Randomized Evaluation in Chile, J-PAL Working Paper.
9. Cunha F., J. Heckman and L.Lochner (2006), Interpreting the Evidence on Life Cycle Skill Formation, *Handbook of the Economics of Education*, Vol (1), pp. 697812.
10. Contreras D., G.Elacqua, M.Martinez, A.Miranda, (2016), "Bullying, identity and school performance: Evidence from Chile", *International Journal of Educational Development*
11. Contreras D., G.Elacqua, M.Martinez, A.Miranda, (2015), "Income Inequality of Performance Gap?", A Multilevel Study of School Violence in 52 Countries, *Journal of Adolescent Health*.
12. Contreras D., E.Thievos, (2014), Determinants of Psychomotor Development with Special Attention to Maternal Employment and Enrollment in Preschool during the First Three Years: Evidence from the Early Childhood Longitudinal Survey in Chile, Department of Economics, University of Chile

13. Contreras D. and Gonzalez S., (2013), Determinants of Early Child Development in Chile: Medical, Cognitive, and Demographic Factors, Department of Economics, University of Chile.
14. Coates C., A. Weisleder, B. Dreyer, M. Johnson, A. Seery, C.Canfield, J.Berkule, S.Mendelsohn, L.Alan, (2017) Early Reading Matters: Long-term Impacts of Shared Bookreading with In- fants and Toddlers on Language and Literacy Outcomes”, Unpublished, Pediatrics, NYU School of Medicine
15. Dowson M., C. Fernet, F. Guay, H. Marsh, C. Senecal (2008), The Work Tasks Motivation Scale for Teachers (WTMST), *Journal of Career Assessment*, Vol.16, pp.256-279.
16. Duflo E., R. Hanna (2005), Monitoring works: getting teachers to come to school, NBER Working Paper 11880.
17. Glewwe P., N. Ilias and M. Kremer (2010), Teacher incentives, *American Economic Journal: Applied Economics* 2, pp. 205-227.
18. Hart, B., y Risley, T., (2003), "The Early Catastrophe: The 30 Million Word Gap by Age 3", *American Educator*, 4-9.
19. Heckman, J., Pinto, R., Savelyev, P., (2012), "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Children", *The American Economic Review*, pp. 2052-2086.
20. Muralidharan K. and V.Sundararaman (2011), Teacher performance pay: experimental evidence from India, *The Journal of Political Economy*, Vol. 119, pp. 39-77.
21. OECD (2016), "Skills Matter: further results from the Survey of Adult Skills"
22. Patrinos H.A. (2016, June 30). Heres the evidence that low cost reading programs can have a big impact. [Web log post]. Retrieved from <http://blogs.worldbank.org/education/here-s-evidence-low-cost-reading-programs-can-have-big-impact>.
23. Roland F. (2011), Teacher incentives and student achievement: evidence from New York City public schools, *Journal of Labor Economics*, Vol. 31, pp.373-407.
24. Rosenbaum, P. R., Rubin, D. B. (1983), The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 4155.
25. Urzua, S. and G. Veramendi (2010), "The Impact of Out-of- Home Childcare Centers on Early Childhood Development", Northwestern University, for IDB Research Network on Early Childhood Development.
26. Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

X. Notes

¹Simce is the most important standardized test for school age students in Chile. These results are based on the question about learning standards (insufficient, elementary or adequate) in 2015 Simce for 4th grade students.

²Simce evolution in terms of learning standards shows no significant difference over the last 5 years.

³For more information on this kind of programs see: <http://iris.peabody.vanderbilt.edu/module/rti03-reading/cresource/q2/p03/>

⁴The evaluation of this surpasses the scope of this paper but will be possible in 2018 when 4th grade Simce results become available.

⁵Patrinos H.A. (2016, June 30). Here's the evidence that low cost reading programs can have a big impact [Web log post]. Retrieved from <http://blogs.worldbank.org/education/here-s-evidence-low-cost-reading-programs-can-have-big-impact>.

⁶SJ data is very similar to PAC, with lower poverty rates (6,7% and mean household income of 1.080.000 CHL pesos) and no significant differences in terms of educational outcomes or ethnic composition

⁷We used the individual level data-base of the 2nd grade Simce for the year 2015 (one grade above the treated sample). The variables used were: school size, average in Simce language, home books, parents education, income level, teacher expectations, teacher educational level and municipality.

⁸See Appendix 3 to see the take-up in each evaluation for the Control and control groups.

⁹When comparing parents that answered the questionnaire to parents of other evaluation such as 2nd grade Simce 2015 (one grade above treated students), parents who answered have higher levels of income and education.

¹⁰Randomization was not possible.

¹¹The test can be divided in 3 sections that will be analyzed as separate results: reading comprehension, text production and management of the language

¹²To see the completion of the assumptions of Conditional Independence and Common Support see Appendix 5.

¹³The results do not drastically change when using alternative methods such as the kernel-based

¹⁴To see the completion of the assumption of Parallel Trends Appendix 6.

¹⁵For more information on ordered logit regression see Wooldridge (2001), section 15.10.

¹⁶To see all the WTMST results see Appendix 12

¹⁷The selection of this sample was made using a matching procedure using the 2nd grade Since 2015 results. For more information on this see Appendix 2