Pontifícia Universidade Católica
do Rio de Janeiro

Vitor Azevedo Pereira

From early childhood to high school: Three essays on the economics of education

Tese de Doutorado

Thesis presented to the Programa de Pós-graduação em Economia of the Departamento de Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia.

Advisor: Prof. Claudio Abramovay Ferraz do Amaral

Rio de Janeiro

August 2016

Pontifícia Universidade Católica
do Rio de Janeiro

Vitor Azevedo Pereira

From early childhood to high school: Three essays on the economics of education

Thesis presented to the Programa de Pós-graduação em Economia of the Departamento de Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia. Approved by the undersigned Examination Committee.

Prof. Claudio Abramovay Ferraz do Amaral
Advisor
Departamento de Economia - PUC-Rio

Prof. Gustavo Maurício Gonzaga
Departamento de Economia - PUC-Rio

Prof. Miguel Nathan Foguel
Instituto de Pesquisa Econômica Aplicada – IPEA

Prof. Ricardo de Abreu Madeira
Faculdade de Economia, Administração e Contabilidade - FEA-USP

Prof. Naércio Aquino Menezes Filho
Insper Instituto de Ensino e Pesquisa-INSPER

Prof. Mônica Herz
Coordenadora Setorial do Centro de Ciências Sociais-PUC-Rio

Rio de Janeiro, August 23rd 2016

Vitor Azevedo Pereira

Vitor Azevedo Pereira graduated in Economics from PUC-Rio in 2005. He received his Master's degree in Economics from PUC-Rio in 2007 and from University of Maryland, College Park in 2011. During his PhD, Vitor was a visiting student researcher at the Lemann Center at Stanford University.

Bibliographic data

Pereira, Vitor Azevedo

From Early Childhood to High School: Three Essays on the Economics of Education/ Vitor Azevedo Pereira; Advisor: Claudio Abramovay Ferraz do Amaral -2016

278 f.: il; 30 cm

Tese (Doutorado em Economia) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2016.

Inclui Bibliografia

1.Economia – Tese. 2. Primeira infância; 3- Incentivos financeiros 4- Aversão à perda. I. do Amaral, Claudio Abramovay Ferraz. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. III. Título.

CDD: 330

For Teca and Beatriz.

Acknowledgments

This thesis would not be possible without the continuous support and encouragement from my adviser Claudio Ferraz. He has shaped me as a researcher, taught me how to choose interesting research topics, how to approach interesting questions, how to be always rigorous, how to write well a paper, among so many things. His support many times exceeded a lot what was expected from an adviser, and I am deeply grateful for all I learned from him.

I am also deeply thankful to my co-adviser in chapter 2, prof Eric Bettinger, who found time he did not have during early mornings to receive me in Stanford every week and provided key insights on the paper.

I am grateful to all the members of my committee: Gustavo Gonzaga, Ricardo Madeira, Naércio Menezes and Miguel Foguel, who read carefully each line of the papers. Their insightful suggestions helped me to improve the papers to another level of quality.

I am deeply indebted to Sebastian Martinez e Sophie Naudeau for giving me the opportunity to work with them in Mozambique, first as a field coordinator, and them as their co-author. I am also thankful to thank Barbara Bruns for presenting me to the world of economics of education, and for her many advices.

I could not forget to mention all my professors which helped me have a strong and solid formation in economics during my undergrad, masters and PhD, especially professors Juliano Assunção, Rodrigo Soares, Judi Hellerstein, Mark Duggan and Gustavo Gonzaga, who reminded during the defense has been following my career since the beginning of my undergrad studies. It was a great pleasure to have him my committee.

Maria das Graças Silva has been my better angel during all these years, and her help has been more valuable as far as I have been, as in Salvador, Menlo Park and Brasília. Flavia Manfrim has offered outstanding support when I need to sign the agreement between PUC-Rio and the SEASDH to have access to the data for the second paper. Cris Antunes has made a superb job in helping me join the Lemann Center in

Stanford in a very short time. Finally, I am grateful for the support of all the staff from the Department of Economics during all these years in PUC-Rio.

These three papers would not be possible without the collaboration from many people who run the programs studied here, or those responsible for their administrative data. My thanks to Ana Tenório, David Wright, Damião Mungói, Domingos Manhangue, Patricia Cavagnis e Melissa Kelly, Margareth Zaponi, Roberta Mary, Epifânia Valença, Madalena dos Santos (who does not run the program in Pernambuco but who helped me a lot in Recife), Antônio Claret, Lucia Modesto, Ana Vieira, Luis Gustavo Martins, Enzo Tessarollo, Júlia Modesto, Rodrigo Capeans, Wilson Risolia, Daniela Ribeiro, Vânia Mendonça and Rosi Figueiredo.

I am thankful to John Bunge for coordinating the follow up data collection in Mozambique, and for the research assistance from Michele Perez for cleaning Mozambique's data and for helping me reach the first findings on the paper.

I am indebted to many comments and suggestions from seminar and workshop participants at PUC-Rio, Stanford, USC-Dornfield, Universidade Nova de Lisboa, Itaú Social and SBE, especially to Cecília Machado, Peter Serneels, Sergio Urzua, Francisco Ferreira, Leandro Carvalho, Silvia Helena Barcellos, Juan Esteban Saavedra, Susanna Loeb, Martin Carnoy and Hessel Oosterbeek. Eric Taylor gave me valuable insights on how to deal with big data and merge large administrative data sets.

I cannot forget to mention my graduate colleagues, who helped me so much during the PhD. My thanks to my UMD friends with whom I spend great moments and who helped me learn so much: Jeronimo Carballo, Pablo Cuba, Filippos Petroulakis, Alvaro Pedraza and Ben Zou. My thanks to Rafael Dix Carneiro, Denise Kassab and Nelson Camanho for the companionship in DC, and to Romero Rocha and Marcio Firmo for the companionship in Rio. I would also like to thank my colleagues from PUC: Arthur Bragança, Bruno Ottoni, Joana Costa, Amanda Schutze, Raphael Ornellas and Pedro Forquesato. and my friends in Stanford: Luana Marotta, Danilo Dalmon, Felipe Michel and Joyce Toyota.

This thesis would not be possible without the financial support from CNPq, CAPES and the Lemann Center at Stanford. Data collection on Mozambique was financed by grants from 3ie, BNPP and SIEF.

My fellow brother Gabriel Buchmann is proud in heaven reading this thesis. I am deeply thankful for all I learned with him. Vamos com tudo!

Finally, my thanks to my parents Vagner and Luci, who always supported my choice to pursue a PhD, even without understanding very well why that was so important to me, to my brother Vinicius and to my grandma Dalva.

Beatriz has changed my life. Thanks to make me such a happier person every day. Teca knows so well my strengths, my fears, my weakness, and who has contributed so much to this thesis. It would not be possible to write this thesis without her.

Abstract

Pereira, Vitor Azevedo; do Amaral, Claudio Abramovay Ferraz (Advisor). **From early childhood to high school: Three essays on the economics of education.** Rio de Janeiro, 2016. 278 p. Tese de Doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

In the first chapter, we analyze the impact of a randomized preschool program on children's development and schooling in rural Mozambique. Children who attend preschool experience gains in cognitive development and socio-emotional skills, and are also more likely to be enrolled in primary school, at the appropriate age. The preschool intervention also had positive spillovers on the schooling of older siblings and labor supply of caregivers. These results suggest that community led preschools are a promising policy option for helping children meet their development potential. In the second chapter, I study the impacts of an innovative attainment award targeted to disadvantaged secondary students. The payment is made through students' bank accounts, and the full amount can only be withdrawn upon timely high school graduation. By exploiting the phased in expansion of the award policy, I find that award eligibility substantially decreases dropout and increases test scores and high school completion. In the third chapter, we analyze an empirical case of loss aversion in public policy. While previous papers have identified evidence of loss aversion in laboratory experiments or in sports, it is still unclear whether these findings could generalize to other domains. We test for loss aversion by analyzing teacher reactions to receiving a bonus based on a continuous underlying measure of school performance. Consistent with loss aversion, we find sizable improvements on student scores at schools that barely fail to receive the bonus. We investigate the mechanisms behind these results and we find significant changes in teachers' pedagogical practices.

Keywords

Early childhood; financial incentives; loss aversion.

Resumo

Pereira, Vitor Azevedo; do Amaral, Claudio Abramovay Ferraz (Orientador). **Da primeira infância ao ensino médio: três ensaios sobre a economia da educação.** Rio de Janeiro, 2016. 278 p. Tese de Doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

No primeiro capítulo, analisamos o impacto de um programa pré-escolar, aleatorizado entre comunidades rurais moçambicanas, sobre o desenvolvimento infantil. Crianças que foram à pré-escola tiveram ganhos no desenvolvimento cognitivo e sócio-emocional, e maiores chances de estarem na escola primária, na série adequada à sua idade. A intervenção também teve efeitos sobre a escolaridade de irmãos mais velhos e sobre a oferta de trabalho de seus cuidadores. No segundo capítulo, estudo os impactos de uma inovadora política de bônus estudantil focalizada em estudantes secundários vulneráveis. O pagamento é feito através de uma conta poupança, em nome do estudante, e o total da conta só pode ser sacado após a conclusão no ensino médio. Ao explorar a expansão gradual do programa, encontro que a elegibilidade ao bônus diminui substancialmente a evasão escolar e aumenta as notas dos estudantes. No terceiro capítulo, analisamos um caso empírico de aversão à perda em políticas públicas. Enquanto outros artigos identificaram evidências de aversão á perda em experimentos de laboratório ou em esportes, ainda não é claro se tais achados podem ser generalizados para outras áreas. Nós testamos a aversão à perda ao analisar a reação de professores ao receber um bônus de desempenho baseado em uma medida contínua de performance escolar. Consistente com a aversão à perda, encontramos grandes melhoras nas notas de estudantes de escolas que perdem o bônus por muito pouco. Investigamos os possíveis mecanismos por trás desses resultados e encontramos significantes mudanças nas práticas pedagógicas de professores.

Palavras-chave

Primeira infância; incentivos financeiros; aversão à perda.

Sumário

Para ser grande, sê inteiro: nada
Teu exagera ou exclui.

Sê todo em cada coisa. Põe quanto és
No mínimo que fazer.

Assim em cada lago a lua toda
Brilha, porque alta vive.

Ricardo Reis, heterônimo de Fernando Pessoa, in "Odes"

1
Preschool and child development under extreme poverty: Evidence from a randomized experiment in Mozambique

1.1. Introduction

It is well known that the foundations for a healthy and productive future are formed in the early years of a child's life. Yet in Sub-Saharan Africa an estimated 61% of children fail to meet their development potential because of poor health and poverty (Grantham McGregor et al, 2007). Inadequate health and nutrition, cultural practices that limit communication between parents and children, and home environments with few books, toys, and other learning opportunities may all contribute towards inadequate cognitive growth and overall child development. These early deficits can have life-long consequences, including lower levels of school participation and performance, lower future earnings and income, increased reliance on the health care system and higher rates of criminality (Walker et al., 2001; Naudeau et al., 2010). The effects of poor development in the early years can thus be deleterious and long lasting, reinforcing the intergenerational transmission of poverty and constraining economic development.

In developed countries, investments in early childhood development (ECD) have been shown to be cost-effective and to have a higher rate of return than investments later in life (Heckman, 2008; Heckman, Stixrud, and Urzua, 2006). In fact, evidence in the United States suggests a potential rate of return of 7-10 percent annually from high quality ECD interventions targeting vulnerable groups (Heckman et al., 2010; Rolnick and Grunewald, 2007), while a model of the potential long-term economic benefits of increasing preschool enrollment to 25% or 50% in every low-income and middle-income country showed high benefit to cost-ratios ranging from 6.4 to 17.6 (Engle et al., 2011). Furthermore, investments during early childhood are highly complementary

with investments made later in life, making future investments more efficient and yielding significant benefits to both individuals and society (Engle et al., 2007).

In the short to medium term, various types of ECD interventions have been shown to enhance school readiness and related educational outcomes, improve physical and mental health, and reduce engagement in high-risk behaviors (Barnett, 2011; Nores and Barnett, 2010; Engle et al., 2007). Significant long-term labor market returns of an early stimulation intervention have also been documented in Jamaica (Gertler et al., 2014)

Positive effects of pre-primary programs on schooling and child development have been demonstrated in several Latin American countries, including Argentina (Berlinski and Galiani, 2007; Berlinski et al, 2009), Bolivia (Behrman et al, 2004), Colombia (Bernal and Fernández, 2013; Attanasio et al. 2013), and Uruguay (Berlinski et al, 2008). On the other hand, an experimental study of preschool interventions in Cambodia finds no positive effects on child development, citing concerns over program implementation and quality (Bouguen et al., 2014). Thus, outside of the Latin American context, there is scarce evidence on the effectiveness of preschool and its viability as a cost-effective model for improving child development outcomes in low income countries[1].

As such, the potential benefits as well as the costs and feasibility of investing in preschool in low-income settings remain largely open questions. This study helps fill this gap by conducting what, to our knowledge, is the first randomized experiment of preschool in an African context. We analyze the effects of a community based preschool intervention randomly assigned to 30 out of 76 eligible rural communities in the Gaza province of Mozambique. At a cost of USD$ 3.09 per child per month, the program provided up to three preschool classrooms per community, community mobilization activities, learning materials, instructor training, and monthly parenting meetings. We collected a baseline survey of 2000 households with preschool aged children just prior to the roll-out of the program in 2008. We additionally measured

[1] Existing preschool evaluations in Latin America have relied exclusively on quasi-experimental identification strategies (Leroy et al, 2012). Other preschool studies compare participating and non-participating children without establishing causal attribution, for example Mwaura et al., 2008 for Kenya, Uganda, and Zanzibar, and Rao et al., 2012 for Cambodia.

primary school performance on a sample of first graders and interviewed local community leaders. We conducted a follow-up survey two years after the start of the intervention, following the original sample of children plus cross sections of first graders and community leaders in the 76 evaluation communities.

We measure the effects of preschool on key dimensions of children's development and school readiness, including the cognitive (numeracy, working memory), linguistic (receptive language, use of gestures, sounds and movements), psycho-social and behavioral (personal and social) and physical (fine and gross motor skills, health and nutrition) domains. We then analyze the effects of preschool attendance on subsequent primary school enrollment. To explore causal pathways in more depth, we analyze the effects of the program on parenting practices and knowledge. Additionally, given that preschools may free up caregiver time at home, we explore potential spill-over effects of the program on other household members.

We find that children who attend preschool demonstrate large and significant improvements in cognitive and problem-solving abilities, communication and receptive vocabulary, fine-motor skills and socio-emotional and behavioral outcomes. As such, children are better prepared for primary school and outperform their peers in the control group on these dimensions. The preschool program has a large impact on transition to primary school. Children who attended preschool are 21.2 percentage points more likely to be enrolled in primary school and spend an average of 5.86 additional hours per week on schooling and homework related activities. Preschoolers are also significantly more likely to enroll in first grade by age 6, an important achievement in the Mozambican context where many children enroll in primary school well after the appropriate age.

In addition to effects on preschoolers, we find evidence of modest but positive spillovers from the program on school enrollment of older children and labor force participation of caregivers. Children 10 to 14 years old are 2.8 percentage points more likely to have attended school and report spending fewer hours caring for younger children. Adult caregivers are 3.7 percentage points more likely to report working in the past 30 days (effect significant at the 10% level). These results suggest that center based models, where children are cared for out of the home, may produce added

benefits compared to alternative home-based models by freeing up time for other household members. Taken together, these results suggest that low-cost preschool interventions such as the one studied here are a promising policy alternative for promoting child development and school enrollment in some of the most resource deprived parts of the world.

The remainder of the paper is structured as follows. The next section describes the country context and preschool program. Section 3 presents the randomized evaluation design and sections 4 and 5 describe the data and identification strategy, respectively. In section 6 we present the main results, explore causal pathways and discuss validity checks. Section 7 concludes.

1.2. Context and Intervention

Mozambique is a Sub-Saharan African country with a population of approximately 25 million people. Soon after gaining independence from Portugal in 1975 the county descended into a civil war that lasted from 1977 to 1992. By the end of the civil war, Mozambique was one of the poorest countries in the world. As of the early 1990s, Mozambique embraced a new constitution enshrining a market oriented economy and a multiparty political system with free elections. By 2010 the country was one of the fastest growing economies of the world, though GNI per capita was still just $460 USD[2]. Among the list of countries surveyed by the United Nations Development Program's Human Development Report of 2013, Mozambique had the third lowest Human Development Index in the World, only higher than Niger and the Democratic Republic of the Congo. Poverty levels in rural areas are particularly severe, with over 57% of the population living below the official poverty line. At the same time, Mozambique has made substantial progress in expanding primary education, with net primary school enrollment rates increasing from 45% in 1998 to 95.5% by 2010. Yet children tend to enroll in school past the appropriate age, and many times struggle

[2] GDP growth has averaged around 7% per year in the period 2005-2015 (http://data.worldbank.org/country/mozambique).

with the new social and academic environment, resulting in increased repetition and low academic performance.

The Southern part of the country, including the Gaza province where this study is conducted, has higher incomes and tends to have better human development outcomes compared to other provinces in the central and northern areas of Mozambique. However, the province was also one of the most severely affected by the country's civil war. Partially due to high migration rates of adult males for employment in the mining sector, the province also has the highest rate of young adults living with HIV-AIDS (Crush et al, 2010). One quarter of the Gaza population aged 15 to 49 is HIV positive, while the national rate is estimated to be around 11% (Instituto Nacional de Saúde et al. 2010). Women are more severely hit by the HIV epidemic, with a prevalence rate of 30% in Gaza province.

While population based child development indicators in Mozambique are scarce, the 2011 Demographic and Health Survey (DHS) reports child growth in terms of height and weight[3]. Nationally, 42.6% of children under age five are stunted, 14.9% are underweight and 5.9% are wasted. The conditions in Gaza province are relatively better off, with 26.8%, 6.3% and 1.0% respectively[4] (Ministério da Saúde, 2013). However, in our sample of 2000 children ages 3 to 5 from 76 poor rural communities, pre-program growth indicators show that 42% of children were stunted, and 5% of children wasted, suggesting that our study population is closer to the national average. Nationally representative statistics on child development outcomes other than growth were not available at the time of this study. However, standardized child screening tests from our baseline survey showed that around half the children in the sample were at risk of delays in fine motor skills and problem resolution, and more than 20% scored below age-appropriate levels in communication (Bruns et al, 2010) [5].

Baseline descriptive statistics of the study population are presented in Table 2. Most children in the sample live in poor households, dependent on subsistence

[3] The positive association between children's linear growth and cognitive development is well documented (Sudfeld et al, 2015).
[4] The DHS also showed that nationally, 91.7% of newborns were breastfed, and the median duration of exclusive breastfeeding was 3.5 months, well below the recommended 6 months. The situation in Gaza is comparable.
[5] With no country specific reference group, we compare to available reference populations outside the Africa Region.

agriculture or informal employment, and ten percent of children in the sample were orphaned, primarily single parent. Caregivers have only 3.3 years of education on average, 38% can't read or write, and 51% of caregivers do not speak Portuguese (the country's official and primary language for education). Furthermore, cultural practices and norms may play a role in limiting caregiver interactions with children that promote cognitive development and school readiness. At baseline, only about half of caregivers report reading, drawing objects or playing games with children at home. Under these circumstances it is not surprising that children frequently experience delayed entry into school and are not prepared for the new academic and social environment in primary school.

While there are a multiplicity of demand side constraints that may limit investments in human capital of young children, in 2008 the supply of education services for pre-primary aged children was virtually inexistent in rural Mozambique[6]. Preschools were concentrated in urban and more affluent areas, and available national estimates put the proportion of children enrolled in preschool at the time of the intervention at about 4%. Therefore, while children under 2 years are eligible for growth monitoring services through the public health system, and children 6 years and older can enroll in primary school, the majority of children in the pre-primary age range of 3-5 years remained without access to age-appropriate child development services and thus at a higher risk of deviating from normal development pathways undetected.

To help address the lack of child development services for preschool aged children in rural Mozambique, the non-governmental organization Save the Children began implementing a community based preschool program in three districts of the Gaza Province starting in 2008.[7] The program focused on early stimulation, emergent literacy and numeracy instruction and psychosocial support, with the objectives of improving children's cognitive, social, emotional, and physical development, and facilitating transition to primary school. In addition, the program organized monthly parent group meetings to strengthen positive parenting practices in the home.

[6] Starting in 2014, a national preschool program rolled out to 84,000 children in 800 communities.
[7] The preschool model was initially piloted in 12 communities starting in 2005. Based on this initial experience and having obtained additional financial resources, the model was scaled up to 30 new communities in early 2008.

Funding constraints limited the intervention to a maximum of 30 communities. Save the Children provided the seed capital and technical assistance to build and equip the preschools, train instructors and implement a standardized curriculum. Each preschool was built with one to three classrooms, washrooms and a playground.[8] Communities donated land, labor and locally available construction materials, and appointed a 10-member committee to manage and supervise preschool activities.[9] Given the scarcity of qualified instructors in the area, preschool teachers were not formally trained educators. Instead, most instructors were women recruited from within communities, provided basic training and supervision by Save the Children, and paid a nominal fee of $10US per month. Instructors were more educated than the average caregiver in our sample (6.1 years of education compared to 3.4) and many had children of their own enrolled in the preschool.

Preschools operated 5 days a week for 3 hours and 15 minutes per day, following a structured daily routine designed to stimulate child development through play and learning activities[10]. Each classroom held up to 35 children and was staffed by two instructors. While enrollment was limited to children between the ages of 3 and 5 years, classrooms were mixed by age and gender to promote peer-to-peer interaction. The primary language of instruction was Changana, the local vernacular, and the curriculum gradually introduced Portuguese into learning activities to help children prepare for primary school.

To complement classroom activities with additional stimulation and caregiving practices at home, parents and caregivers of enrolled children committed to participating in monthly meetings to discuss child development topics such as health, nutrition, and literacy[11]. Meetings were facilitated by Save the Children staff with

[8] A total of 67 classrooms were financed through the program. Physical requirements included 1.2 to 1.5 meters of space per child, adequate ventilation and light, and clean and dry floor surfaces. Classrooms were built using both traditional and conventional building materials and were typically built as single standing rooms with cement floors, wood or straw walls and thatched or tin roofs.

[9] Preschool management committees were appointed by the community. Each committee was composed of a president, secretary and treasurer, and was assisted by other community members responsible for mobilizing the community to assist with construction, participate in caregiver meetings, and conduct preschool maintenance activities such as cleaning and providing safe water.

[10] see Online Appendix table 1 for the daily schedule of activities.

[11] Meetings were open to anyone in the community. Parents of preschoolers were meant to attend meetings as a condition to enrolling their children in the program, though attendance was not strictly enforced in practice.

assistance from preschool instructors and community health workers, following an appreciative inquiry approach in which knowledge is built from existing positive parenting practices and harmful practices are brought to light and modified with strategies such as the use of positive deviants to model new behaviors.

The program was implemented in three districts of Gaza Province (Manjacaze, Xai Xai and Bilene) where Save the Children had an established operational presence. Program requirements narrowed eligibility to rural communities with between 500 and 8000 residents, located within operational areas of sufficient geographic proximity so program field teams could travel between assigned communities in a single day.[12] As a pre-condition for funding, communities committed to the contribution of land, materials and labor. The intervention rolled out in 2008 with the formation of preschool committees, recruitment and training of teachers and construction of classrooms. Some communities initiated teaching activities prior to completion of the physical infrastructure, meeting outdoors or in other community structures while the preschool facilities were built.

Enrollment was restricted to children between three to five years old with residence in the community, though age and residency requirements were difficult to monitor and were not always strictly enforced. Enrollment in the preschool program was voluntary and participating parents were encouraged to send their children to preschool daily, to attend the monthly parenting meetings, and help with preschool maintenance activities such as supplying clean drinking water and cleaning. Monetary contributions from families were minimal, as the program covered the costs of instructor stipends and basic materials during the first two years of operations. While in principle the program was targeted to the poorest and most vulnerable children in each community, in practice we observe that children who enroll in preschool are more likely to speak Portuguese, score higher on some child development indicators, and

[12] The population criteria were established to reach enough children for at least one classroom at the lower limit, while enabling community mobilization in larger communities. The program financed 5 field teams, each of which was responsible for implementing the program in 6 communities. To reduce travel time and costs, each group of 6 treatment communities needed to be located within sufficient proximity so that a field team could travel between its assigned communities within the same day. Operational areas corresponded closely with administrative posts, the intermediate administrative unit between district and community.

tend to have more favorable nutritional indicators. We also find differences in caregiver characteristics and behaviors in our baseline survey, with parents of enrolled children more likely to speak Portuguese, read and write, and report pro-active parenting practices such as playing games with the child[13].

1.3. Experimental Design and Sample

As discussed above, funding constraints limited the project to a total of 30 intervention communities, assigned at random within the pool of eligible communities in the intervention districts. Following operational requirements set by the program, the random assignment protocol was established in the following steps. First, the program compiled a list of all eligible communities in the three districts[14] and identified "operational areas" based on the geographic proximity and access between communities. To maximize the number of eligible communities in the sample, the five operational areas with the largest number of communities were selected, and each area was assigned six treatment communities. Within each operational area, communities were stratified by population size, forming blocks of two or three communities. For each block, one treatment community wandom, resulting in 30 treatment communities and 46 control communities. Six of the original 46 control communities turned out to be neighborhoods in treatment communities where a preschool was built, and were re-classified to their treatment counterparts in our analysis.[15] We test for and confirm that our main results are robust to the ex-post re-classification of these six control communities[16].

The household survey collected data on a random sample of 2000 households with preschool age children in the 76 evaluation communities. In the absence of a

[13] See Online Appendix Table 7.

[14] The list of communities was based on the best available information at the time of the survey. With no official roster of communities in the three districts, a consultant was hired by the program to update existing lists based on interviews with officials at the administrative post level.

[15] The sample selection, randomization procedure and the ex-post corrections of neighborhoods is discussed in detail in online appendix section 3. Re-classification was possible thanks to precise geo-location data collected on the original sample of 76 treatment and control communities in 2014.

[16] Results are presented in Online Appendix Table 28.

household level sample frame, we conducted a door to door census to identify households with at least one child in the preschool-eligible age range of 36 to 59 months. We then drew a random sample of 23 households with eligible children per community. In addition, in each of the 4 largest treatment communities where oversubscription to the program was likely,[17] an additional 63 households were selected, yielding a total sample of 2,000 households. In our analysis we re-weight the data with community level population sample weights equal to the inverse probability of selection, though results are generally robust to the exclusion of weights as would be expected from block randomization based on community population size.[18]

In each sampled household we collected a detailed baseline survey including a battery of child development tests and anthropometric measurements for one preschool aged child per household, identified as the "target child." In households with more than one preschool aged child, the youngest child was selected. We additionally interviewed the target child's primary caregiver and the head of household to collect demographic and socio-economic information about parents, caregivers and other household members. In addition, in each community we conducted a community leader survey and a primary school survey, interviewing school principals and first grade teachers to collect school performance indicators on a sample of 1st graders.

Baseline characteristics are balanced between treatment and control communities, with no significant differences for community infrastructure (online appendix Table 6), most child characteristics (sex, age, language, orphan, health status or anthropometrics), child development tests (described in detail in section 4) or caregiver and household characteristics (Table 2).

In 2010, approximately two years after the start of the program, we conducted a follow-up survey on the panel of target children and their households as well as the current community leader and a cross-section of 1st graders in the same primary schools

[17] Individual level randomization was initially proposed for communities with oversubscription, though ultimately this was not systematically implemented and was abandoned as an evaluation strategy. Nevertheless, oversubscription did occur in several larger communities.
[18] The modified Breusch-Pagan test suggested by Solon et al (2013) indicates the presence of heteroskedasticity associated with cluster sizes. In this case weighting, can improve the precision of the estimates. Unweighted estimates are presented in online appendix Table 28.

interviewed at baseline. We also visited the preschools in treatment communities to collect information on the status of the program's operation. An intensive tracking effort was made to locate the target child and minimize sample attrition. If the child had moved from his or her original place of residence we attempted to interview the child (and their current household) so long as he or she maintained residence in Gaza Province (including outside the three intervention districts) or had moved to the capital city, Maputo. Overall, we successfully located 94.9% of the baseline sample, for an average attrition of approximately 2.5% per year and with no differential attrition between treatment and control (94.8% re-contact in treatment, 94.9% re-contact in control).

Figure 1 shows preschool enrollment over time as reported by primary caregivers for children ages 3 to 9 in 2010, in the treatment and control groups. We observe that prior to 2007 preschool enrollment was virtually non-existent for children in both groups. There is a slight increase in preschool enrollment in treatment communities in 2007, though still under 4%.[19] Starting in 2008 when the program is fully operational, we observe a sharp increase in enrollment amongst children in treatment communities, with 29% of children enrolled by January 2010. We also observe a slight positive trend in preschool attendance in control communities in the period between 2008 and 2010, though total enrollment rates never surpass 6%. We attribute this primarily to the construction of 6 government and church preschools in control communities over this period, but cannot rule out some contamination from program preschools.

To confirm program effects on preschool participation we disaggregate enrollment by age (online appendix Table 8). We find significant differences in enrollment rates for children in the cohort of 3 to 7 year old children who were eligible for program participation in the period between 2008 and 2010, but no differences for children 8 to 11 who were at least 6 years old in 2008 and thus too old to enroll in preschool. For the sample of target children (3 to 5 years old at baseline), enrollment

[19] The baseline survey was timed prior to the construction of any preschool classrooms, however some communities had already started the community mobilization process and had recently begun operating preschools in outdoor or temporary spaces at the time of the baseline survey. Some of the reported preschool participation in the pre-program period may also be attributed to recall bias. However, it is likely that some children in treatment communities had already been enrolled when the baseline survey took place. Given the very short exposure to treatment on this group of children, we do not expect this would significantly alter longer term measures of child development collected at baseline.

in treatment communities was 41% compared to 8% in control, resulting in a treatment effect of 33 percentage points in preschool enrollment. Amongst children enrolled in preschool, on average children attend 4.9 days a week, for a total of 3.7 hours per day. Average travel time from home to the preschool is 20 minutes and average reported fees are 5 meticals ($0.16 USD) in the treatment group and 23 meticals ($0.76 USD) per month in the control group.

We also asked caregivers of children that did not enroll in preschool about access to a preschool in their area. Approximately 77% of households in treatment communities report having access to preschool compared to 27% in control communities. This result suggests that about a quarter of households in the treatment communities were either unaware of the preschools in their community or viewed them as being too far or otherwise inaccessible. When analyzing the primary reason given for not enrolling their preschool-aged child in preschool, the three most common reasons were that the distance to the preschool was too great, that the child was too young (suggesting misinformation of eligibility rules or a perception that younger children are better off staying home) and that preschool was too expensive. 5% of non-participating households in treatment areas reported applying to the preschool but were not accepted, while 13% applied but were not accepted in control areas.

We attribute this to oversubscription in some treatment communities, where total demand exceeded the number of spots. Children who were not accepted into preschools in control communities may have attempted to enroll in Save the Children financed preschools in neighboring (treatment) communities, but were not granted admissions based on the community residency requirements established by the program.

1.4. Measuring Child Development

The primary objective of the intervention is to improve children's development along the domains of cognitive, social, emotional, and physical development, thus facilitating transition to primary school. We apply a set of standardized tests to measure cognitive ability (including problem-solving skills, memory, and early math skills),

gross motor skills (e.g., running, jumping), fine motor skills (e.g., picking up objects, holding a pencil), language and communication (e.g., production and understanding of words, ability to identify letters), and socio-emotional development (e.g., getting along with peers and adults, following directions and cooperating, capacity to regulate emotions positively in stressful situations). We also collect children's anthropometric measurements (height and weight) and caregiver-reported morbidity.

The specific child development tests are based on adapted versions of: (i) the "Ages & Stages Questionnaires®" (ASQ), (ii) the "Teste de Vocabulario por Imagens Peabody" (TVIP);[20] and (iii) the Early Development Instrument (EDI). All tests were applied at baseline[21] and again at endline,[22] using age specific versions of the tests when appropriate. The adapted versions of the ASQ and TVIP were collected on the panel of target children. The adapted version of the EDI is collected on a repeated cross section of a random sample of 20 first graders in primary schools in treatment and control communities.

The ASQ is a child monitoring system used to assess whether children have reached certain developmental milestones across the domains of language, cognitive, gross motor, fine motor, and socio-emotional development. For this study, the questionnaire was translated into Portuguese and adapted for the local context. The adapted version of the ASQ was administered in Changana.[23] Some questions were asked directly to the target child, while other questions involving child behaviors that are difficult to observe in the context of a household visit were asked to the child's mother or caregiver. Each domain includes a series of individual questions, and is scored based on the ability of the child to perform the task in question. Scores for each domain are aggregated to form a total score and sub-score by domain.

The TVIP is a test of receptive language applied to all target children in the sample. The TVIP was originally adapted and normalized for Spanish speaking

[20] The TVIP is an adaptation of the Peabody Picture Vocabulary Test (PPVT)
[21] See Naudeau, Martinez, Premand, & Filmer (2011) for a detailed review and discussion of TVIP findings at baseline.
[22] A fourth test, the Strengths and Difficulties Questionnaire (SDQ) was added to the endline survey. We do not present the results due to a coding error present in the data which impaired the authors' ability to carry out meaningful and reliable analysis.
[23] Changana is a vernacular language. Therefore, it was important to have a standardized written version in Portuguese before a common Changana translation could be agreed upon by all surveyors (who spoke both Changana and Portuguese but not English).

populations in low-income settings and has been widely used in Latin America. In the test, the child is shown a series of 4 pictures or items at a time (e.g., fork, table, dog, doll). The surveyor asks the child to point to one of the pictures (the doll, for example) and then records whether the child pointed to the correct picture. The test stops when the child makes 6 errors within 8 consecutive responses. For this study, the TVIP was translated into both Portuguese and Changana, and some items adapted to fit the local context. All target children were given the test in both languages, with Portuguese being administered first.

In addition to raw TVIP scores, we analyze standardized scores using age specific norms published by the test developers. The norms take as reference a sample of 1219 Mexican children and 1488 Puerto Rican children (Dunn et al, 1986). According to those norms, the age specific mean is always 100, and one standard deviation is 15, such that a score of 70 is two standard deviations from the mean of the reference population of Mexican and Puerto Rican Children. As observed in Figure 4, as of 60 months the mean TVIP score falls well below the 70-point mark. As a point of reference, children aged 66 months in our sample perform on average close to the 25% of poorest children from the sample of young and poor Ecuadorian families in a study by Paxon and Schady (2007).

The Early Development Instrument (Janus & Offord, 2007) is completed by a first grade primary school teacher[24] who reports information on a random sample of 20 first graders enrolled in his or her class.[25] While potential biases in teachers' reporting (on the basis of socio-economic background, for example) can be a legitimate concern, the reliability and validity results of studies conducted with the EDI in diverse areas of

[24] In each school the survey team interviewed the principal and administered the EDI with one first grade teacher. In schools with more than one first grade teacher, the survey field supervisor selected one first grade teacher randomly. Once the teacher was selected, the supervisor randomly selected 20 first graders using a random table. The supervisor then filled in 3 questionnaires (i.e., for the first 3 first grade students) with the teacher to familiarize the teacher with the instrument. The supervisor left the 17 remaining questionnaires with the teacher, for him/her to complete, and returned within 2 weeks to collect all completed surveys.

[25] For the EDI we observe only the subset of children who enroll and are attending primary school. Given that the preschool program had a large and significant effect on primary school enrollment in treatment communities (section 6.1), it is likely that the composition of first graders in treatment communities changed relative to controls. If the program led otherwise lower-performing or more disadvantaged children to enroll in primary school, then the results of the EDI reported here are likely lower-bound estimates of impact (given that the "lower-performing" counterparts in control communities are simply not enrolled in primary school).

Canada and in British Columbia (where a potential racial bias towards Aboriginal children was considered possible) dispute this contention (see a summary of these studies in Janus et al., 2007). For the purposes of this study, the EDI was translated into Portuguese, and some of the items were dropped or adapted to fit the local context.

1.5. Estimation Strategy

We first present intention to treat (ITT) estimates of the effect of offering preschools in treatment communities. Given that about 41% of eligible target children enroll in preschool, the ITT estimate is a weighted average effect of enrolled and unenrolled children in treatment areas. The ITT estimate represents the relevant treatment effect from the point of view of the policy maker interested in replicating this intervention model in similar contexts, where some parents will not enroll their children based on personal preferences, information, capacity constraints or other reasons. The basic regression model for the ITT estimator is:

$$Y_{ijt} = \alpha + \beta_1 T_j + \sum_{n=2}^{N} \beta_n X_{nit-1} + \sum_{j=1}^{J} \phi_j + \varepsilon_{it} \qquad (1)$$

where Y_{ijt} is the outcome for individual i in community j at time t. T_j is an indicator variable for the treatment status of the community, based on random assignment. X_{nit-1} are a series of n individual and household level baseline controls included to reduce residual variance, ϕ_j are block-level fixed effects based on the random assignment protocol (district, administrative post and block), and ε_{it} is the random error. We estimate all regressions using population weights[26] and robust standard errors, clustered at the community level. The key parameter of interest is β_1, the causal effect of offering preschool in treatment communities.

Our second estimate of interest is the average impact of the program on children who enroll in preschool. If preschool enrollment is endogenous, depending for example on the preferences and information of parents, then a simple regression of outcomes on

[26] Sampling weights are calculated as the inverse of the probability of selection based on the sample design. Estimates are robust to weighting (see online appendix Table 28).

an individual child level indicator for preschool attendance will yield a biased estimate of the impact of preschool attendance. To identify the unbiased effect, we use an instrumental variables (IV) approach, instrumenting individual preschool participation status with the randomized treatment status at the community level. We estimate a two stage least squares model:

$$Y_{ijt} = \eta + \gamma_1 \hat{D}_{ijt} + \sum_{n=2}^{N} \gamma_n X_{nit-1} + \sum_{\text{j=1}}^{\text{J}} \lambda_j + \xi_{it} \qquad (2)$$

$$D_{ijt} = \alpha + \theta_1 T_j + \sum_{n=4}^{N} \theta_n X_{nit-1} + \sum_{\text{j=1}}^{\text{J}} \phi_j + \varepsilon_{it} \qquad (3)$$

where D_{ijt} is an indicator variable for whether child i attended preschool. Our key parameter of

interest is γ_1, the local average treatment effect (LATE) interpreted as the effect of preschool for the subset of children who enroll in preschool thanks to the program. For the purposes of our analysis on a binary enrollment variable, we classify a child as having participated in preschool if they were reported to have enrolled and attended any length of time. In addition, to capture the effects of differential exposure, we analyze the number of months a child is enrolled in preschool as a proxy for "treatment intensity." We implement a similar IV approach, instrumenting the number of months a child attended preschool Tij by the random allocation of preschools at the community level.

Ideally, we would also like to estimate the average effect of the Treatment on Treated (TOT), the average effect of attending preschool. To interpret the IV estimate as the TOT, however, we require no preschool enrollment in control communities (Bloom, 1984). Yet 8% of target children enroll in preschools in control communities. This attendance is due primarily to the construction of seven non-program preschools,[27] rather than enrollment of children from control communities in Save the Children

[27] Through additional field work conducted in 2014, community leaders confirmed that two communities had existing preschools before 2008 and four other control communities had preschools built between the period of 2008-2010. These preschools were managed by other NGOs or Churches, and not Save the Children. One additional community, Muwawasse, had enrollment rates above 25% although a local preschool was not confirmed by the community leader. See online Appendix Table 5.

preschools. We add a dummy variable to control for those communities in the regression analysis, and interpret the IV estimates as LATE instead of TOT effects. Nonetheless, we propose three additional approaches to validate our results and approximate a TOT in the presence of potential treatment contamination in control communities.

First, taking advantage of the block randomization procedure, we identify and drop from the analysis entire blocks where preschools were built by churches or other NGOs in control communities, thus reducing the number of blocks to 23. Limiting the analysis to this subset of blocks, average preschool enrollment in control communities falls to 4%.

Second, we redefine the participation dummy variable to include as compliers only the children enrolled in preschool and who live in a treated community. In other words, we replace D_{ijt} by D'_{ijt} that is equal to one if and only if $D_{ijt} = 1$ and $T_j = 1$ in equations 2 and 3. As we did not identify any preschools other than those built by Save the Children in the sample of treated communities, we are confident that preschoolers in treatment communities attended a Save the Children preschool. In this set up, enrollment is zero in control communities by construction. Henceforth, it is an alternative, albeit likely lower bound approximation of the true TOT[28].

Third, we make use of GPS data to calculate the distance from a child's house to the nearest preschool built by Save the Children.[29] Figure 2 shows preschool participation by distance to the nearest Save the Children preschool. Enrollment clearly falls with distance and no child travels more than 5 km to a Save the Children preschool. For the GPS analysis, we ignore the community treatment status and use the distance between a child's home and the nearest Save the Children preschool as an instrument for preschool participation. Mechanically, this means replacing T_j by the continuous distance variable S_{ij} in equation 3. The estimated treatment coefficient is the

[28] To see why this alternative is a lower bound to TOT we note that the TOT= [E(Y|T=1)-E(Y|T=0)]/[E(T Z=1)-E(T|Z=0)]. Assuming E(Y|T=0) under full compliance is greater than E(Y|T=0) with treatment contamination in control communities, then our estimated LATE is a lower bound estimate of the true TOT.

[29] GPS data was collected at baseline and endline, as well as in a short re-contact survey collected in 2014. Of the 1897 households with complete surveys, 1529 households have valid GPS locations.

average effect of attending preschool for those affected by the proximity to a Save the Children preschool. These likely include all enrolled children in treatment communities and the set of children in control communities who attended a Save the Children preschool. We estimate program effects using these different constructs of treatment assignment and verify that results are largely robust to the definition of treatment assignment[30].

To summarize information and avoid data mining, we summarize our main outcomes though indices, performing factor analysis when appropriate, as the case of cognitive development and parenting, or by aggregating outcomes as in Kling et al (2001).

1.6.Results

1.6.1. Child Development

We begin by analyzing our most comprehensive measure of child development, the aggregate ASQ score, represented as a z-score transformation of the aggregate score in standard deviations from the mean of the control group. Table 3 presents the results of the ITT and IV regression specifications[31]. Each coefficient is estimated from a separate regression. The intent to treat (ITT) effect in model (1) represents a 0.184 standard deviation (σ) increase in the average total ASQ score from offering the program in treated communities (significant at the 1% level). The IV estimate in model (1), interpreted as the LATE on children enrolled in preschool because of the program's presence, is an increase of 0.37σ in the total development score. Effects are positive and significant for the sub-domains of communication, problem solving and precise motor coordination, in the range of 0.30σ to 0.35σ in the IV models (columns 2 through

[30] See online appendix section 8 and Table 28.

[31] For all tables, we present the control complier mean as a base rate to assess the program impact of the program for the relevant group of children that is affected by the policy. We present an approximation of the implied mean outcome for individuals in the control communities that would have enrolled their child in a preschool if they had the possibility, by assuming the proportion of always takers is low. We calculate the control complier mean by subtracting the estimated effect from the mean of the group of enrolled children in treated communities. For all ASQ domains, our approximation of the control complier means imply that the group of children who would have enrolled in preschool in control communities have lower development scores than other children in their communities.

4). The exception is gross motor coordination, for which there is a positive but insignificant effect of preschool participation.

Table 4 reports results on receptive vocabulary as measured by the TVIP. We report effects on the raw score, within sample standardized score, and standardized score as per the test developers. While all three coefficients are positive, the raw score is estimated imprecisely. Effects on the standardized scores are positive and significant. Preschool increases an average participant's TVIP normed score based on developers tables by 1.8 points (significant at 5%), a relative increase of about 3% relative to the control group. Figures 3 and 4 plot the raw and the normed TVIP scores by age for treatment and control groups (Changana version of the test shown). Consistent with the regression results, we observe higher scores for children in treatment communities throughout the distribution of ages.

One concern with the use of the normed TVIP score is censoring, since for each age there is a minimum raw score that can be normed[32]. About 50% of all children in our sample have censored normed scores, and the probability of censoring increases with age (Figure 5), which could downward bias the estimate of program effect, if positive. When we restrict the sample to children whose scores are not censored, the estimate of the program effect on the TVIP normed score increases to 3.18 points (column 5), and we do not find that the program changes the probability of censoring (column 4). In addition to presenting only the impact on the raw and normed scores, we can circumvent the censoring problem by following the procedure suggested by Schady et al (2014), Paxson and Schady (2007), by calculating a z-score of the raw TVIP score for each child's age in month. Column 5 of table 4 show a program impact of 0.26σ for the raw score over the control mean.

We report results from the EDI on children enrolled in primary school in table 5. We observe large and statistically significant effects on the domains of physical health, cognitive development and communication, and large but imprecisely estimated effects on other domains. Preschoolers enrolled in first grade have a 0.301σ increase in physical health, a 0.439σ increase in the cognitive domain score, and a 0.373σ increase

[32] For example, a raw score of 1 translates into a normed score of 55 for children aged 60 months, which is the same normed score for all children aged 79 months with a raw score below 16.

on the domain of communication and general knowledge[33]. While the estimated impacts on the domains of social competence and emotional maturity are large, results are estimated imprecisely and are not statistically significant at conventional levels.[34]

Overall, the results from child development tests applied to children and caregivers in the household (ASQ and TVIP) and to teachers in primary school (EDI) consistently demonstrate robust positive effects of preschool on child development as measured in the domains of cognition, communication, precise motor and socio-emotional development of young children. To obtain an aggregate measure of preschool's effects on child development, we combine the ASQ and TVIP into a single summary variable via factor analysis, summarizing the 4 ASQ domains and the internally standardized TVIP score into a "cognitive factor". The first factor ("principal", or "g") explains 92% of the variance. As shown in model 1 of Table 9, the presence of a preschool in the community increases the mean index by 0.17σ, while attending a preschool increases the index by 0.337σ.

1.6.2. Schooling

By stimulating child development, the preschool program aimed to improve school readiness and facilitate the transition of children into primary school. Table 6 presents the ITT and LATE impacts of preschool on the probability of currently being enrolled in primary school, of ever enrolling in primary school, or enrolling at the appropriate age, and of dropping out of primary school. Children who enroll in preschool have an increased likelihood of being enrolled in primary school at the time of the survey of 21.2 percentage points and an increased probability of ever enrolling of 18.2 percentage points. Particularly important in the Mozambican context is that preschool increases the probability of enrolling at the appropriate grade for age (defined as 6 years old in 1st grade). Children who attend preschool are 14.9 percentage points

[33] Results are sensitive to the exclusion of controls such as child age and the time elapsed between the start of classes and the day of the interview. On line appendix Table 28 shows the estimates without the inclusion of controls.

[34] Online Appendix Table 28 presents select individual response categories for the domain scores presented in Table 5. Significant items include being interested in mathematics, being able to count, ordering objects, recognizing geometric shapes, writing simple words, the overall social-emotional development, the ability to get along with peers, the probability of comforting other children are higher for children in the treatment group.

more likely to enroll in school at the appropriate age. The effect of preschool on primary school dropout is negative but not significant. This is not surprising given that at endline, children had only a short exposure to primary school, and dropout rates are below 4%. We also calculate an aggregated schooling index, summarizing those outcomes in a single variable. We observe an increase of 0.36σ on the aggregated schooling outcome (Table 9, model 2).

Another dimension of interest is the amount of time spent by children on school related activities. Table 7 analyzes the impact of preschool on time use. We observe that time reported on schooling and homework activities increases by 5.89 hours per week for children who enrolled into preschool, above the average of 15.7 hours spent by children in the comparison group. While the other time categories are not statistically significant, based on the magnitude of the estimated coefficients it appears that increased time on school related activities comes at the expense of "other activities" and not play, work and chores or sleep.

1.6.3.Parenting

Children spend most of their time with their parents. Parenting behavior has been linked to child cognitive and language development (Hart and Risley, 1995), and interventions aimed to improve parent cognitive stimulation have shown positive results in Jamaica (Gertler et al, 2014) and in Colombia (Attanasio et al, 2013). Through its monthly caregiver meetings, the intervention sought to build positive caregiving practices of parents and primary caregivers that would complement learning activities conducted in preschool. We construct an index of parenting practices that combines activities in which parents actively interact with their children, such as playing with toys and balls, reading to the child, telling stories, singing songs, playing games or naming objects, among other activities (see online appendix Table 14 for the full set of practices). For caregivers whose children were enrolled into pre-school, the caregiving index increases by 0.23σ (Table 9), suggesting that the program successfully promoted such practices.

1.6.4. Health

The program could affect child health by instilling self-care practices such as hand washing (heavily promoted as part of the daily routine at preschool) as well as by changing care giving practices. On the other hand, increased daily exposure to children from throughout the community could also facilitate the transmission of infectious diseases. Table 7 presents impacts of the program on self-reported health outcomes, as reported by caregivers for the Target Child. Children are 14.5 percentage points more likely to be reported as sick in the last 4 weeks (model 1), primarily having had a cough (model 4). This increase could simply reflect the healthy maturation of children's immune systems in reaction to their first real exposure to a range of viruses in the context of a group setting, but could also be viewed as a negative side effect of the program. We do not find significant health effects on other self-reported measures of illness. Appendix tables 24 and 25 explore indicators of hygiene practices and healthcare (nutritional supplementation, deworming and vaccination) and find no significant differences between the treatment and control groups on any measures, suggesting that health effects on self-reported colds likely come from increased contact between children at preschool and not changes in health practices of children and caregivers.

Appendix table 26 presents effects on anthropometric measures of height and weight. While the program initially proposed a goal of improving children's physical growth, we find no effects on measures of height or weight (models 1 and 3) or the prevalence of wasting or stunting (models 2 and 4). 32.4% of children remain stunted and 9.1% show signs of wasting at the time of the endline survey. Given that children start the preschool program at 3 years or later, well past the critical period for growth during the first 1000 days, and there is no feeding component of the program, the only plausible mechanism for influencing children's growth is through parent meetings, where nutrition was one of multiple topics covered by the program. Thus, the absence of a detectable impact is hardly surprising.

1.6.5. Impacts on Siblings and Caregivers

Having discussed the primary impacts of preschool on children who attend preschool, we now turn to effects of the preschool program on older siblings and caregivers. Having a young child in the household enrolled in preschool may free up time for older siblings and caregivers who would otherwise help with child care. Furthermore, the preschool program may have influenced parents' views on the importance of school, encouraging enrollment of other children in the household. Table 11 presents the estimated impacts of having a preschool aged child enrolled in preschool during the treatment period on the school enrollment status of children 10 to 14 years old in the same households. Children 10 to 14 were too old to enroll in preschool at the start of the program, so any impacts of the program are indirect effects. We observe a 4.9 percentage point increase in the likelihood that an older child was ever enrolled in school (model 2). While the effect on current enrollment is not statistically significant, it is in the same direction and similar magnitude. On the other hand, we see no effects on appropriate grade for age, as would be expected for the cohort of children past primary enrollment age. The positive spillover on school attendance may be explained, at least in part, by a decrease in the time older children spend taking care of younger siblings (model 4). Older children whose sibling went to preschool spend 1.2 fewer hours per week taking care of children, and spend an additional 2.7 hours on schooling and homework related activities (model 5).

Finally, we explore the effects of preschool on adult labor supply. While there is no effect on aggregate for adult household members, we find a 7.1 percentage point increase in labor supply for the primary caregiver (significant at the 10% level), representing an almost 30% increase relative to the control.[35]

[35] We find a positive and significant correlation in control communities between having a caregiver working and child's cognitive factor in control communities. We rule out, though, the possibility that our results on child cognitive gains and the impacts on child enrollment in primary school are driven by labor markets and income effects. A simple back of the envelope calculation yields an impact of 0.006σ (=0.037* 0.183) on cognitive factor from changes in labor supply, well below the estimated impact of 0.171 σ. Caregivers labor supply is not significant when include all communities and we add the treatment dummy. The lack of significant results on child anthropometrics also rule out the possibility that our main findings are driven by income effects.

1.6.6. Heterogeneous Effects by Child Characteristics

The effects of attending preschool may vary according to a child's initial level of human capital and the amount of investments made by parents. We analyze treatment heterogeneity on our summary indices of child development, schooling, parenting and health outcomes, disaggregated by child characteristics and baseline conditions including gender, orphan status, wealth, parenting skills and cognitive development. Table 10 summarizes the results. The first row presents the OLS estimates (first stage) of the impact of offering preschool on the probability of preschool enrollment. We observe significant effects for all sub-groups of children. However, enrollment is larger for girls[36] and children with ex-ante low probability of enrolling in primary school,[37] which might reflect Save the Children's targeting of vulnerable children for enrollment.

Next, for each subsample, we estimate the effect of preschool attendance on the summary index for cognitive development, schooling, parenting and health. In terms of cognitive development, preschool appears to at least partially compensate for the loss of a parent, with large effects on the cognitive factor for orphans and an insignificant effect amongst non-orphans. Children with higher initial cognitive skills at baseline have large gains in the cognitive index, while those in the lower half of the distribution experience no gains, consistent with the notion that skills are self-productive (Cunha, 2007, Cunha, 2010). Preschool effects are also larger amongst children with low predicted probability of attending primary school and for children with parents in the lower half of the distribution on the parenting skills index.

For schooling, although boys are less likely than girls to be enrolled in preschool, those who attend preschool appear to benefit more than girls. Boys who went to preschool experience an increase of 0.43σ in the schooling index, while for girls the increase is 0.33σ. Children under the median of the distribution of the asset index also show larger gains on the schooling index, as do children with a lower ex-ante

[36] We test if the effects are the same between boys and girls, as well as all other subsamples of children shown in table 10. We omit the t-statistics for lack of space.

[37] To calculate the ex-ante probability of enrolling into primary school, we estimate a Probit model for children in control communities, using information as dwelling conditions, parents' assets, parents' education, child sex, orphan status, age in months, baseline anthropometrics and baseline scores from ASQ and TVIP. We then use the forecast of the probability of being enrolled in primary school, for both children in control and treatment areas, to rank children. We split the sample at the median.

probability of attending primary school, for whom the schooling index increases by 0.62σ, while we observe no effects of preschool on the subset of children who had a high predicted chance of attending primary school. Finally, the impact on schooling attendance is also larger for children under the median value of the parenting index, which means that preschool might be compensating for low parental investments in children.

Figures 7 and 8 plot a non-parametric regression of the actual probability of primary school enrollment (ever enrolled and currently enrolled, respectively) against the predicted probability.[38] The difference between the red (treatment) and the dashed (control) lines shows the effect of preschool on primary school enrollment for a given level of ex-ante probability of being into primary school. For both figures 7 and 8, the effect is larger at lower predicted levels, suggesting that preschool is in fact most effective at promoting school enrollment and attendance for the most disadvantaged (i.e. least likely to enroll) children.

The final two rows of Table 10 present heterogeneous effects on the aggregate parenting and health indices. We observe that effects on parenting skills were larger for non-orphans and parents of children above the median cognitive factor. For health, no sub-groups present statistically significant effects apart from non-orphans, for whom the program appears to have generated a positive health effect.

[38] The predicted probability of going to primary school is calculated through a Probit regression of the variable that indicates that the child has even been to primary school on a set of controls, only at control communities. The model is then used to extrapolate the probability for children in treatment communities.

1.7. Conclusion

We present experimental evidence on the effectiveness of a community based preschool intervention in rural communities in Mozambique. By age 3, a large fraction of children in the study areas present severe delays in physical growth (as evidenced by the high rates of stunting) and signs of strong lacunas in vocabulary development. We find that children who attend preschool improve a number of important dimensions of child development, including cognitive, fine motor and socio-emotional, leading to higher levels of school readiness and significantly increased primary school enrollment at the appropriate age. The program also produced positive impacts on the school enrollment of older siblings and increased the labor supply of primary caregivers. Taken together these results suggest that low-cost community based preschool interventions such as the one studied here are a promising policy alternative for investing in early child development. At US$ 3,09 per student per month, the intervention is an affordable and effective way to improve the lives not only of young children who attend preschool, but also to improve the welfare of families of preschool aged children.

While the results discussed here are very encouraging, a number of caveats are in order. While the first randomized experiment of a preschool intervention in rural Africa, with rich data, large sample sizes, results must be extrapolated with caution.

First, whether or not the results of the small and well implemented program studied here can be reproduced at a national level or by a government agency should be tested using rigorous evaluations of similar interventions in other countries and contexts.

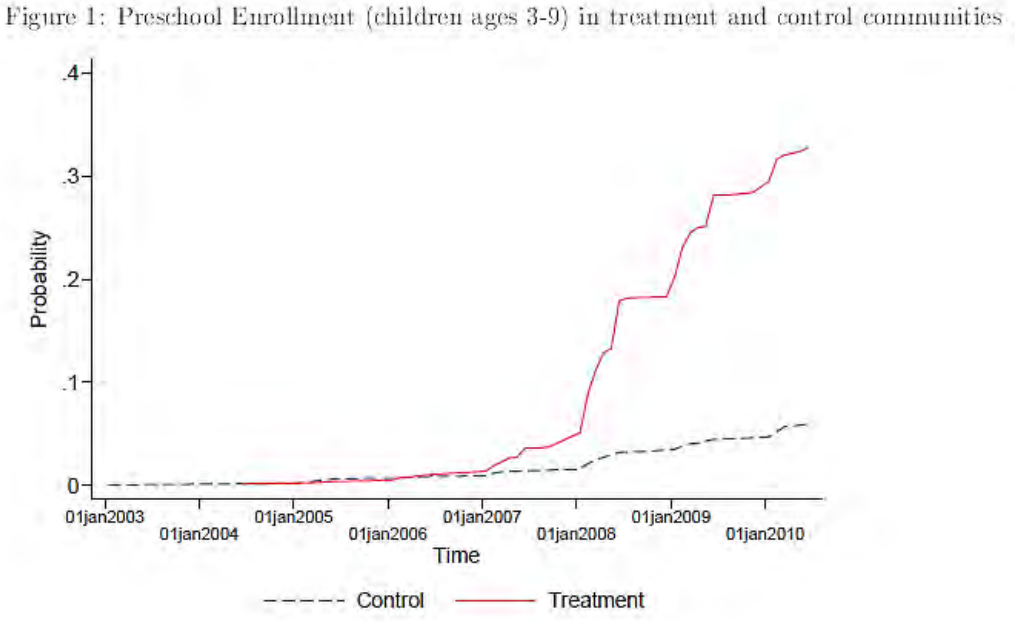
Second, if children who voluntarily enroll in preschool differ from those who do not, for example if enrolled children are those who expect to benefit most, then including a wider distribution of the population could produce different results. As documented in the paper, several demand-side constraints exist that prevent children from participating in ECD programs even when these are locally available.

Further research will be needed to better understand how to alleviate these constraints, so as to ensure that all targeted children, especially the most vulnerable,

can benefit. Finally, it is important to note that the preschool program had only mild impacts on children's language development and there are mixed results on children's health. These aspects of the program design merit further consideration.

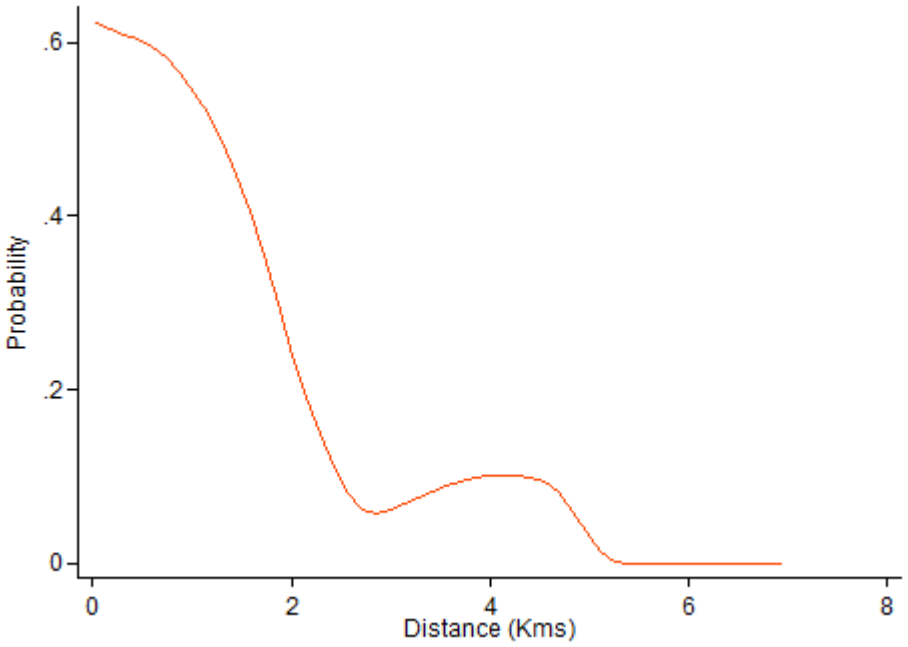
1.8. Figures

Figure 1- Preschool Enrollment



Notes: Figure 1 presents the probability of enrollment on preschool for each month from January 2003 until June 2010, for controls and treatment communities. Probability is the proportion of children aged 3 to 6 at endline survey who ever attended preschool. The probability is constructed through caregivers' report of the month and year the child started attending preschool.

Figure 2- Preschool enrollment by distance



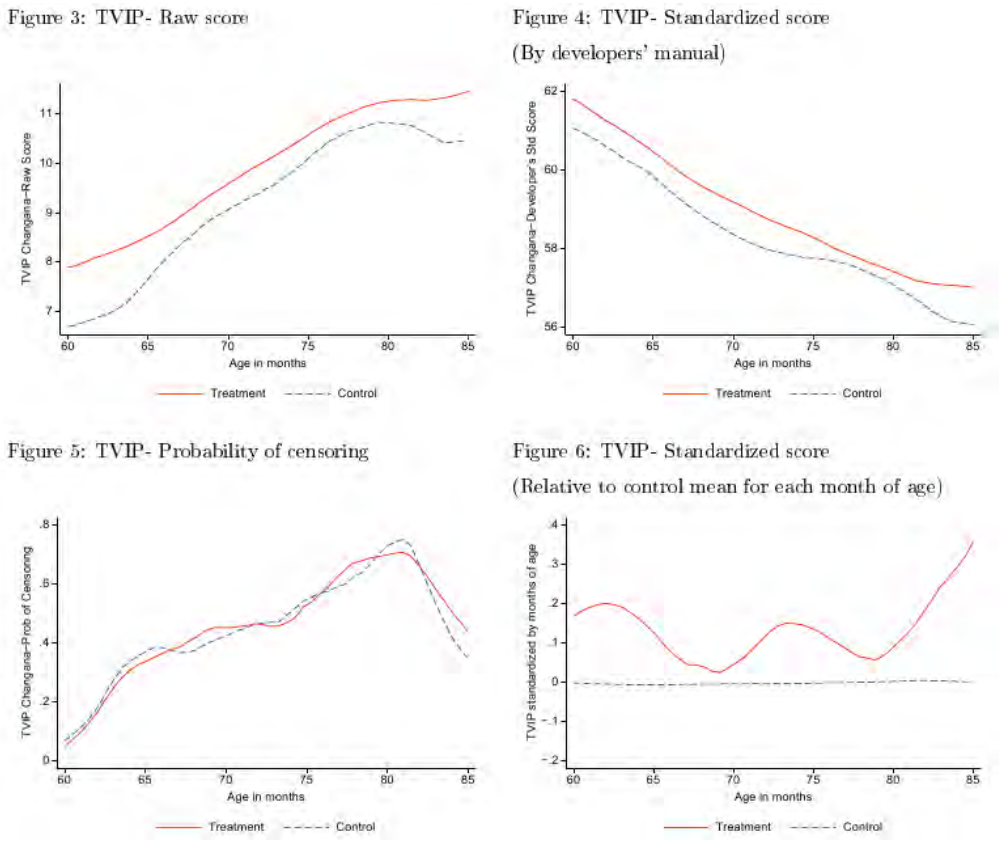
Notes: Figure 2 shows the proportion of children aged 3 to 9 at endline who have ever been to preschool, by distance to the nearest preschool operated by Save the Children.

Figure 3-TVIP raw score

Figure 4-TVIP standardized score

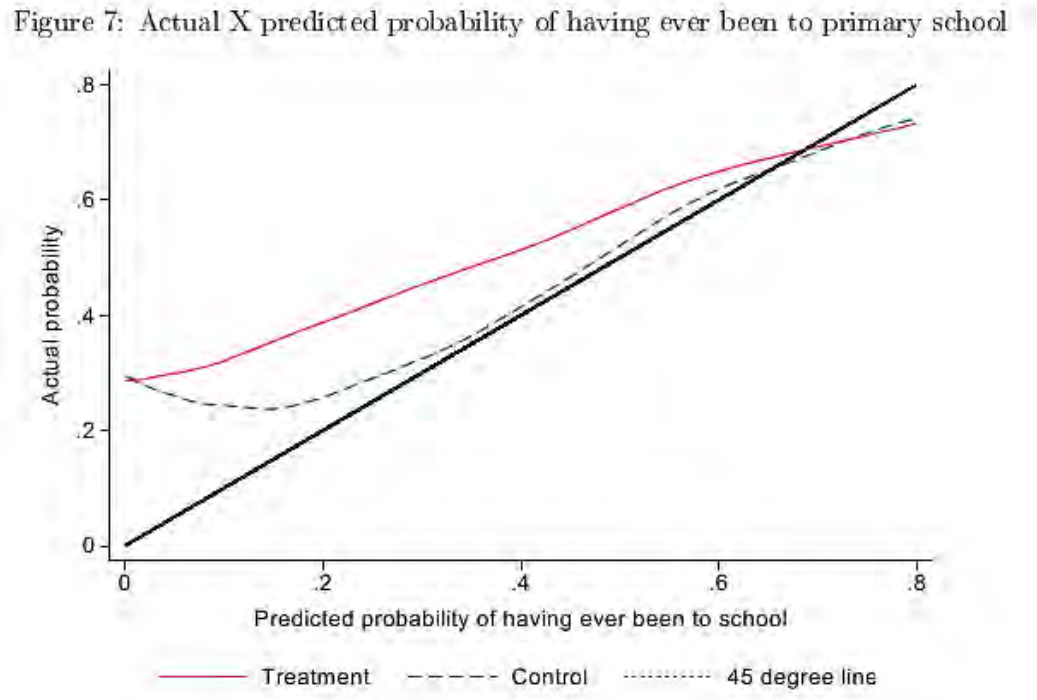
Figure 5-TVIP probability of censoring

Figure 6-TVIP standardized score



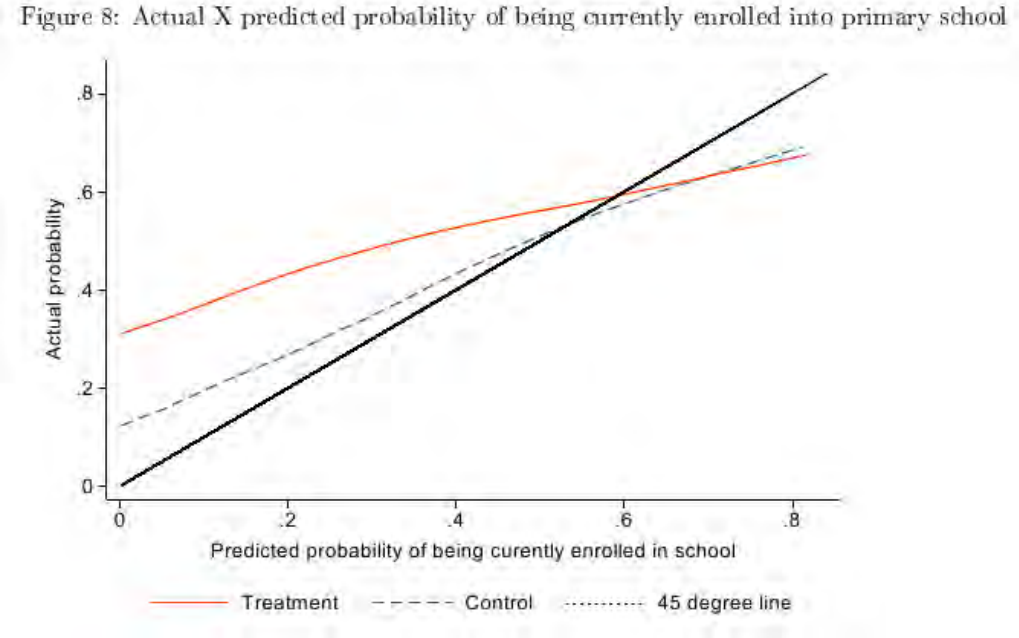
Notes: Figures (3), (4) and (6) show non parametric regressions of TVIP scores on age in months, by treatment and control communities. Figure (5) shows a non-parametric regression of the probability of censoring on age in months. At figure (4), the score is normalized according to the test developers' standard. At figure 6, we calculate the within sample standardized score by subtracting the control mean and dividing by control standard deviation, for each age in months.

Figure 7- Heterogenous impact on probability of having ever been to primary school



Notes: Figure 7 shows a non-parametric regression of the actual probability of having ever been to primary school against the predicted probability of having ever been to primary school, for all children aged 5 to 9 at endline. The predicted probability of going to primary school is calculated through a Probit regression of the variable that indicates that the child has even been to primary school on a set of controls, only at control communities. The model is then used to extrapolate the probability for children in treatment communities.

Figure 8- Heterogeneous impact on probability of enrolling into primary school



Notes: Figure 8 shows a non-parametric regression of the actual probability of being currently enrolled into primary school against the predicted probability of being enrolled into primary school, for all children aged 5 to 9 at endline. The predicted probability is calculated through a Probit regression of the variable that indicates that the child is currently enrolled into primary school on a set of controls, only at control communities. The model is then used to extrapolate the probability for children in treatment communities.

1.9.Tables

Table 1

PRESCHOOL CHARACTERISTICS

Teacher characteristics (N=98)	
Female	93.22 %
Age	33
Years of Education	6.16
Married or partnered	70.69 %
Household size	5.98
Number of own children	3.05
Own child attends preschool	54.39 %
Hours spent at preschool per day	3.46
Hours spent on training, meetings and other preschool related activities per month	3.64
Checklist for items present at the classroom in the last 30 days (N=57)	
Blackboard	96.55 %
Chalk	91.38 %
Notebooks or sheets to write on	89.66 %
Pencils and pens	93.10 %
Picture books	86.21 %
Picture cards	89.66 %
Cards games	75.86 %
Construction blocks	93.10 %
Dolls/puppets	79.31 %
Other toys	91.38 %
Attendance lists	93.10 %
Chairs	29.31 %

Mats	72.41 %
Checklist for items present at the preschool in the last 30 days (N=27)	
Running water	39.66 %
Soap	72.41 %
Swing	87.93 %
Kids climber	79.31 %
Seesaw	68.97 %

Notes: Author's calculation from endline survey

Table 2- Baseline balance

BASELINE BALANCE				
	Treatment mean N=1028	Control mean N=879	Means difference	T-stat
Household characteristics				
Number of household members	5.085	4.937	0.148	0.930
Asset index	-0.213	0.080	-0.293	-1.034
Number of rooms at home	2.084	2.224	-0.140	-1.466
Improved latrine at home	0.154	0.118	0.036	1.547
Adobe walls at home	0.662	0.679	-0.017	-0.441
Dirty floor at home	0.806	0.838	-0.033	-1.198
Target child characteristics				
Female	0.509	0.497	0.011	0.521
Age (years)	3.456	3.480	-0.025	-0.926
Speaks Portuguese	0.133	0.118	0.015	0.489
Orphaned	0.100	0.101	-0.001	-0.041
ASQ total Score	199.129	196.569	2.560	0.774
TVIP raw score	5.783	5.590	0.194	0.553
TVIP score-within sample standardized score	0.073	0.154	-0.080	-0.505
TVIP normed score by developers' table	78.858	78.637	0.221	0.367
Child had skin problems in the last 4 weeks	0.156	0.103	0.053	0.838
Child had swallowing difficulties in the last 4 weeks	0.038	0.029	0.008	0.659
Respiratory illness (flu, pneumonia, asthma) in the last 4 weeks	0.139	0.115	0.024	1.021
Child had diarrhea in the last 4 weeks	0.064	0.032	0.031	3.045
Child slept in mosquito net the night before	0.148	0.103	0.046	1.494
Child has been dewormed in the last 12 months	0.114	0.098	0.016	0.907
Child received vitamin A (Health card)	0.424	0.398	0.025	0.762

Child was diagnosed with malaria in the last 4 weeks	0.074	0.063	0.011	0.933
Weight for age z-score	-0.314	-0.257	-0.057	-0.654
Height for age z-score	-1.553	-1.506	-0.048	-0.481
Weight for height Z-score	1.298	1.244	0.054	0.389
Caregiver characteristics				
Age (years)	36.027	36.330	-0.304	-0.297
Female	0.859	0.820	0.039	1.122
Speaks Portuguese	0.487	0.490	-0.003	-0.058
Read and write	0.611	0.632	-0.021	-0.537
Years of education	3.236	3.410	-0.174	-0.693
Married or partnered	0.660	0.692	-0.032	-1.009
Reads/skims through books with child	0.532	0.521	0.011	0.246
Plays with child in the garden	0.457	0.412	0.045	1.183
Spends time naming and drawing objects with child	0.404	0.370	0.034	0.780
Plays games with child	0.421	0.468	-0.047	-1.053
Practices self-sufficiency activities with child	0.581	0.579	0.002	0.054

Notes: T-stats computed through simple linear regression with standard errors clustered at community level. Asset index calculated by principal components using a list of household assets. Dirty floor includes mud, sand, and adobe. Within sample standardized TVIP score calculating by subtracting the age in months controls average and dividing the age in months standard deviation.

Table 3- Child development- Ages and stages questionnaire

CHILD DEVELOPMENT-AGES AND STAGES QUESTIONNAIRE

Dep var:	Total ASQ Score (1)	Communication (2)	Problem Solving (3)	Precise Motor Coordination (4)	Gross Motor Coordination (5)
OLS: Treatment community	0.184***	0.174***	0.166***	0.152***	0.080
	(0.043)	(0.054)	(0.037)	(0.044)	(0.054)
IV: Ever been to preschool	0.370***	0.350***	0.334***	0.307***	0.161
	(0.096)	(0.116)	(0.078)	(0.098)	(0.111)
Observations	1,831	1,831	1,831	1,831	1,831
Control Mean:	0.000	0.000	0.000	0.000	0.000
Control Std:	1.000	1.000	1.000	1.000	1.000
Control Complier Mean:	-0.257	-0.285	-0.212	-0.189	-0.140

Notes: This table reports estimates of the effects of the provision of preschool centers at community and the estimates of the effects of preschool attendance. Only the target children were tested. The first line reports the estimates of an OLS regression of each section of the Ages and Stages Questionnaire on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance. Preschool attendance is instrumented by the community treatment status. Each variable was standardized by subtracting the mean at control communities and by dividing by the standard error. Control complier mean calculated as in Kling et al (2001). Total ASQ score is the sum of all 4 section scores. All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in months, sex, height for age at baseline, weight for age at baseline, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, stunted at baseline, child with risks of communication deficits at baseline, child with risks of motor coordination deficits at baseline, child with risks of precise motor coordination at baseline, child with risks of problem resolution deficits at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old and household -age equivalent- size. See the online appendix for alternative specifications and alternative samples.

Table 4-TVIP scores

TESTE DE VOCABULARIO POR IMAGENS PEABORY (TVIP) SCORES

Dep var:	Raw Score (1)	Normed Score (All) (2)	Probability of Censoring (3)	Normed Score (Non-censored) (4)	Within-sample standardized score (5)
OLS: Treatment community	0.664	0.910*	0.001	1.716**	0.130**
	(0.429)	(0.456)	(0.022)	(0.742)	(0.063)
IV:Ever been to preschool	1.313	1.800**	0.001	3.184**	0.258**
	(0.831)	(0.862)	(0.044)	(1.312)	(0.123)
Censored Observations	x	X	x		x
Observations	1,801	1,801	1,801	925	1,801
Control Mean:	8.962	59.249	0.472	63.045	0.000
Control Standard Deviation:	6.739	6.942	0.500	7.794	0.983
Control Complier Mean:	8.634	57.936	0.452	60.484	-0.135

Notes: This table reports estimates of the effects of the provision of preschool centers at community and the estimates of the effects of preschool attendance. Only the target children were tested. The first line reports the estimates of an OLS regression of each section of the Teste de Vocabulario por Imagens Peabody (TVIP) on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance. Preschool attendance is instrumented by the community treatment status. The raw score is calculated by taking the number of questions answered by child and subtracting the number of wrong answers. The within sample standardized score is calculated by subtracting the average of raw score and dividing by the standard deviation for each month of child age, as in Schady et al (2014). The standardized score according to developers table reflects the relative position of the child from a sample of Mexican and Puerto Rican children. According to those norms, the average is 100 and one standard deviation is 15, for all ages. For each age, there is a minimum score that can be normed, and last column only contains observations that are higher than the minimum score. Control complier mean calculated as in Kling et al (2001). All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in months, sex, height for age at baseline, weight for age at baseline, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, stunted at baseline, child with risks of communication deficits at baseline, child with risks of motor coordination deficits at baseline, child with risks of precise motor coordination at baseline, child with risks of problem resolution deficits at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old and household -age equivalent- size. See the online appendix for alternative specifications and alternative samples.

Table 5-EDI

TABLE 5
EARLY DEVELOPMENT INDEX- BY DOMAINS

Dep var:	Physical Health and Well-being (1)	Communication and General Knowledge (2)	Cognitive Development and Language (3)	Social Competence (4)	Emotional Maturity (5)
OLS: Treatment community	0.301*	0.373**	0.429***	0.329	0.300
	(0.154)	(0.153)	(0.148)	(0.233)	(0.222)
Observations	919	919	919	919	919
Control Mean:	-0.054	-0.030	-0.094	-0.070	-0.057
Control Standard Deviation:	0.986	1.013	1.051	1.031	0.890

Notes: This table reports estimates of the effects of the provision of preschool centers at development domains of first graders, as measured by the Early Development Index. Sample consists of first graders of primary schools, randomly chosen from the list of first graders from each primary school operating at the sampling area. Each domain is standardized with mean zero and standard deviation equal to one. The first line reports the estimates of an OLS regression of each development domain on the dummy that indicates that a preschool was built at the community where the primary school is located. All regressions include dummies of randomization blocks, local district and local administrative post. Standard errors clustered at class level. Controls include child age in years, sex, time elapsed since the start of school year and the date of the interview, flag for date of start of classes not reported, date of interview, number of students at class, teacher's sex, teacher's highest grade completed, flag for highest grade not reported, teacher's subjective familiarity with students. See online appendix for item to item regressions.

Table 6-Primary school outcomes

PRIMARY SCHOOL OUTCOMES -CHILDREN AGED 5 TO 9

Dep var:	Currently Enrolled at School (1)	Ever gone to School (2)	Appropriate Grade for Age (3)	Dropout from School (4)
OLS: Treatment community	0.082***	0.070***	0.056***	-0.007
	(0.023)	(0.022)	(0.019)	(0.009)
IV:Ever been to preschool	0.212***	0.182***	0.149***	-0.020
	(0.061)	(0.056)	(0.050)	(0.027)
Observations	2,591	2,686	2,891	1,872
Control Mean:	0.635	0.676	0.474	0.040
Control Standard Deviation:	0.482	0.468	0.499	0.196
Control Complier Mean:	0.414	0.487	0.336	0.059

Notes: This table reports estimates of the effects of the provision of preschool centers at community and the estimates of the effects of preschool attendance. Sample includes all children aged 5 to 9 at endline survey. The first line reports the estimates of an OLS regression of each outcome on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance. Preschool attendance is instrumented by the community treatment status. Schooling index calculated by standardizing each one of the variables from columns (1) to (4) and by averaging, as in Kling et al (2007).All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Control complier mean calculated as in Kling et al (2001). Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in years, sex, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size. See online appendix for alternative specifications and alternative samples.

Table 7-Time use

TABLE 7
HOURS ON EACH ACTIVITY DURING LAST WEEK -CHILDREN AGED 5 TO 9

Dep var:	School and Homework	Play	Work at Family's Plot	Household Chores	Caring for children elders and sick	Community Meetings	Sleep	Other Activities
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
OLS: Treatment community	2.214***	0.468	-0.381	-0.078	-0.034	-0.312	0.110	-1.986
	(0.817)	(0.898)	(0.248)	(0.136)	(0.126)	(0.225)	(0.814)	(1.610)
IV: Ever been to preschool	5.869***	1.240	-1.011	-0.207	-0.091	-0.827	0.290	-5.263
	(2.189)	(2.395)	(0.645)	(0.362)	(0.334)	(0.619)	(2.149)	(4.219)
Observations	2,891	2,891	2,891	2,891	2,891	2,891	2,891	2,891
Control Mean:	15.708	21.819	2.597	0.749	0.567	1.054	61.407	64.100
Control Standard Deviation:	15.120	15.797	6.681	3.358	2.385	5.371	16.861	32.204
Control Complier Mean:	12.739	21.277	2.284	0.712	0.590	1.640	62.955	65.804

Notes: This table reports estimates of the effects of the provision of preschool centers at community and the estimates of the effects of preschool attendance. Sample includes all children aged 5 to 9 at endline survey. The first line reports the estimates of an OLS regression of hours on each activity during the week the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance, instrumented by the community treatment status. Time on each activity measured in hours during the week. All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Control complier mean calculated as in Kling et al (2001). Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in years, sex, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size. See online appendix for alternative specifications and alternative samples.

Table 8-Health

TABLE 8
HEALTH OUTCOMES OF TARGET CHILDREN

	Child was sick in the last four weeks (1)	Child had skin problems in the last four weeks (2)	Child had Diarrhea in the last four weeks (3)	Child had cough in the last four weeks (4)	Child had fever in the last four weeks (5)	Child had breathing problems in the last four weeks (6)	Child was diagnosed with malaria in the last four weeks (7)	Child had swallowing difficulties in the last four weeks (8)
OLS: Treat. community	0.072**	-0.023	-0.015	0.077**	0.028	-0.038	-0.009	0.009
	(0.030)	(0.020)	(0.012)	(0.038)	(0.021)	(0.038)	(0.017)	(0.009)
IV: Ever been to preschool	0.145**	-0.047	-0.031	0.155*	0.056	-0.068	-0.018	0.018
	(0.065)	(0.040)	(0.023)	(0.082)	(0.043)	(0.068)	(0.034)	(0.018)
Observations	1,836	1,837	1,832	1,839	1,833	829	1,828	1,829
Control Mean:	0.365	0.146	0.080	0.443	0.283	0.388	0.169	0.040
Control Std:	0.482	0.353	0.271	0.497	0.451	0.488	0.375	0.196
Control Complier Mean:	0.291	0.178	0.086	0.351	0.242	0.390	0.182	0.024

Notes: This table reports estimates of the effects of the provision of preschool centers at community and the estimates of the effects of preschool attendance. Sample includes only target children. The first line reports the estimates of an OLS regression of each outcome on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance, instrumented by the community treatment status. All health outcomes reported by caregiver. Control complier mean calculated as in Kling et al (2001). All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in months, sex, height for age at baseline, weight for age at baseline, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, stunted at baseline, child with risks of communication deficits at baseline, child with risks of motor coordination deficits at baseline, child with risks of precise motor coordination at baseline, child with risks of problem resolution deficits at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old and household -age equivalent- size. See the online appendix for alternative specifications and alternative samples.

Table 9-Summary indeces

TABLE 9
SUMMARY INDECES

Dep var:	Cognitive Index (1)	Schooling index (2)	Parenting index (3)	Health index (4)
OLS: Treatment community	0.171***	0.136***	0.119**	0.064
	(0.050)	(0.044)	(0.052)	(0.062)
IV:Ever been to preschool	0.337***	0.360***	0.232**	0.126
	(0.107)	(0.117)	(0.104)	(0.125)
Observations	1,686	2,891	1,630	1,697
Control Mean:	0.075	-0.039	-0.011	-0.002
Control Standard Deviation:	0.956	1.010	0.994	1.017
Control Complier Mean:	-0.183	-0.402	-0.141	-0.099

Notes: This table reports estimates of the effects of the provision of preschool centers at community and the estimates of the effects of preschool attendance. At columns (1), (3) and (4), sample includes only target children. For column 2, sample are all children aged 5 to 9 at endline survey. The first line reports the estimates of an OLS regression of each outcome on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance. Preschool attendance is instrumented by the community treatment status. See on line appendix for the construction of each index. All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Control complier mean calculated as in Kling et al (2001). Estimates weighted by community population size. Standard errors clustered at community level. Controls at column (2) include child age in years, sex, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size. For columns (1), (3) and (4), child age in months, height for age at baseline, weight for age at baseline, dummy for being under median of asset index at baseline, stunted at baseline, child with risks of communication deficits at baseline, child with risks of motor coordination deficits at baseline, child with risks of precise motor coordination at baseline and dummy for child with risks of problem resolution deficits at baseline are included in controls, in addition to controls from column (2). See online appendix for alternative specifications and alternative samples.

Table 10- Heterogeneity

TABLE 10
HETEROGENEOUS EFFECTS BY SUBGROUPS OF CHILDREN

Groups:	Boys	Girls	Non orphans	Orphans	Under median Asset Index	Above Median Asset Index	Low prob prim school	High prob prim school	Under median cognitive factor	Above median cognitive factor	Under median parent Index	Above median Parent Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep var:												
OLS: Child ever	0.349***	0.401***	0.372***	0.378***	0.384***	0.368***	0.461***	0.287***	0.369***	0.378***	0.352***	0.388***
been to preschool	(0.026)	(0.028)	(0.038)	(0.026)	(0.023)	(0.035)	(0.038)	(0.023)	(0.076)	(0.025)	(0.032)	(0.028)
IV: Cognitive factor	0.387***	0.249	0.053	0.426***	0.322***	0.417***	0.376***	-0.045	-0.768	0.418***	0.536***	0.218**
	(0.130)	(0.154)	(0.285)	(0.125)	(0.117)	(0.159)	(0.101)	(1.007)	(0.613)	(0.102)	(0.208)	(0.096)
IV: Schooling index	0.431**	0.334***	0.333	0.381***	0.445***	0.245	0.625***	-0.099	0.499	0.319***	0.714***	0.138
	(0.191)	(0.113)	(0.316)	(0.109)	(0.165)	(0.194)	(0.131)	(0.147)	(0.578)	(0.120)	(0.177)	(0.173)
IV: Parenting index	0.302**	0.324**	0.815**	0.232*	0.299**	0.268	0.233*	0.365*	-0.192	0.324***	0.345	0.225*
	(0.143)	(0.148)	(0.392)	(0.129)	(0.146)	(0.200)	(0.125)	(0.197)	(0.453)	(0.118)	(0.276)	(0.120)
IV: Health index	-0.211	0.106	0.613**	-0.142	0.021	-0.148	-0.111	0.117	0.247	-0.071	-0.186	0.001
	(0.209)	(0.185)	(0.302)	(0.172)	(0.197)	(0.215)	(0.168)	(0.240)	(0.641)	(0.160)	(0.223)	(0.211)

Notes: Panel A shows the estimates of an OLS regression of the probability of enrollment into preschool on the dummy that indicates community treatment status. Panels B, C, D and E show the estimates of an instrumental variables regression of each outcome on the dummy indicating preschool enrollment. All children aged 5 to 9 are included in models (1), (2), (3), (4), (5), (6), (7), (8), (11), (12), panels A and B. Models (9) and (10), and Panels C, D and E contain only target children. Columns (1) and (2) split the sample between boys and girls. Columns (3) and (4) splits the sample between orphaned (father or mother deceased, or both) and non-orphaned children. Columns (5) and (6) splits the sample by wealth. Column (5) contains only children who are under the median of the asset index. Refer to the on line appendix for the construction of the index. Columns (7) and (8) split the sample by the probability of having ever been to primary school. Probability is calculated by estimating a probit model for having ever been to primary school on a set of controls, for children in control communities. The model is then used to extrapolate the probability to children in treatment communities. Median probability is 0.76. Columns (9) and (10) split the sample by cognitive factor. Refer to the on line appendix for the construction of the factor. Columns (11) and (12) split the sample by the parenting index. Refer to the on line appendix for the construction of the index. Refer to the on line appendix for the construction of the health index. ll regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in years, sex, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size.

Table 11-Spillover to siblings and caregivers

SPILLOVERS: SCHOOLING AND LABOR SUPPLY OF OTHER HOUSEHOLD MEMBERS

	10-14 Year Old Children					Caregiver
Dep var:	Ever gone to school (1)	Appropriate grade for age (2)	Dropout from school (3)	Time taking care of children (4)	Time on School and Homework (5)	Worked in last 30 days (6)
OLS: Treatment community	0.028**	0.002	0.013	-0.659**	1.477*	0.037*
	(0.012)	(0.035)	(0.012)	(0.252)	(0.809)	(0.021)
IV: Younger household member has been to preschool	0.049**	0.003	0.022	-1.205**	2.703*	0.071*
	(0.023)	(0.060)	(0.021)	(0.458)	(1.537)	(0.040)
Observations	1,660	1,372	1,544	2,035	2,035	1,726
Control Mean:	0.926	0.484	0.054	2.075	2.075	0.240
Control Standard Deviation:	0.262	0.500	0.227	4.735	4.735	0.428
Control Complier Mean:	0.892	0.492	0.033	3.237	-0.671	0.184

Notes: Sample for models (1)- (5) includes siblings of target children who are aged 10 to 14 at endline and who had not been enrolled in preschool. Model (6) includes primary caregivers. The first line reports the estimates of an OLS regression of each outcome on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of having a younger sibling who went to preschool. The endogenous variable is a dummy that is equal to one if any younger household member has been to preschool. Instrument is the community treatment status. Time in models (4) and (5) measured in hours per week. All regressions include dummies for randomization blocks, local district and local administrative post, and non-Save the Children preschools. Control complier mean calculated as in Kling et al (2001). Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in years, sex, parents speak Portuguese at baseline, mother deceased at baseline, father deceased at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size.

2
Paying Students to Graduate from High School: Evidence from Brazil

2.1. Introduction

Despite substantial gains associated with high school completion, high school dropout rates are still large both in the US as well as in many countries of the developing world. Previous research has consistently documented negative consequences from high school dropout, such as lower adult earning (Oreopoulos, 2006), higher chances of being unemployed (Oreopoulos, 2007), higher chances of committing crime and ending up in jail (Lochner and Moretti, 2004), higher likelihood of teenage pregnancy (Black et al., 2004) and lower overall satisfaction with life (Oreopoulos, 2007).

The inconsistency between such large returns from high school completion and high dropout rates has led many economists to suggest that adolescents either ignore or highly discount the benefits of completing high school. This reasoning is consistent with the literature on neurology and psychology (Spear, 2000) that indicate adolescence as a period especially susceptible to myopic behavior. If this is true, incorporating insights from behavioral economics on the design of financial incentives for adolescents could help increase high school completion rates.

While anti-dropout experimental programs have generally failed to increase high school graduation rates (Dynarski and Gleason, 2002), recent experiments with cash incentives have shown promising results of increased enrollment in secondary and tertiary levels (Barrera-Osorio et al. 2011; Angrist and Lavy 2009). In Latin America, where poor youths face high risks of dropping out of high school, being jobless, engaging in substance abuse, behaving violently, and engaging in unsafe sexual practices (Cunningham et al., 2008), government have been recently experimenting with programs that pay students upon high school completion.

This paper presents evidence on the impacts of Renda Melhor Jovem program, an attainment award targeted at socio-economic disadvantaged secondary students in the State of Rio de Janeiro, Brazil. The program consists of annual

deposits in the student's bank account for each year of high school successfully completed. Balance accrues interest and can only be fully withdrawn upon high school graduation in three years' time, which is the expected duration of secondary education in Brazil.

On top of offering cash incentives to students, the program also tries to motivate students by exploiting their loss aversion: At any moment, students can see their full balance, but can withdraw only up to 30% of each year's transfer. The penalty for failing a grade or dropping out of school is strong. All the remaining balance is lost, and the student becomes ineligible for future transfers.

I take advantage of the phased in expansion of the program across municipalities and of program eligibility rules to compare educational outcomes for schools with different participation rates. For each percentage point increase on the proportion of eligible students, I estimate a 0.095 pp gain on grade passing and a decrease of 0.056 pp on dropout rates.

The combination of program details such as transfers made to students' bank accounts, the delayed payment and the penalty for not passing grade is a unique feature of the program that has not been explored yet by previous research. Equally important, unlike previous research that focused on small scale experiments (Barrera-Osorio et al. 2011; Angrist and Lavy 2009), as far as I know this is the first paper to explore the impacts of a high school attainment award implemented at large scale, as a real public policy. Scaling up such interventions can involve several logistical issues, and a successful implementation depends crucially on the cooperation of teachers and principals.

Not only there is scarce evidence on the effects of programs that pay high school students for attainment and at the same time motivates student by exploring loss aversion, but there is also very little evidence on programs that can effectively increase high school graduation rates of very poor and vulnerable students.

The context studied here involves students under extreme poverty and schools with low academic standards and without a strong tradition of accountability. I present evidence of an intervention with strong positive results on a population of very poor adolescents vulnerable to many risky behaviors including early pregnancy, use of drugs and involvement with criminal activities and violence (Cunningham et al., 2008). In that sense, the positive results presented here show a

promising way to increase educational attainment for at-risk students for which early intervention is seen no longer as possible.

This paper contributes to at least two more strands of the literature. First, this paper complements the literature on programs that reward students for completing specific tasks, such as reading books during summer (Fryer, 2011), attaining good grades in regular exams (Fryer 2011; Bettinger 2012), scoring above certain threshold on advanced placement exams (Jackson, 2010), or maintaining good grades in college (Angrist et al., 2009). This literature has found mixed results from those interventions. While some experiments had no impact on observed outcomes, others only had impact on specific school subjects or on restricted subgroups of students.

This work is also related to the literature on the effects of CCTs, and the effects of its design features. Program Renda Melhor Jovem was implemented over and and above two CCTs, and the incentive scheme can be thought as a complementary feature that can included on the design of CCT programs. While CCT programs can in part affect family decisions and student behavior through income effects, incentives play an important role in motivating students (Baird et al. 2011; Kazianga et al. 2012). Right choosing what to incentivize (Barrera-Osorio et al., 2011), how to frame the incentive (Barrera-Osorio and Filmer, 2013) or how label it (Benhassine et al., 2015) has important consequences over the magnitude of the program impact. Correctly targeting and designing the incentive portion of CCT programs is crucial in order to maximize its cost effectiveness.

Two papers are closest to mine: Angrist and Lavy (2009) analyze the effect of a school based randomized trial implemented in 20 Israeli schools. The achievement award incentivized students to pass their high school certification exam, the Bagrut, by offering a cash transfer for students who get certified. While the Israeli experiment explicitly incentivized learning, program Renda Melhor Jovem incentivizes timely high school graduation. The difference in context is subtle but important, as in Rio de Janeiro promotion rules do not depend on achievement on standardized test scores. In principle, this could create a perverse incentive for teachers to promote students with low academic records. We test this hypothesis and find no evidence that standardized test scores at the senior year decrease after the introduction of the program.

The other related paper is the work of Barrera-Osorio et al. (2011), who investigate the effects of a pilot program that tested alternative designs of conditional cash transfers (CCT) in Bogota. In one of the tested settings, part of the transfers was made upon enrolling in a tertiary institution or after one year from graduating from high school, in a lump sum transfer. Bogota's experimental program did not require students to open a bank account, neither threatened students to lose all their balance for not passing grades. In Rio, students who open their account and are promoted to the next grade always see the remaining balance blocked. Previous studies have shown that financial incentives can be leveraged by the "endowment effect" arising from loss aversion (Kahneman et al., 1991). Losing the money that has been deposited in the account should be more painful than receiving it is pleasurable. Consequently, according to Prospect Theory (Kahneman and Tversky, 1979), announcing that the student has some money, but that can be lost, can be much more effective than only promising to pay a lump sum transfer upon high school graduation.

The level of implementation is a second important difference between the two interventions. The pilot program from Bogotá was randomized at the individual level, while Rio's program was implemented at the school level, i.e, all high school students under extreme poverty at the school were eligible to earn the attainment award. Implementation at the school level involves the collaboration of the principal and teachers in communicating program rules and making sure that all eligible students open a bank account, which in practice can translate in lower program effects. By the other hand, the implementation at the school level can benefit from peer effects at the classroom that can multiply the impact of the program (Glaeser et al, 2003).

The remainder of the paper is organized as follows: At section 2 I make a brief description of the program. At sections 3 and 4 I present the data and I describe the process of merging them. Section 5 presents the empirical strategy, while section 6 presents the results. In section 7 I test whether the observed effects are due to income effects or to the incentive, while in section 8 I present the conclusions from the study.

2.2. Institutional environment: The Program Renda Melhor Jovem

The Government of the State of Rio de Janeiro launched in 2011 a new strategy for fighting extreme poverty and increasing high school completion rates called Rio Sem Miséria (Rio Without Misery). The strategy was comprised of two main arms. By one side, extreme poor beneficiaries of federally-run Bolsa Família program receive additional transfers to match Rio's line of extreme poverty (R$ 100 per capita per month), without any additional conditionality to the families. Beneficiary families are chosen via an algorithm that predicts per capita income. This first arm of the strategy is called Renda Melhor program. By the other side, high school students from extreme poor families are incentivized to graduate from high school by participating from a savings incentive scheme, called Renda Melhor Jovem program.

Renda Melhor Jovem program was inspired by existing experiences from Mexico (Jovenes con Oportunidades) and from Bogota (Subsidios Condicionados a la Assistencia Escolar), and its final goal is to make vulnerable youth scape from poverty when adults by incentivizing them to graduate from high school. The program consists on an incentive scheme that awards extremely poor students for passing each grade on high school.

Each participating student receives a transfer for each grade passed in high school. The annual award is deposited in a bank account owned by the student, but the account has a special feature: Students can only withdraw up to 30% of the balance that was deposited in each given year. Balance accrues interest at the rate of Brazilian traditional tax-free savings accounts. The full amount, however, can only be withdrawn upon timely high school graduation (3 years at regular schools). If the student fails to enroll at any grade at the following year, or repeats a grade, or has any criminal conviction, all the remaining balance is lost, including the 30% that could have been withdrawn.

At any moment, the student can check his balance, with the corresponding interests earned, and the amount that can only be withdrawn after graduating from high school.

This is a special and important feature of the program that was not included either in the Israeli nor the Colombian experiments. If students are loss avert, students would react more strongly to the incentive due to the "endowment effect".

The total amount at stake is sizeable. Students earn R$ 700 for passing 10th grade, R$ 900 for 11th grade and R$ 1000 for 12th grade. At a few vocational schools that include one additional year, passing the 13th grade yields an additional R$ 1200. In addition, graduating students earn an extra R$ 500 for performing above the national average at the National High School Exam (ENEM), the Brazilian equivalent for the SAT. For a beneficiary student enrolled in a regular 3-year high school, the total amount of R$ 3100 at stake is equivalent to more than 2.5 times the annual per capita income of their families, about two thirds of the annual minimum wage they would earn if employed in the formal sector and 43% of the income that they could potentially earn by joining drug tracking activities (Carvalho and Soares, 2013).

In order to fully participate in the program, students have to open a bank account in their name, at a branch that is designated by the bank. Opening the account involves a series of steps, from presenting documents at the school to signing a contract at the bank agency. Until signing the contract with the bank, the student cannot receive the award, even if is eligible and is promoted to the next grade. As no retroactive payments are allowed, if the account is not open until the date of the payment, the transfer is lost.

The series of steps required to open the account end up excluding some students from incentive scheme. In fact, less than a third of the eligible students end up opening the account. Indeed, until 2013, 53% of the eligible students who passed their grades and who could have received the award did not open their account on time and did not receive their award.

2.3. Data

I combine rich administrative datasets from the Secretariat of Education of Rio de Janeiro, the Secretariat of Social Assistance of Rio de Janeiro and the Ministry of Social Development containing individual student cores, enrolment, age, social background and participation in social programs including Bolsa Familia, Renda Melhor and Renda Melhor Jovem. Student flow outcomes comes from publicly available data at the school level on school passing rates, grade failure rates, dropout rates and grade-age distortion, released by the Ministry of Education.

Table 1 shows some characteristics of schools at 2010, the year before program Renda Melhor Jovem started to be implemented. Schools where the incentive scheme was implemented in 2011 are similar to the ones in which the program was implemented in 2012. Dropout rates were 16% and 17%, respectively, while grade passing rates were close to 65% for both and grade failure rates were about 20%. More than half of students at these schools were at least two years older than expected for their grade.

Control schools, where Renda Melhor Jovem was only implemented in 2013, had slightly better numbers on student flow, with 12% of dropout rates and 72% of pass rates. These schools also had lower age-grade distortion rates, smaller classes, and less students. Almost all schools have at least one TV, and more than 90% of schools have one meeting room for teachers and one copy machine.

2.4.Merging the data

The school enrolment records from the Secretariat of Education contain observations from all 1,432,387 students enrolled in regular public schools from 2010 to 2012. The administrative dataset from Renda Melhor Jovem program contains information on 58,883 students who were enrolled in regular public schools in 2011 and 2012 and who were eligible to receive the award, as well as their account status, i.e. if they had opened or not a bank account.

I start by matching the eligibility data and enrollment records. I managed to match 58,600 students by their Matriculation ID, and 194 students by their name, grade and school, successfully matching 99.84% of eligible students. Having matched these students, I calculate the proportion of eligible students for each school and grade, as well as the proportion of students who opened their accounts. Finally, I match this data with grade-school level data on grade passing rates, grade failure rates, dropout rates and grade-age distortion, as well as the information on school characteristics provided by the Ministry of Education.

2.5.Empirical strategy

Programs Renda Melhor and Renda Melhor Jovem were expanded at the same time to all municipalities, except for the city Rio de Janeiro, where program Renda Melhor was not implemented due to the existence of another municipal

program that already matched Bolsa Familia transfers up to Rio de Janeiro's poverty line (Cartão Familia Carioca).

Renda Melhor and Renda Melhor Jovem programs were first implemented in the three poorest municipalities of the Metropolitan Region of the capital of Rio de Janeiro in 2011. They were then expanded to 49 additional municipalities in 2012, distributed throughout the State, and by 2013 the programs were finally expanded to all the 92 municipalities in the State of Rio. In its first year of implementation, 5,638 students from regular schools were eligible for the incentive scheme in the 3 pilot municipalities. In its second year, 53,800 students were eligible, of whom 4,660 were from the 3 initial municipalities.

The phased in expansion of the two programs allows us to estimate their impact through a difference in differences strategy. Our main assumption is that the trends of student outcomes were similar before the program implementation.

Data on high school dropout and approval rates are available from 2007 to 2013, allowing us to check if trends are parallel. Figures 3 to 5 show the trends of drop out, passing rates and failing rates for treatment and control municipalities. All data come from INEP and correspond to the student situation by the end of the school year, in December.

In both the 3 cases, the evolution of outcomes is U shaped. For all outcomes, the patterns from treatment and control municipalities follow about the same trends until 2010. After 2010, though, there is a break on the trend of approval rates for schools located in the 3 pilot municipalities that received the program in 2011. At these municipalities, approval rates start to increase faster after 2011, relative to both the other 2 groups of municipalities. Relative to control municipalities, we also observe a faster increase of approval rates right after the second wave of municipalities receive the program in 2012.

This pattern of student outcomes improving right after the implementation of the program can also be observed for dropout and failing rates. This provides a visual check of the main assumptions underlying our differences-in-differences analysis, as well as a first visual evidence of the program effect.

In addition to explore the gradual expansion of the program across municipalities, we can also explore the variation of program coverage across schools from the same municipality. The proportion of eligible students for each school is determined only by the pre-existing proportion or poor students (according

to an unknown algorithm that predicts poverty) and by the implementation status of the program in the municipality where the school is located. Neither school participation nor the proportion of eligible students could be manipulated or anticipated by schools.

Therefore, in addition to the variation of the implementation of the program over time across municipalities, we can also use the variation on the proportion of eligible students across schools in the same municipality to identify the effect of the incentive. The proportion of eligible students at the school is zero before the introduction of the program, and then suddenly increases when the program is implemented at the municipality where the school is located.

We can thus measure the impact of the incentive award on student by estimating the following equation:

$$Y_{st} = \alpha_s + \mu_t + \rho Eligibles_{st} + \gamma X_{st} + \varepsilon_{st}$$

Where Y_{st} denotes the outcome variable (e.g., dropout rate) for school *s* ,at year *t*, $Eligibles_{st}$ is the proportion of eligible students at the school (i.e. the proportion of high school students whose predicted per capita family income falls below R$ 100 at municipalities where the program has been implemented) , X_{st} is a vector of controls at the school level (e.g., number of computers at school), α_s are school fixed effects, μ_t is a full set of year dummies, and ε_{st} are the unobserved school heterogeneity and idiosyncratic shocks. Notice that $Eligibles_{st}$is zero before the implementation of the program, and can vary between zero and one.[39]

Our main coefficient of interest is ρ, which indicates, on average, how a marginal change of the proportion of students for which the incentive scheme if offered changes student average outcomes at the school. The coefficient, consequently, can be thought as an intention to treat (ITT) effect of the program, since part of the eligible students end up not participating from the incentive scheme, as they do not open a bank account. Finally, as all regressions are estimated at the school level, the estimated impact captures spillover effects that can be present inside the school.

[39] In all regressions, I cluster the standard errors at the school level. Clustering at the municipal level leads to very similar results (Table 7).

2.6.Results

Table 2 shows the estimates of equation 1 for student dropout, approval rates and repetition rates. As suggested by the previous figures, the introduction of the program decreases dropout, increases passing rates and decreases failing rates. For each percentage point increase on the proportion of eligible students, average school dropout falls by 5.6 percentage points (significant at 5%), from a mean control average of 15 percentage points. If we were to extrapolate this result, the program could be responsible for reducing high school dropout by one third if all students became eligible.

Students are also being allowed to be promoted to the next grade more often. For each percentage point increase on the proportion of eligible students, I estimate a causal increase of 9.5 percentage point on passing rates. Students receive the transfer for each grade passed. Higher pass rates should lead to higher high school graduation rates.

The incentive scheme also decreases failure rates by 3.9 percentage points (significant at 10%). Lowering grade failure rates in a context where 18% of students fail every year is not only important per se, but might have several other consequences. First, combined with the reduction in dropout and the increase in passing grades, it decreases the amount of time students take to graduate. From an expenditures perspective, it means to decrease the cost spent for each student that ends up graduating from high school. Second, keeping constant the number of teachers and the existing facilities, at the school, repetition tends to increase class sizes. By lowering repetition rates, there is less need to hire extra teachers, and it is possible to have smaller classes. Third, repetition is shown to have a negative causal impact on educational attainment (Jacob and Lefgren 2009; Manacorda 2010). Consequently, in addition to the immediate impact on high school dropout, decreasing repetition rates today can also further decrease future dropout.

In fact, the table 3 shows a decrease of 13.5 percentage points on grade distortion rates (the proportion of students who are at least two years older than the expected age for their grade), from an average of 53%. The number of students per class at the following year is also reduced by 2.7, from an average of 30.9 students per class.

In Brazil, grade promotion is determined by a combination of tests that are written and graded by the teacher and student assignments. There is a great scope for subjectivity on retaining or not students (Botelho et al, 2006), especially those at borderline. Consequently, one could be worried about the possibility that, in the presence of the attainment award, teachers would simply promote beneficiary students with low grades. If that was the case, the program could increase graduation rates without increasing the human capital accumulation of beneficiary students. By promoting less prepared students to the next grade, test scores taken at the end of 12th grade should decrease, as the pool of tested students would now include less prepared students who otherwise would have dropped or would have stayed at lower grades.

Columns 3 and 4 of table 3 shows the effects of the program on Language and Math test scores from Rio's standardized test (SAERJ), taken by all students at the end of 12th grade. For both Language and Math scores, the impact of the program cannot be differentiated from zero.

Overall, all these estimates should be interpreted with some caution. First, while the impact is being identified from small changes in the proportion of eligible students across participating and non-participating municipalities, the coefficient shows the effect of changing the proportion of eligible students from 0 to 100% and only less than 5% of schools have more than 30% of eligible students. The impact of the program has at dropout rates of each school, computed for all eligible and non-eligible students, is thus smaller. In addition, in the presence of positive spillovers inside the school, the regression at the school level estimates a combination of the individual impact of the policy and of the social interactions(Glaeser et al., 2003).

2.7.Income or incentive?

As programs Renda Melhor and Renda Melhor Jovem were expanded to the same municipalities at the same time except for the capital of Rio de Janeiro, the question of whether 20 the above results estimate the effect of the incentive award or simply an income effect remains open. In the capital of Rio de Janeiro, the complementation of Bolsa Familia transfers is made through program Cartão

Família Carioca, which was launched in December 2010, and that was fully operative in 2011.

The incentive scheme from Renda Melhor Jovem was only expanded to the capital in 2012. The separate implementation of the cash transfer and the incentive scheme in allows us to investigate whether changes in student attainment are due to the incentive or to the income effect.

Let's assume that student outcomes are driven by the following equation:

$$Y_{st=}\alpha_s + \mu_t + Incentive(t1) + Income(t1) + Incentive(t2) + Income(t2) + \phi X_{st} + \xi_{st}$$

Where α_s are school fixed effects, μ_t are year fixed effects, $X_{st}$are time varying school characteristics, $Income(t)$ is the income effect after t years of implementation of the policy, and $Incentive(t)$ are the incentive effects after t years of implementation.

At the beginning of 2011, the municipal government of the capital of Rio implemented the program Cartão Família Carioca, a cash transfer that matched Bolsa Familia transfers up to Rio's poverty line. The program imposed no additional conditionality for Bolsa Familia beneficiaries who had a youth in high school, but families who had a child in primary school received an extra incentive for good grades at primary school. At about the same time, the State government implemented programs Renda Melhor and Renda Melhor Jovem in the 3 first municipalities. Consequently, families from students in both places received an income boost of roughly the same amount. If we compare the evolution of student outcomes until 2011 between these 3 municipalities and Rio's capital city, then according to equation 2 all the difference should be attributed to the incentive effect of given by RMJ program during its first year of implementation.

Column 1 of table 4 presents the results. We estimate an increase of 6 pp on passing grade rates, and a decrease of 5.8 pp on failure grade rates for the first 3 municipalities when compared to Rio's capital. That suggest that student flow started to improve for these municipalities in 2011, but not as much in Rio's capital when Cartão Família Carioca was introduced.

In order to isolate the incentive effect, we can also investigate the introduction of Renda Melhor Jovem program in the capital of Rio in 2012, comparing schools from control municipalities that had not receive the program in 2012. As the poor families from Rio's capital were already receiving the cash

transfer from Cartão Família Carioca in 2011, a change on student flow trend in Rio after 2012 should be either attributed to non-linear dynamic income effects from Cartão Família Carioca or to the incentive effect of Renda Melhor Jovem.

Column 2 of table 4 presents the results. We find no significant impact on dropout, but we estimate an increase of 1.8 pp on grade passing rates and a reduction of 2.2 pp on grade failure rates, significant at 10%.

Finally, we can estimate the income effect from the implementation of the Cartão Família Carioca. All we need are control municipalities that received no treatment and 2011, and check if high school student flow improves in Rio's capital, relative to the schools in control municipalities. Column 3 of table 5 presents the results from this exercise. We find no effect on student dropout. Surprisingly, we find a 2.5 pp decrease on pass rates and a corresponding increase in failure rates, both significant at 10%. While significance is not too strong, trying to explain the possible reasons for this result (as a intrahousehold relocation of resources towards younger kids) would be out of the score of this paper. From this exercise, however, we find no evidence that the introduction of additional cash to extreme poor households would benefit high school students. Lumped together, the 3 exercises suggest a very modest or even null contribution of income in explaining the observed impacts of Renda Melhor Jovem program among high schoolers. The strong incentive provided by the program, framed as a possible loss once the transfer is made to the student account, seems to play a strong role in making students stay in high school.

2.7.1. Robustness

In table 5 we test the robustness of the main results to the exclusion of controls or the inclusion of school specific linear trends. First column presents the estimates without including any control, while the second column presents our preferred specification that includes controls. The results for dropout rates and grade passing are roughly unchanged by the inclusion of controls, while the estimate for grade failure rates becomes significant. The inclusion of linear trends decreases the point estimates for the 3 main outcomes, but the estimated impact for grade passing rates and grade failure rates remain significant.

In table 6, I test for the clustering level and I include alternative proxies for grade passing, grade failure and dropout constructed from the school census. The promotion rate is the proportion of students who are matriculated in the next grade in the subsequent year. I also calculate the repetition rates, i.e. the fraction of students who are found in the same grade in the subsequent year. Finally, I calculate the fraction of students who are not found in the data following year. When compared to the results from tables 2 and 3, the level of clustering does not seem to make much a difference, while the alternative proxies for student flow show a negative and significant decrease on repetition rates.

One could also be worried that the First 3 municipalities are too different from the remaining ones, and that all our identification could come from a very specific set of municipalities. Table 7 shows that results are also fairly robust to the exclusion of the First 3 pilot municipalities.

Finally, I test for a placebo effect before the introduction of the Renda Melhor Jovem program. I estimate the effects for leads and lags of the dummy that indicates that the municipality participates in Renda Melhor Jovem program. The estimates are plotted in figures 6 to 8. Because of collinearity, though, one of the leads ends up not being estimated. The variance of the estimated effect of the program ends up increasing, and the estimated effect of the program encompasses zero for all the three outcomes.

2.8.Conclusion

This paper presents evidence that attainment awards that pay poor high school students for graduating from high school can substantially reduce dropout and increase high school graduation rates among economically disadvantaged youths. The program design exploits students' loss aversion by paying for each grade passed, but by locking the amount transferred in a bank account, and only releasing the full amount to the student after timely graduation from high school. This rule possibly creates a strong incentive for students who are loss averse, and is likely to influence students who highly discount the future or who present myopic behavior.

In our preferred specification, being eligible to receive the incentive award decreases dropout rates by roughly 37%, and increases grade passing rates by 14%. Eligibility, however, does not guarantee that the student can receive the award in

case he is grade promoted. Only about one third of students open their account, and half of students that could receive the transfers end up not receiving for not opening the account. This suggest that the program effect could be much higher if implementation issues were resolved and all eligible students were able to fully participate from the incentive scheme.

These effects are sizeable when compared to more traditional conditional cash transfers in which the family receives a monthly transfer conditional on the student enrollment and presence at school, especially in the context of urban areas. A recent meta-analysis of Conditional Cash Transfers on secondary school dropout (Saavedra, 2016) found an average reduction of 5 percentage points, similar to the effects found here. However, the average secondary enrollment rate on the countries with evaluations of Condition Cash Transfer is very low, of about 52%, and marginal gains are easier to achieve then in Rio de Janeiro. In Brazil, the expansion of Bolsa Família program to adolescents aged 16 and 17 was found to have positive and significant impact on enrollment and working decisions, but the results is fully concentrated on rural areas and is absent in the urban setting (Chittolina et al, 2016). The same expansion resulted in lower violence around schools with greater proportion of poor adolescents, but the main mechanism does not rise from increased time at school (Chioda et al, 2016).

The Latin American evidence of the effects of traditional conditional cash transfers casts doubts on the idea that universal high school attendance could be achieved by expanding traditional transfers to adolescents or by increasing the monetary value of the transfers. Indeed, Saavedra and Garcia (2016) find evidence that the impact of conditional cash transfers is unrelated to the generosity of payments or the frequency of payment. That suggests that factors other than liquidity constraints might influence youths' decision to drop out of school.

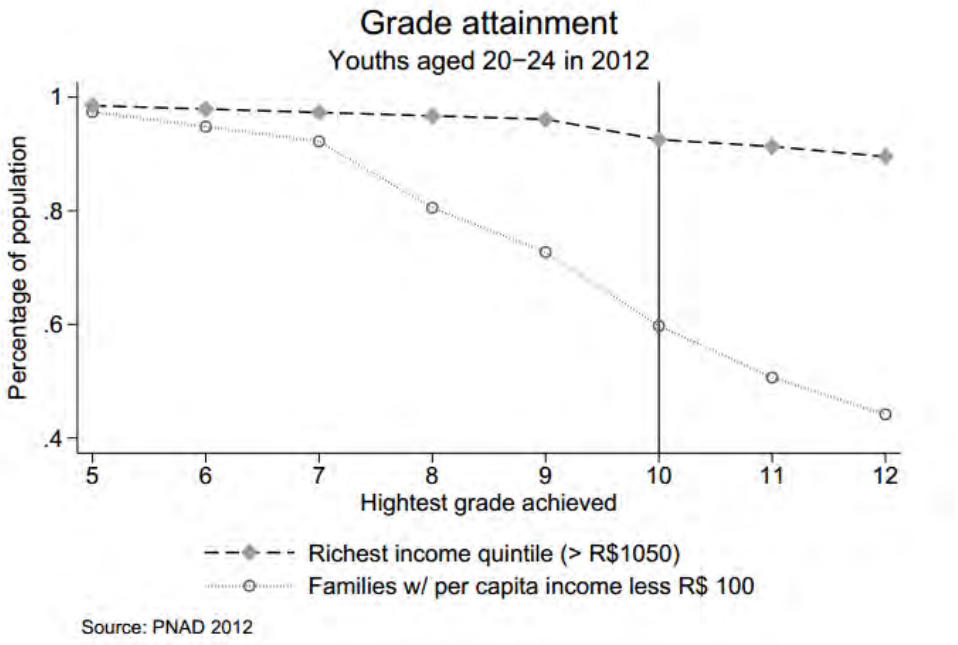
The literature from neuroscience shows that adolescence is a period of intense structural chance in the brain (Spear, 2000). The maturation changes in the brain during adolescence contribute to some of the behavioral differences from adolescence to other ages, as the higher predisposition to take risky attitudes and a tendency to focus on the present and ignore future consequences. In the presence of this myopic behavior, financial rewards for high school graduation can effectively decrease dropout. In the case of Renda Melhor Jovem program, the incentive can

avoid dropout by increasing the perceived returns of graduating from high school, or by making the importance of high school completion more salient to the student.

In a context where involvement with drugs, violence and risky sexual behavior can be a daily threat to the group of targeted students, the decrease in dropout rates induced by the program can have substantial effects on human capital accumulation and welfare to the society. These results indicate that attainment awards framed as a loss and targeted to poor students can be a promising way of reducing dropout rates among high schoolers.

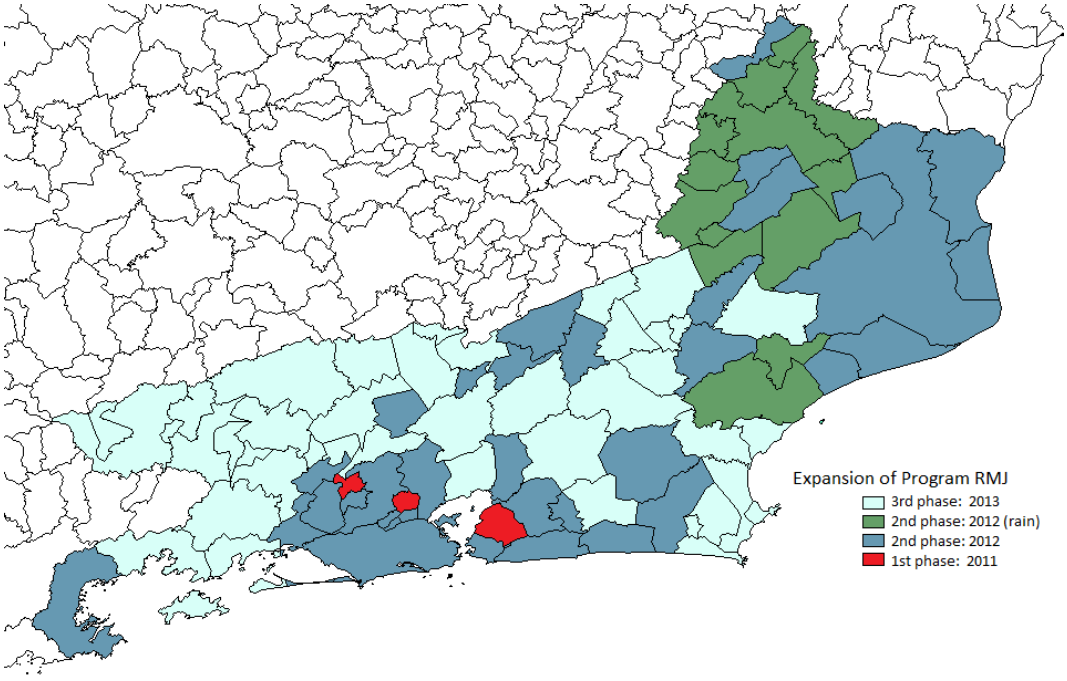
2.9. Figures

Figure 1-Grade attainment



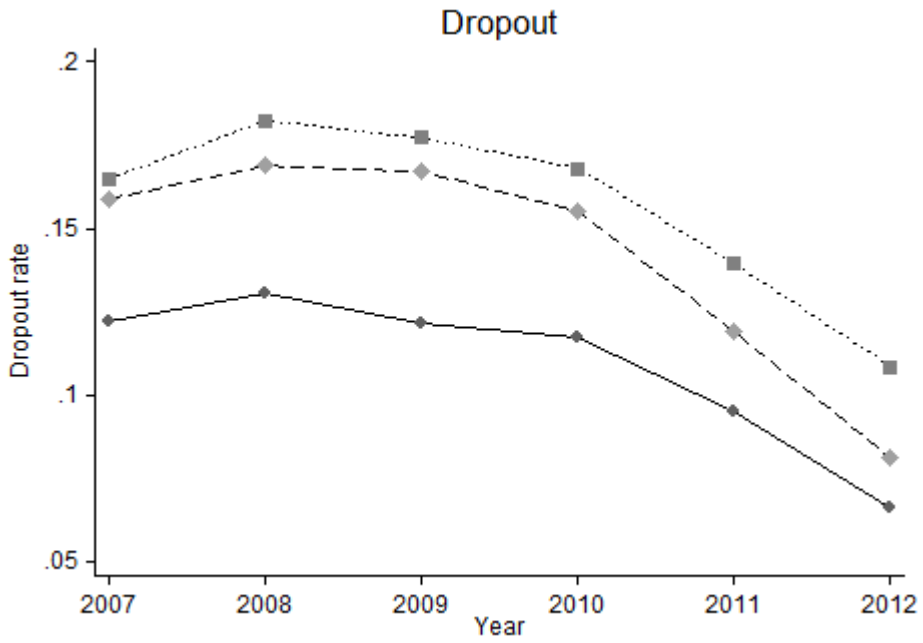
Notes: Individual data from PNAD 2012. Figure 1 shows the average grade attainment of individuals aged 20 to 24, living in the state of Rio de Janeiro, in 2012. Vertical axis presents the percentage of youths that reached each grade level represented in x axis. Dashed line represents individuals whose family per capita lies in the First quintile of income distribution, i.e, above R$ 1050 per month. Dotted line represents individual living under R$ 100 per capita per month. Vertical line at 10th grade indicates the start of high school

Figure 2- Program expansion



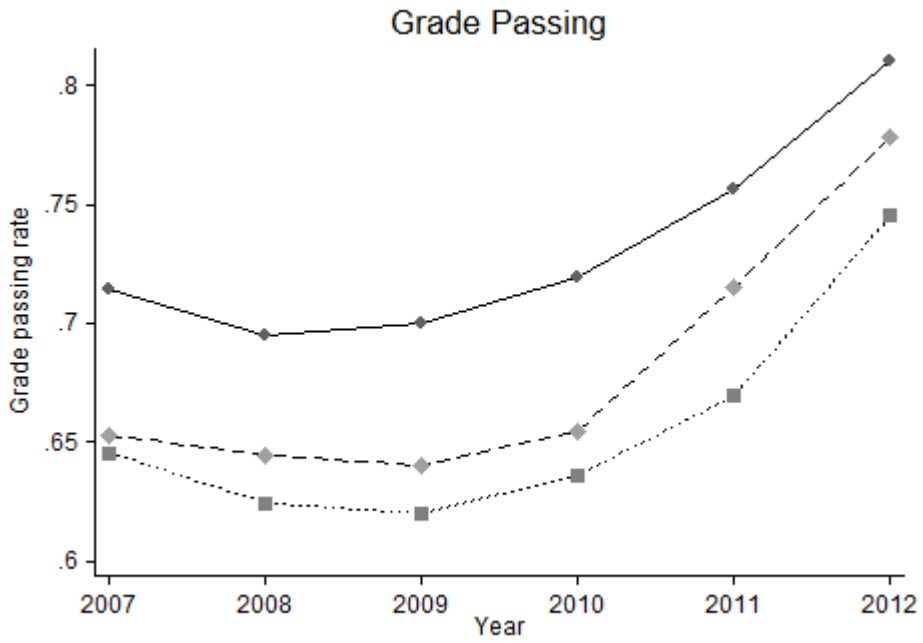
Notes: Figure 2 shows the expansion of program Renda Melhor Jovem across municipalities. Municipalities that received the program in 2011 are colored in red, municipalities that received the program in 2012 colored at dark blue, while municipalities in light blue received the program in 2013.

Figure 3-Dropout by year



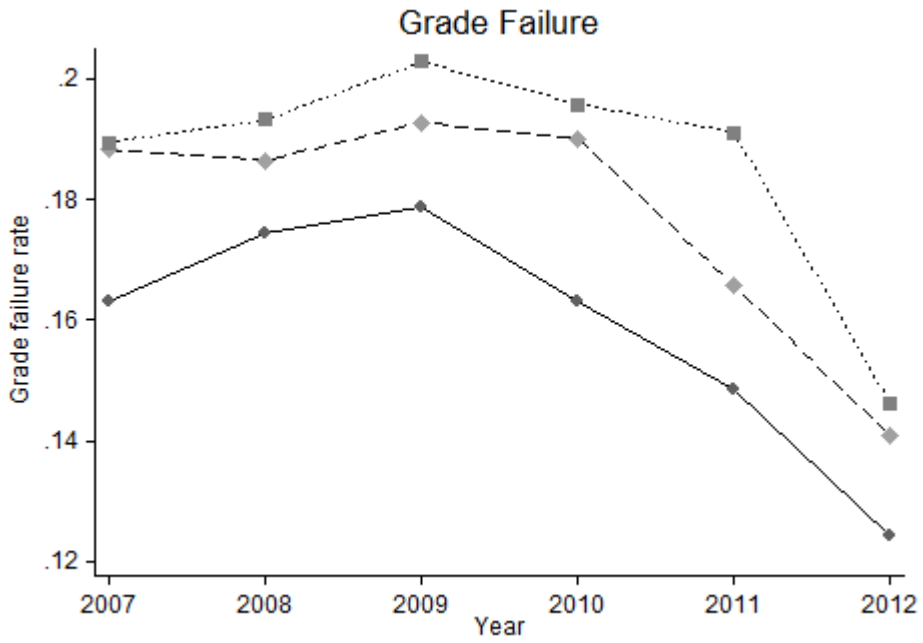
Notes: High school data from INEP, grades 10 to 12. Control municipalities only received the program in 2013, and are represented by the continuous line. Pilot municipalities that received the program in 2011 are represented by the dashed line, while municipalities that received the program in the First wave of expansion in 2012 are represented by the dotted line

Figure 4-Grade passing by year



Notes: High school data from INEP, grades 10 to 12. Control municipalities only received the program in 2013, and are represented by the continuous line. Pilot municipalities that received the program in 2011 are represented by the dashed line, while municipalities that received the program in the First wave of expansion in 2012 are represented by the dotted line

Figure 5-Grade failure by year



Notes: High school data from INEP, grades 10 to 12. Control municipalities only received the program in 2013, and are represented by the continuous line. Pilot municipalities that received the program in 2011 are represented by the dashed line, while municipalities that received the program in the First wave of expansion in 2012 are represented by the dotted line

Figure 6-Placebo-dropout

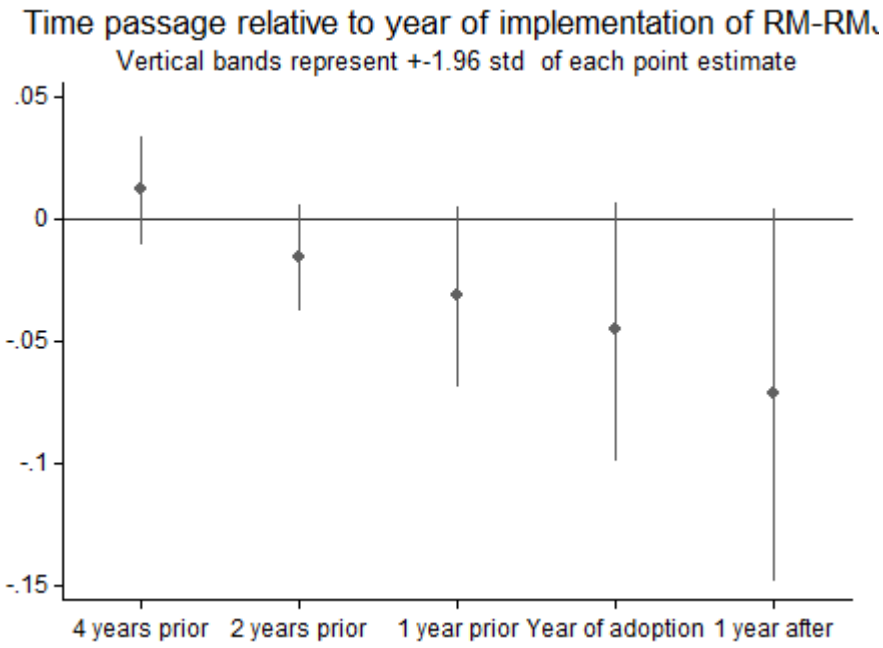


Figure 6 shows the coefficient of a regression of dropout rate on leads and lags of the treatment dummy that indicates if the municipality is a treatment one. Data from INEP at school level, grades 10 to 12. Standard errors clustered at school level. Controls include proportion of male students in school, proportion of students born outside the state of Rio, proportion of students born outside school municipality, proportion of students who live outside the school municipality, school has room for principal, teacher's meeting room, sciences lab, bathroom for girls, TV, DVD player, copy machine, printer, number of computers at the school, number of computers for school management, number of computers for students use, number of staff, school offers classes for adults, flags for missing data.

Figure 7-Placebo-grade passing

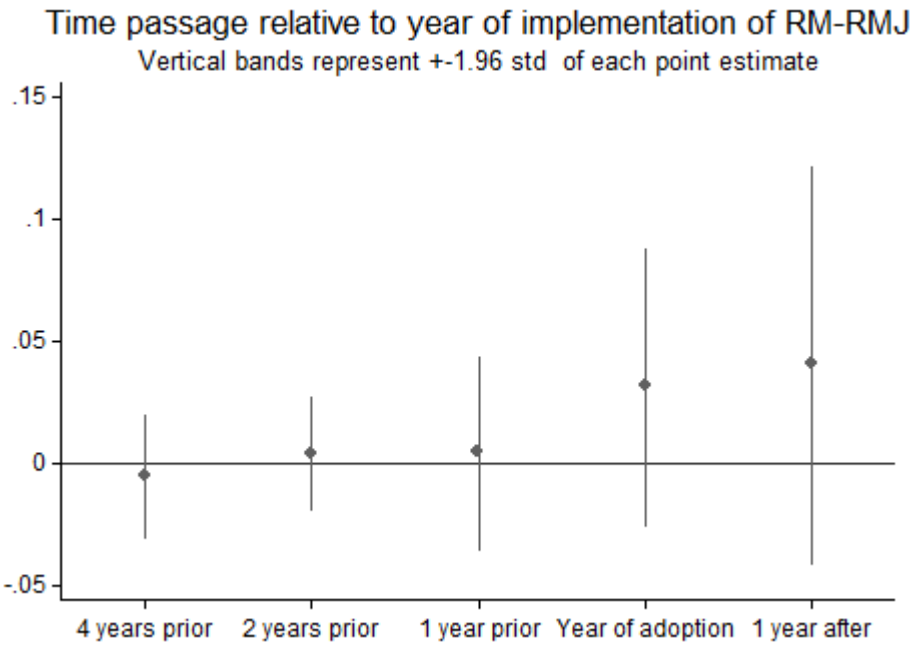


Figure 7 shows the coefficient of a regression of passing grades rate on leads and lags of the treatment dummy that indicates if the municipality is a treatment one. Data from INEP at school level, grades 10 to 12. Standard errors clustered at school level. Controls include proportion of male students in school, proportion of students born outside the state of Rio, proportion of students born outside school municipality, proportion of students who live outside the school municipality, school has room for principal, teacher's meeting room, sciences lab, bathroom for girls, TV, DVD player, copy machine, printer, number of computers at the school, number of computers for school management, number of computers for students use, number of staff, school offers classes for adults, flags for missing data.

Figure 8-Placebo-grade failure

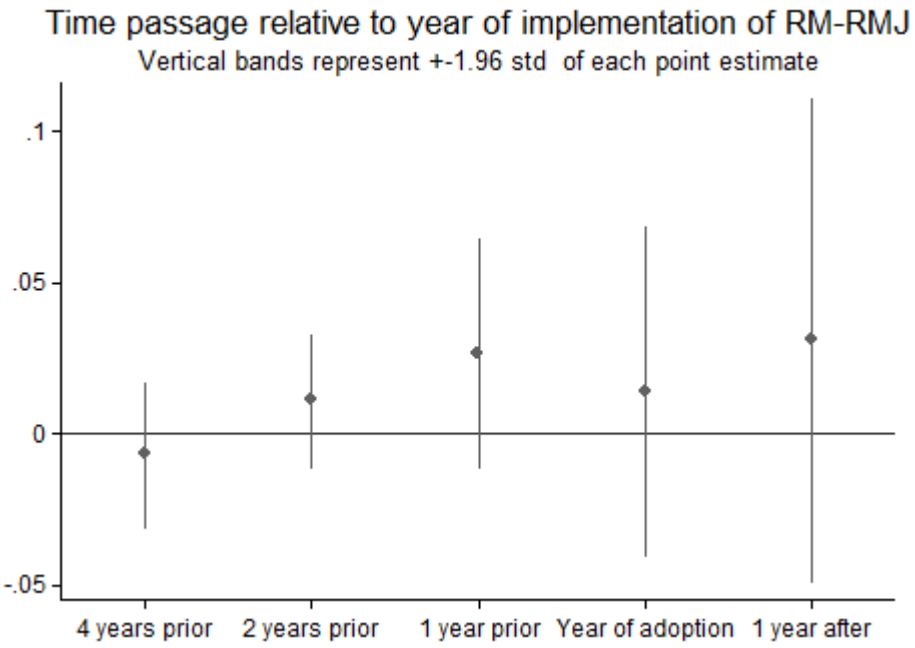


Figure 8 shows the coefficient of a regression of grade failure rate on leads and lags of the treatment dummy that indicates if the municipality is a treatment one. Data from INEP at school level, grades 10 to 12. Standard errors clustered at school level. Controls include proportion of male students in school, proportion of students born outside the state of Rio, proportion of students born outside school municipality, proportion of students who live outside the school municipality, school has room for principal, teacher's meeting room, sciences lab, bathroom for girls, TV, DVD player, copy machine, printer, number of computers at the school, number of computers for school management, number of computers for students use, number of staff, school offers classes for adults, flags for missing data.

Table 1- School characteristics (2010)

SCHOOL CHARACTERISTICS BY TREATMENT AND CONTROL AREAS (2010)

	Treated in 2011	Treated in 2012	Controls	Treated in 2011 x Controls		Treated in 2012 x Controls	
	n= 147	n=851	n=306	Diff	T-stat	Diff	T-stat
Dropout	0.16	0.17	0.12	0.04	2.65	0.05	6.07
Grade passing	0.65	0.64	0.72	-0.06	-4.29	-0.08	-8.64
Grade failure	0.19	0.20	0.16	0.03	2.43	0.03	4.80
Age-grade distortion	0.52	0.57	0.46	0.05	2.73	0.11	8.20
Number of students	899.29	874.83	690.21	209.07	3.73	184.62	5.16
Students per class	32.93	32.27	28.47	4.46	5.56	3.80	7.31
Male students	0.48	0.47	0.47	0.01	1.08	0.00	-1.15
Black students	0.15	0.17	0.11	0.03	2.80	0.05	7.37
School has teachers meeting room	0.94	0.93	0.94	-0.01	-0.24	-0.01	-0.80
School has TV	0.99	0.99	1.00	-0.01	-1.00	-0.01	-2.24
School has copy machine	0.95	0.91	0.95	-0.01	-0.24	-0.04	-2.76
Total staff (including teachers and principal)	77.22	78.84	71.20	6.02	1.35	7.63	2.56
School has high speed internet	0.87	0.84	0.84	0.03	0.89	0.01	0.21

Note: School data from Inep. All data from 2010. Controls are the schools that were treated in 2013. Mean differences calculated via OLS regression, clustering standard errors at the school level.

Table 2- Impact on student flow

EFFECTS ON DROPOUT, GRADE PASSING AND GRADE FAILURE

	(1)	(2)	(3)
	Dropout	Grade passing	Grade failure
Prop students eligible RMJ	-0.056**	0.095***	-0.039*
	(0.026)	(0.027)	(0.023)
Observations	6,328	6,328	6,328
Control Average:	0.150	0.666	0.184
Control Standard Deviation	0.121	0.149	0.111
Year Fixed Effects	✓	✓	✓
School Fixed Effects	✓	✓	✓
Controls	✓	✓	✓

Note: All regressions at school level. Data from INEP. Standard errors clustered at school level. Controls include the proportion of male students in school, proportion of students born outside the state of Rio, proportion of students born outside school municipality, proportion of students who live outside the school municipality, school has room for principal, teacher's meeting room, sciences lab, bathroom for girls, TV, DVD player, copy machine, printer, number of computers at the school, number of computers for school management, number of computers for students use, number of staff members, school offers classes for adults, flags for missing data.

Table 3- Impacts on student age, class size and test scores

EFFECTS ON AGE, CLASS SIZE AND TEST SCORES

	(1) Age-Grade Distortion (t+1)	(2) Students per class (t+1)	(3) Language test score	(4) Math test score
Prop students eligible RMJ	-0.135***	-2.726**	-0.246	0.119
	(0.026)	(1.318)	(0.168)	(0.150)
Observations	6,342	6,353	3226	3238
Control Average:	0.532	30.896	0.007	0.010
Control Standard Deviation	0.203	8.260	0.537	0.514
Year Fixed Effects	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Controls	✓	✓	✓	✓

Note: All regressions at school level. Age-grade distortion data and class size data from INEP school census. A student is considered to be grade-distorted if is at least two years older than the expected age for his grade. Standardized test scores from SAERJ/SEEDUC-RJ. Test scores applied to 12th graders. Data on test scores only available for 2010, 2011 and 2012. Standard errors clustered at school level. Controls include the proportion of male students in school, proportion of students born outside the state of Rio, proportion of students born outside school municipality, proportion of students who live outside the school municipality, school has room for principal, teacher's meeting room, sciences lab, bathroom for girls, TV, DVD player, copy machine, printer, number of computers at the school, number of computers for school management, number of computers for students use, number of staff members, school offers classes for adults, flags for missing data.

Table 4- Income or Incentive?

INCOME OR INCENTIVE EFFECT?			
Effect:	Incentive (t1)	Incentive (t1)	Income (t1)
Treatment:	First 3 municipalities* (year=2011)	Rio's capital city * (year=2012)	Rio's capital city * (year=2011)
Control:	Rio's capital city	Controls (only received RMJ in 2013)	Controls (only received RMJ in 2013)
Years:	2007-2011	2007-2012	2007-2011
Dropout	-0.001	0.004	0.000
	(0.030)	(0.010)	(0.014)
Grade passing	0.060**	0.018*	-0.025*
	(0.025)	(0.009)	(0.013)
Grade failure	-0.058**	-0.022*	0.025*
	(0.027)	(0.011)	(0.014)
Observations:	1,847	3,054	2,549
Year fixed effects	✓	✓	✓
School fixed effects	✓	✓	✓
Controls	✓	✓	✓

Notes: All regressions at the school level. Standard errors clustered at school level. Controls include the proportion of male students in school, proportion of students born outside the state of Rio, proportion of students born outside school municipality, proportion of students who live outside the school municipality, school has room for principal, teacher's meeting room, sciences lab, bathroom for girls, TV, DVD player, copy machine, printer, number of computers at the school, number of computers for school management, number of computers for students use, number of staff, school offers classes for adults, flags for missing data.

Table 5- Robustness to inclusion of controls and trends

SENSITIVITY TO INCLUSION OF CONTROLS AND SCHOOL LINEAR TRENDS

	No Controls	Including Controls	Controls + Linear trend	Control Avg and Std
Dropout	-0.061**	-0.056**	-0.031	0.150
	(0.026)	(0.026)	(0.023)	0.121
Passing grade	0.091***	0.095***	-0.039*	0.666
	(0.026)	(0.027)	(0.023)	0.149
Failing grade	-0.002	0.113***	-0.083**	0.184
	(0.007)	(0.039)	(0.033)	0.111
Year Fixed Effects	✓	✓	✓	
School Fixed Effects	✓	✓	✓	
Controls		✓	✓	
School Linear Trend			✓	

Notes: All regressions at school level. Data from INEP. Standard errors clustered at school level. Controls include the proportion of male students in school, proportion of students born outside the state of Rio, proportion of students born outside school municipality, proportion of students who live outside the school municipality, school has room for principal, teacher's meeting room, sciences lab, bathroom for girls, TV, DVD player, copy machine, printer, number of computers at the school, number of computers for school management, number of computers for students use, number of staff mmbers, school offers classes for adults, flags for missing data

Table 6 -Robustness to clustering and alternative proxies

ROBUSTNESS TO CLUSTERING LEVEL AND ALTERNATIVE PROXIES OF STUDENT FLOW						
	Dropout	Passing grade	Failing grade	Dropout (next year)	Promotion Rate	Repetition rate
School level data, cluster at municipal level						
Treated Municipality*Post	-0.063*	0.105***	-0.042**	-0.011	0.044	-0.033**
	(0.032)	(0.032)	(0.019)	(0.036)	(0.041)	(0.015)
Observations	6,364	6,364	6,364	6,231	6,231	6,242
Control Average:	0.150	0.666	0.184	0.325	0.514	0.139
Control Standard Deviation	0.121	0.149	0.111	0.133	0.151	0.075
Year Fixed Effects	✓	✓	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

Notes: All regressions at school level. Data from INEP. Standard errors clustered at municipal level. Controls include the proportion of male students in school, proportion of students born outside the state of Rio, proportion of students born outside school municipality, proportion of students who live outside the school municipality, school has room for principal, teacher's meeting room, sciences lab, bathroom for girls, TV, DVD player, copy machine, printer, number of computers at the school, number of computers for school management, number of computers for students use, number of staff mmbers, school offers classes for adults, flags for missing data

Table 7- Robustness to exclusion of municipalities

EXCLUDING FIRST 3 MUNICIPALITIES

	Dropout	Passing Grade	Failing Grade
Prop students eligible RMJ	-0.054*	0.079***	-0.025
	(0.029)	(0.028)	(0.027)
Observations	5,560	5,560	5,560
Number of pk_cod_entidade	981	981	981
Year Fixed Effects	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
School Linear Trend	No	No	No
Control Average:	0.151	0.664	0.185
Control Standard Deviation	0.119	0.148	0.111

Notes: All regressions at school level. Data from INEP. Standard errors clustered at municipal level. Controls include the proportion of male students in school, proportion of students born outside the state of Rio, proportion of students born outside school municipality, proportion of students who live outside the school municipality, school has room for principal, teacher's meeting room, sciences lab, bathroom for girls, TV, DVD player, copy machine, printer, number of computers at the school, number of computers for school management, number of computers for students use, number of staff mmbers, school offers classes for adults, flags for missing data

3
Can students benefit if teachers lose their bonus? Behavioral biases inside the classroom

3.1. Introduction

Motivating public workers to perform and deliver high quality services is one of the main challenges of modern governments, and an extensive literature has devoted attention to the use of incentives to improve the quality of service delivery (Bandiera, Barankay, and Rasul, 2013; Banerjee, Glennerster, and Duflo ,2008; Glewwe, Illias, and Kremer, 2010; Khan, Khwaja, and Olken, 2014; Muralidharan and Sundararaman, 2011; Olken, Onishi, and Wong, 2014[40]). Most of this literature, however, is based on traditional models whose main assumptions have recently been questioned by a series of behavioral deviations usually found in laboratory experiments (see Della Vigna, 2009 for a review of the field evidence).

One of these empirical findings is that people react to being behind a goal or a point of reference (Kahneman and Tversky, 1979), by either reducing productivity when disappointed (Mas, 2006), or by exerting extra effort (Berger and Pope, 2001; Pope and Schweitzer, 2006) when slightly behind their goal. However, the question of whether these behavioral biases can improve motivation and performance in other domains of public policy remains largely unanswered.

If being slightly behind a goal really motivates workers to exert more effort, then it might be possible to design public policies that combine incentives, targets and the strategic release of information in order to improve the quality of public services. The field of education, in which pay for performance schemes have been increasingly implemented, is especially suitable for experimentation with these policies (Koch et al, 2015).

In this paper, we investigate whether losing a pay for performance bonus by small margin motivates teachers to exert more effort. We explore a discontinuous rule from a teacher incentive scheme in the state of Pernambuco, Brazil, that

[40] See Finan, Olken and Pande (2015) for review of the literature.

determines that only teachers from schools with higher than expected performance gains receive the bonus. Although there is no treatment per se and winning or losing the bonus brings no new information about school performance once the performance index is known, we find that students from schools slightly behind the bonus threshold do score significantly higher in Portuguese and Math at the state standardized exam only 6 months after the release of the list of bonus winners and losers.

We also investigate some of the possible mechanisms behind these results. We find that teachers from schools that barely failed to win the bonus change their pedagogical practices in a number of ways. Teachers assign and grade more homework, for example. We were also able to investigate changes on teacher practices from the point of view of the students, who report that teachers correct more homework and are more attentive to students' opinions. On the other hand, we find no changes on principal management practices, neither on measures of teacher cooperation nor trust.

These results are in stark contrast with those from Alexandre Mas (2006), who finds that effort from police members depend on the degree in which salary raises fall below their expectations. By contrast, when pay raise exceeds expectations, there is no effect on police officers' performance, which would be an indication of police officers being loss averse. It could be the case, for example, that teacher's morale would decrease after losing the bonus and demotivate teachers to exert effort in the subsequent periods. Disappointment could thus create a discouragement effect and decrease student scores.

Our results, however, go on the same direction of previous studies that investigate how individuals exert higher effort to avoid losses, especially when they are below a reference point. Physicians who face shortfalls from their reference income take actions to boost their income (Rizzo and Zeckhauser, 2003), golf players try harder if they are at risk of coming worse than a par[41] (Pope and Schweitzer, 2011) and basketball players have higher chances of winning a game if they end the first half one point behind (Berger and Pope). In all these cases, being behind the reference point induces more effort.

[41] In golf, a par is a predetermined number of strokes in which a golfer should complete a hole.

We attribute our results to loss aversion (Kahneman and Tversky, 1979). After the release of the bonus results, teachers from schools that almost won the bonus are suddenly placed at the loss side relative to the bonus allocation threshold, and exert more effort than teachers who barely won the bonus in order to avoid missing out again.

We also investigate whether teachers change their behavior relative to other candidates for reference points. The rules of the pay for performance scheme, for example, allow some schools to receive the bonus even if the school presents no improvement. Teachers from schools that are required to improve (loss side) make no additional effort. We find no evidence that teachers take into account other reference points that are important when strategically deciding how to allocate effort to maximize the expected bonus value.

We argue that winning or not the bonus increases the salience of the bonus allocation threshold. Teachers that barely won the bonus suffer the emotional consequences of losing and change their expectations about future school performance, consistent with the theory of rational choice of reference points from Koszegi and Rabin (2006). Fearing to lose the bonus again, they change their pedagogical practices and exert more effort.

Our paper contributes to the literature linking behavioral economics and public policy by showing strong reactions from almost winning the bonus. As pointed by List (2003) and Levitt and List (2008), biases would likely to be extinguished by large stakes, competition and repeated exposure to situations in which biases could arise. We present evidence that these biases are also present in the field of education. Consistent with the skepticism with the existence of behavioral anomalies in the long run, though, we find that the results are concentrated among schools that have never previously failed to win the bonus by a short margin. Teachers from schools that almost won the bonus in previous years see no improvement on its students' test scores, which indicates that in the long run schools would stop reacting to not winning the bonus.

Our paper also relates to Fryer et al (2012) who have shown that teacher incentive schemes can be enhanced by paying teachers in advance and then asking teachers to give back the money if their students fail to improve. This "endowment effect" (Kahneman, Knetsch and Thaler, 1991) is a consequence of teachers' loss

aversion. We present additional evidence of teachers' loss aversion and document positive effects on children.

Our paper also relates to Ahn and Vigdor (2014), who study discontinuities in teacher bonus allocation in North Carolina. They find an increase on math test scores of students from schools that almost won the bonus, but without making reference to Prospect Theory (Kahneman and Tversky, 1979). Our paper presents some advantages in investigating teachers' reactions to almost winning the bonus. First, contrary to the North Carolina context, there are absolutely no sanctions for schools that fail to win the bonus and no room for confusion over the bonus rules. Second, we have data on teacher attitudes that allows us to assess whether teachers changed their practices in reaction to failing to win the bonus. We can thus better describe the mechanisms behind changes in student performance.

This paper is organized as follows: in the first section, we describe the pay for performance policy in Pernambuco. The second section describes the data used on our analyses. In section 3, we document our basic results and we check for their validity. In section 4, we investigate the mechanisms of our main results. Section 5 discusses and interprets the findings, while in section 6 we conclude.

3.2.Institutional background

In 2007, Pernambuco was ranked second to last among the 27 Brazilian States in the national index of educational quality for late primary education. Aiming to overcome this scenario, the Government of the State of Pernambuco set explicit and clear goals for key educational outcomes. The pillar of the actions designed to achieve those goals was the implementation of an ambitious accountability system, rooted on the application of annual standardized tests, the release of an annual index of school quality and the implementation of a pay for performance policy.

Pernambuco's pay for performance policy gives an annual bonus to all teachers from schools that meet the educational targets set by the State Department of Education. Targets of student achievement are set for some grades and subjects, nonetheless all teachers from the school receive the bonus. If a school meets its target, all teachers receive the full value of the bonus. Yet, teachers can still receive part of the bonus if at least half of the school target is met. Teachers from schools

that achieve 50% of the target and less than the full target or more receive a proportional fraction of the full bonus[42].

The amount earned by each teacher depends only on the proportion of the school target achieved and on the quantity of hours worked at the school. The amount is exactly the same for teachers of tested and non-tested subjects. The bonus is deposited directly on teachers' individual bank account without any mediation from principals or state bureaucrats.

Teachers do not know beforehand the value of the bonus they can receive: The Government of the State of Pernambuco allocates the equivalent to at least one month of the teachers' payroll to be integrally spent on teacher bonuses every year. But as the bonuses are paid only to teachers from schools that achieve at least 50% of their targets, the money left over from schools that do not meet their full target is redistributed, increasing the bonus value. Thus the amount received for each teacher depends not only on the achievement of her own school, but also on the performance of all other schools. Since 2008, the value of the bonus varied from 1.5 to 3 monthly wages.

Targets combine both a measure of student progression on all grades within a segment (early primary, late primary and high school) and student scores on standardized tests (at the last grade of each segment: 5th, 9th and 12th grades), with the objective of making teachers and school principals internalize the trade-off between retaining the worst performing students before they reach the tested grades and raising the average score of its students[43].

Targets are set on the scale of Pernambuco's Index of Educational Quality (IDEPE), which is simply the product of approval rates and average test scores, within each segment (IDEPE=SAEPE*Pass rate). The approval rate for a segment is the average from all the grades in that segment: 1st to 5th grade in early primary, 6th to 9th grade in late primary, and 10th to 12th grade in high school. Standardized tests are taken only at the final grade of each segment. The index is calculated for

[42] Pernambuco's bonus design, with collective bonuses, growth targets and winning thresholds, is very similar to the one implemented in North Carolina since 1996/1997 academic year (Vigdor, 2009).

[43] Student retention was still a major problem by 2007. In the 9 grades of primary education, 17% of students were retained on average, while 13% were dropping out of school. With such high retention and dropout rates, a student starting first grade had only a 28% chance of being in school after 9 years, and would have only an 18.6% chance of successfully completing primary education on time.

each subject and segment and then weighted by the proportion of students on each segment to form the school quality index, also known as the Global IDEPE.

The standardized tests, known as SAEPE (Pernambuco's State learning assessment), are applied at the end of school year (November-December) by a specialized firm that hires external evaluators. The tests only test students' Portuguese Language and Mathematics' skills. The first exams were applied at the beginning of the 1990s, but only reached its current format, graded by item response theory in the same scale of the similar national exam called "Prova Brasil" in 2005. Starting in 2008, the tests were administered every year. Additionally, students from 3rd grade are tested, but their scores are not counted in any high stakes index.

Formally, the Global IDEPE index can be written as:

$$IDEPE = \sum_{c} w_c\left(\frac{SAEPE_c(Lang) + SAEPE_c(Math)}{2} * Pass_c\right) \qquad (1)$$

where c stands for segment (1st through 5th, 6th through 9th, 10th to 12th), Pass are average approval rates for each segment and weights w_c are the proportion of enrolled students at each segment.

School targets are set yearly by the Department of Education based on previous levels of each segment-subject IDEPE index. In 2008, the first year of the pay for performance scheme, targets were set for each school-segment-subject based on the last available data on test scores and approval rates dating from 2005. From 2009 on, targets were set taking into account the 2 previous indexes, in an attempt to smooth measurement error from test scores. School targets are set such that schools with lower past indexes are required to make bigger improvements.

Principals receive the targets at the beginning of the second semester in July-August, after the release of bonus winners, IDEPE indexes and test scores for the previous year, which usually happens by the end of the first semester in May or June. After receiving the targets, principals sign a contract with the Department of Education in which they state that they will make an effort to achieve them. There is no clear sanction for failing to achieve the targets besides not getting the bonus.

For schools that present improvement in all IDEPE indexes, the percent of the target achieved is calculated by simply dividing the weighted improvement of the indexes by the improvement targeted. If a school presents a decrease on some index, though, the decrease is disregarded. This generates a lower bound for the

percentage achieved of the target, at zero. Formally, the value of the bonus can be written as:

$$Bonus_{is} = y_i * m * \mathbf{1}(Perc_s) \tag{2}$$

Where

$$Perc_s = \frac{\sum_c w_c * \max(0, \Delta IDEPE_c)}{\sum_c w_c * Target(\Delta IDEPE_c)} \tag{3}$$

$$\mathbf{1}(Perc_s) = \begin{cases} 0 \text{ if } Perc_s < 0.5 \\ Perc_s \text{ if } Perc_s \geq 0.5 \text{ and } Perc_s < 1 \\ 1 \text{ if } Perc_s \geq 1 \end{cases} \tag{4}$$

$$m = \frac{\sum_s \sum_i y_{is}}{\sum_s \sum_i y_{is} * \mathbf{1}(Perc_s)} \tag{5}$$

and where s denotes the school, i the teacher, c the segment, w_c are weights given by the proportion of students on each segment, y is the teacher's wage, m is a redistribution factor that increases the value of the bonus according to the result of the tournament. 1(Perc_s) is a function of the percentage of the target achieved by the school, which indicates that no teacher receives the bonus if the achievement is inferior to half of the target, that teachers receive a fraction of the bonus if the target is not fully met, and the full bonus is received by teachers if the target is met. There are no sanctions, rewards or any intervention for schools failed or to schools that met its targets.

Table 1 shows the value of the bonus for each year. In 2008, the first year of the new policy, 51% of the schools achieved at least 50% of their targets. One month's payroll was allocated toward the bonus. After redistributing the budget from the schools that did not fully meet their targets, the redistribution factor was set at 2.31, which meant that each teacher working in schools that achieved their global target received 2.31 times their monthly wage as the bonus. The mean value of the bonus was R$1,661 (approximately USD$ 790). The payment of the bonus relative to scores and approval rates from 2008 happened in July 2009. Until soon

before the payment, no one knew exactly how much each teacher would receive or which schools would achieve their targets.

In 2009, 81% of schools met at least half of their global target, and teachers were paid in May 2010. In 2010, 70% of schools met their targets. Teachers were paid in July 2011. Each teacher whose school fully met its target received a bonus corresponding to 1.88 of one monthly wage.

3.3. Data

We identify the effect of missing the bonus by a small margin by using student data provided by the Secretariat of Education of Pernambuco. Our main data are the student test scores in math and language from SAEPE, the high stakes exam that is applied to all students from 5^{th}, 9^{th} and 12^{th} grades the last grades of each segment. Tests are taken by the end of the school year, usually by the month of December. When combined with student pass rates, these high staked exams determine which schools receive the teacher bonus. We make use of SAEPE test scores from 2008 to 2011.

We also investigate some possible mechanisms for changes in student test score by using the annual school survey administered by Pernambuco's Secretariat of Education jointly with SAEPE tests. The survey is filled by principals and teachers and students from tested subjects and grades. Questions include some teachers' pedagogical practices, like frequency of homework assignments, usage of textbooks, whether teachers reflect upon the results of standardized tests or whether teachers allocate extra time to students with learning difficulties. Teachers are also surveyed about the cooperation of their colleagues, about their trust in their peers and about the leadership of the principal. By its turn, students are asked to fill a brief survey on their daily learning activities, such as frequency of homework, help from parents when doing homework, as well as some teachers' practices, as assigning homework, correcting homework in class and teachers' absenteeism.

3.4.Empirical strategy

We estimate the effect of missing the bonus threshold by a small margin through a regression discontinuity (RD) approach, exploiting the fact that teachers only receive the bonus if their schools make 50% or more of their global target.

This rule creates a sharp discontinuity rule for bonus receipt. We can thus compare teacher characteristics from schools that barely lost the bonus to schools that achieved slightly more than half of their global target. In the presence of loss aversion, teachers from schools that almost achieved their global target would exert higher effort than teachers from schools that just made it.

The measure of the percentage of the target achieved combines student approval rates and students' test score gains. As Kane and Staiger (2002) point out, test scores provide a noisy measure of school performance. Sampling variation of students, the presence of disruptive students in class, noise during test taking and bad weather can all affect the average school scores, while changes in mean test scores from one year to the next are measured even more unreliably. Schools cannot predict, anticipate or manipulate the indicator of their achievement. Consequently, in the small window around the discontinuity cut-off, the allocation of bonus winners and losers entails a great amount of luck.

We investigate the regression discontinuity effect of not winning the bonus on student test scores through a regression discontinuity approach. Conceptually, our approach can be thought of as we were estimating the following equation through OLS:

$$y_{it} = \alpha + \beta\, Win_{it} + \gamma h(Margin_{it}) \times Win_{it} + \theta_t + \varepsilon_{it} \qquad (7)$$

Where i indexes the schools, t indexes time, Win is a dummy equal to one if teachers from the school receive the bonus, Margin is the margin of victory, i.e. the percentage of the target achieved minus 50%. h( ) is a flexible control function of margin of victory, allowed to differ for each side of the discontinuity.

In practice, we adopt a non-parametric approach and estimate the effect using local linear regression. In our preferred specification, we use the bandwidth proposed by Calonico, Cattaneo and Titiunik (2014) (hereafter referred as CCT). We then check the robustness of our estimates by employing the bandwidth suggested by Imbens and Kalyanaraman (2012) (hereafter referred as IK), and by estimating the impact through OLS.

3.5.Results

3.5.1.Impact on students' test scores

We start by investigating the effect of missing the bonus by a small margin on student test scores at the end of the year. Usually the results from SAEPE and the bonus results are released by the year and teachers are paid right after the release. For the 2008/2009 bonus, for example, students were tested in December 2008 (the school year runs from February to December). The results from the bonus were released by July of 2009, and a new cohort of students was tested in December 2009. We are thus estimating the effect on students' test scores in December 2009 less than 6 months after the release of the bonus results.

Conceptually, the information of whether a school made more or less than 50% of its target should be irrelevant once teachers know exactly the percentage of the target achieved. A teacher from a school that made 51% of its target should believe that the chances of her school winning the bonus again next year should be the same as those from a school that made 49% of the target.

Figures 1 and 2 show a local polynomial estimate of language and math SAEPE test scores on our running variable, which is the margin of bonus victory (the percentage achieved of the global target minus 50%). SAEPE test scores from December 2009, 2010 and 2011 were all pooled together. We observe a positive relationship between the running variable and test scores (which is not obvious since targets are smaller for low performing schools). Visually, schools that barely missed the bonus seem to have higher test scores than schools that won by a small margin.

Table 2 confirms the pattern plotted in figures 1 and 2. Schools just below the bonus discontinuity threshold present higher test scores when compared to schools barely won the bonus. By pooling all tested grades for 2009, 2010 and 2011, we find an RD impact of 0.134 sd (significant at 10%) for language, and 0.119 sd (significant at 10%) for math. The estimated effects are sizeable. These effects are at about the same magnitude of the average annual impact of placing one student in a charter school in the US (Fryer, 2016), but by treating the students for only up to 6 months. These effects are also three to four times larger than the effects of losing the bonus by a small margin in North Carolina (Ahn and Vigdor, 2014).

Table 2 also separates the impact by each one of the three grades tested. While the point estimates for the impacts on 5th grade test scores are similar to the ones for 9th grade, they are not statistically different from zero. Meanwhile, we find a 0.133 sd impact for 9th grade math and 0.146 sd impact on 9th grade language, both significant at 5%. For 12th grade, we find a 0.096 sd impact on math (significant at 10%), and no significant impact on language.

We next analyze the impact of barely missing the bonus targets for each one of the 3 first years of bonus implementation. Figures 3 to 8 show a graphical representation of the RD impact, while table 3 shows the estimated effects and standard errors. We see no evidence of impact for 2009, a small gap in 2010 and a clear impact for 2011. The figures for 2011 shows a local spike right below the bonus threshold, while levels of test scores seem similar when not very close to the threshold. For 2011, we estimate an impact of 0.480 sd for language and of 0.429 sd for math, both significant at 5%. These effects are surprisingly large, especially when taking into account the short time between the announcement of the bonus results and SAEPE exam. When using a larger bandwidth proposed by Imbens and Kalyanaraman, we estimate an effect of 0.160 for math (significant at 5%) and 0.190 for language.

3.5.2.Validity tests

We performed a series of tests to ensure the validity of the assumptions underlying our exercises. First, we start investigating the assumption that schools that barely missed 50% of their target are similar to the schools that reached the 50% mark. We should expect that, at the year of reference for the bonus, school characteristics should be similar on average. In particular, we should expect those characteristics to be continuous around the bonus threshold.

Table 4 confirms this hypothesis. We test for jumps in school characteristics at the bonus allocation threshold, for different specifications. We find no discontinuity on students' age, the proportion of black students, school size and average schooling of teachers. We find only a significant difference for school size when employing a OLS specification containing 3rd and 2nd order polynomials. All other estimates, for all other specifications, cannot be statistically distinguished from zero.

We also performed placebo tests, replacing our main outcomes on test scores for past test scores. If there is a causal effect of not winning the bonus on student learning, we should not observe any effect around the threshold in previous years. In fact, the last two panels of table 4 confirm our conjecture. We find no significant impacts of just missing the bonus threshold when we use test scores from the previous year as the dependent variable.

We test whether schools or the department of education somehow manipulate the percentage of target achieved, which is the index that determines whether schools win or not the bonus. In practice, manipulating the index around the cutoff would be virtually impossible for the reasons outlines below.

The global achievement index is a weighted average of the product of pass rates and student scores. All exams are multiple choice, applied by independent evaluators[44], and graded by automated machines. It is thus virtually impossible for teachers to inflate scores. It is also impossible for the school to produce fake numbers on pass rates, because they are checked by Secretariat of Education by matching students' unique codes from one year to another. While teachers could in principle inflate pass rates by effectively promoting more students, it seems very unlikely that teachers could strategically do so near the cutoff, because pass rates ought to be calculated by December while test scores are only released by the middle of next year. It is not possible to revise approval rates by the time scores are revealed and it would be very hard for school administrators to precisely predict the percentage of the target that would be achieved.

The only possibility of manipulation comes from the process of listing winners and non-winners, which is done by the Secretariat of Education. We actually found 4 schools that achieved 49.5% of their global target and were rounded up, receiving the bonus. Nonetheless, both the graphical evidence and the McCrary tests show no evidence of manipulation of the index around the threshold. After excluding these 4 schools, we get a clear discontinuity on bonus allocation, as seen by figure 9.

In table 5, we also test the robustness of the 2011 estimates to bandwidth choices and functional forms. The magnitude of the impacts on math and language test scores fall by more than a half when a larger bandwidth suggested by Imbens

[44] In NYC, having other professionals then the teacher itself applying the exam to its students was enough to substantially decrease cheating (2016)

and Kalyanaraman is employed, with an estimated impact of 0.172 sd and 0.188 for language and math, respectively. The estimate using the bandwidth calculated from cross validation yields a similar estimate, also significant at 5%, while OLS estimates are more imprecise.

Finally, we test if our basic results appear when using irrelevant thresholds. Instead of using the threshold of 50% of the target, we use false thresholds of 65%, 80%, 35% and 20% of the target. Table 6 shows the estimates. We find no significant effects in any of these falsification tests.

3.6.Mechanisms

What could be driving the increase in students' test scores? In this section, we investigate some possible paths behind the observed changes in students' test scores. As we find no significant changes in test scores for the 2008/2009 bonus, we will concentrate our analysis to bonus of 2009/2010 and 2010/2011. We start by investigating whether not winning the bonus by a small margin changes teachers' pedagogical practices, principals' management practices and teachers' perceptions about other staff by making use of teachers, students and principal surveys from SAEPE.

We start by analyzing teachers' pedagogical practices. As teachers are the key input in the production of students' human capital, teacher pedagogical practices are our first candidate to explain the observed effects on students' test scores. [45] At the SAEPE survey, teachers were asked about their level of agreement or disagreement about a series of statements in a Likert scale. These statements were designed to elicit teachers' perceptions about the frequency with which they engage in some activities.

The questionnaires from 2010[46] and 2011 have different questions about teaching practices. We thus construct two different indexes of teaching practices for each year by employing factor analysis. For each year, we subtract the mean and divide by the standard deviation to facilitate interpreting the results. For 2010, our index of teaching practices encompasses teacher ratings about the amount of

[45] In Israel, for example, the introduction of a pay for performance policy induced teachers to change their pedagogy and to exert more teaching effort (Lavy, 2009).

[46] The questionnaires for 2008, 2009 and 2010 are the same.

homework assigned, homework corrected in class, coverage of school curriculum, engagement with low performing students and class management[47]. For 2011, the index includes self-reported ratings about their own degree of absenteeism, of their capacity for maintaining a climate of respect inside the classroom, the frequency of discussions about the homework, usage of the textbook and agreement about feeling responsible for their students' learning[48].

Schools that lose the bonus by a small margin have a higher index of pedagogical practices for 2010. Using the bandwidth proposed by Calonico, Cattaneo and Titiunik, we estimate an impact of 0.44 of standard deviation for the index of pedagogical practices, significant at 5%. The effect is robust to bandwidth choice and different global polynomial OLS specifications[49]. By repeating the same exercise for each individual question that composes the index, we observe significant shifts on the probability of assigning homework, on the proportion of students who complete the assigned homework, on the probability of covering all the curriculum and the probability of using newspapers and magazines in class.

For the 2010/2011, we do not find significant results on the index containing teachers self-reported practices. When analyzing question by question, we find a positive impact on teachers' agreement with the sentence "I feel responsible for my students learning". We also find a small decrease on reported absenteeism, but only significant at 10% and only for some choices of bandwidth.

The SAEPE questionnaire from 2011 also allows us to capture changes on teacher attitudes from the point of view of the students. We observe positive impact on students' perception that teachers are attentive to their opinion, on the frequency teachers solve students' doubts, on teacher fairness with all students, on the frequency teachers correct the homework, on the frequency teachers use the

[47] Teachers were asked to rate how much they agree with sentences like: "I assign homework", "I check my students' homework", "I use the textbook inn my classes", "I pay extra attention to students with poor performance or those who have learning difficulties", "I use newspapers and magazines in class", "I managed to successfully cover the curriculum this year", "I lose a lot of time organizing the classroom, with roll call, with student warnings and with disciplinary problems", among other sentences. We construct an index of pedagogical practices via factor analysis.

[48] Indeed, some of these aspects seem to be very relevant for student learning. Fernandes (2013), analyzing student scores and teacher surveys from São Paulo, finds that the effect of teacher pedagogical practices is much more relevant than the effect of teachers' content knowledge. In addition, he finds that assigning and correcting homework has a strong and significant predictive power on explaining student score gains.

[49] See on line appendix table 21

textbook, on the probability of explaining the lesson until all students learn and on being clear when explaining the subjects. By aggregating those perceptions in a single index using factor analysis, we find an impact of 0.22 of a standard deviation (significant at 10%). The estimate is also significant at 10% for other bandwidth choices.

Interestingly, the students' questionnaire from 2011 also reveals a significant improvement of students' satisfaction with the school. Students are more likely to say that they are learning new things at the school, to feel safe, to feel well taken care of, to feel valued, to feel proud of their school, of being motivated to go to school. We find an impact of 0.260 of a standard deviation (significant at 10%) for the index aggregating variables about student satisfaction.

These results open the possibility that the impact on student performance are due to changes in school wide factors that are under the control of principals. Principals are key in facilitating teachers' work, fostering teachers' professional development, establishing school routines and can influence a series of factors that are shown to influence student learning, like the overall climate of the school, student discipline and frequent teacher feedback (Loeb, Kalogrides, and Beteille, 2011; Dobbie and Fryer, 2011; Angrist, Pathak, and Walters, 2012)[50].

The questionnaires from SAEPE 2010 and 2011 allow us to form two indexes (one for each year) summarizing several variables directly linked to principals' leadership, such as being encouraged and motivated by the principal, principal absenteeism or having confidence in the principal. We construct a principal leadership index by factor analysis. Column 4 of Table 7 (line 2) shows no significant impact on principals' leadership score. The RD estimates are very small and are statistically indistinguishable from zero, regardless of bandwidth choice or functional form. Improvements in students' test scores do not seem to have been associated to changes in principal leadership.

As Pernambuco's bonus is collective, there is an incentive to free ride on the work of teachers from tested subjects and grades. Bruns and Ferraz (2012), for

[50] The full set of sentences can be seen in the online appendix. Some examples of sentences: "The principal motivates me to work"; "The principal manages to engage teachers", "The principal stimulates innovative practices", "I fell respected by the principal", "The principal implements clear rules", among others. We construct an index of principal leadership via factor analysis. A higher index means a stronger agreement with the positive sentences. All factor loadings have the expected sign and the Keyser-Meyer-Olkin statistic, 0.96, shows excellent sampling adequacy. Finally, we standardize the index to have zero average and standard deviation equal to one.

example, find the introduction of the bonus system in Pernambuco was more effective in raising student performance in small schools, possibly due to the free rider problem. Critics of pay for performance systems often argue that monetary incentives can stimulate competition across schools and between teachers within the same school, thus harming social connections. Missing the bonus by a short margin could incentivize teachers to cooperate and strengthen their relations to overcome the bad results.

We test this hypothesis by creating two indexes from teachers' answers in the SAEPE survey from 2010. First, we create a teamwork index summarizing how teachers work together and collaborate with each other by sharing ideas, suggestions, concerns, as well as how strongly they participate on school decisions.[51] The second index is based on a set of attitudinal measures of trust that were adapted from the World Value Survey to the school context and applied to teachers in Pernambuco. This trust index summarizes teachers' ratings to statements aimed to elicit their degree of trust on other staff from the school. [52]

Columns 5 and 6 of table 7 test the hypothesis that losing the school bonus could harm or favor school teamwork and disrupt trust among teachers. Non bonus winning schools have both lower levels of reported teamwork and trust, but the estimated RD impact of the bonus on teamwork and trust is very small and cannot be statistically distinguished from zero.

We test if slightly missing out on winning the bonus leads to changes in teachers' attitudes towards standardized tests. Teachers who oppose the use of standardized tests should not use test results to revise their teaching practices and to exert more teaching effort. Teachers who do win the bonus by a small margin could then start using the results from SAEPE to revise their practices[53]. Both

[51] This includes rating statements like "I participate in decisions related to school subjects", "In this school, I have difficulty sharing my concerns and frustration", "The team of teachers takes my ideas into consideration", "Few teachers are willing to take on new charges to improve the school" and "I take into account suggestions from other colleagues".

[52] Teachers rated in a Likert scale statements like: "I can trust in the school staff", "I am a person in whom people can trust", "If someone from the school had to borrow R$30 for an emergency, she could borrow", "If I had to borrow R$30, I could borrow from someone from the school" or "People from the school staff all have the same opinion about what is correct and what is wrong".

[53] In a related study, for example, Muralidharan and Sundararaman (2011) find that teachers' support of performance pay policy increased after exposure to an incentive program in India. Consequently, it could be possible that the exposure to the loss of the bonus changes the way teachers deal with standardized tests.

questionnaires from 2010 and 2011 contain questions about how teachers deal with standardized test, although with different questions. We summarize a series of teachers' ratings from SAEPE 2010 and 2011 based on questions that ask whether they use the results of standardized tests to reflect upon their work[54]. Although we find a positive and significant impact on the rate of agreement with a sentence indicating that the results from the tests have been contributing to improve students' performance, the RD impact of losing the bonus on the index summarizing these attitudes cannot be statistically differentiated from zero.

Finally, we test if the results could be explained by teacher turnover. In a field experiment in schools, Fershtman and Gneezy (2011) find that strong incentives might induce teachers to exert more effort and exhibit a higher performance, but also induce to more quitting. Table 8 shows the RD impact of losing the bonus on teacher turnover and teacher characteristics on the next year. Although we find an impact of 2.7 percentage points on teacher turnover, significant at 10% when using Inbens and Kalyanaraman's bandwidth, teacher characteristics do not change in the next year. Indeed, in the context studied here, although quitting is a real possibility, it could hardly explain the impacts on grades, because retirement decisions and teacher transfers between schools happen at the beginning of the school year, in February, while the bonus results are only released by May and June.

3.7.Discussion

3.7.1.Loss aversion and reference points

We find sizeable, significant and robust impacts of not winning the bonus on test scores and teachers pedagogical practices 6 months after teachers and principals are informed about the school results. Schools who fail to win the bonus do not suffer any sanctions and winning schools receive no additional rewards or intervention from the Secretariat of Education. The performance of students along both sides of the bonus allocation threshold are indistinguishable from one another. Given the information on the percentage of the target achieved by school, winning

[54] Following the previous procedure, we also transform teachers answers from a Likert scale into a single index via factor analysis. Teachers' attitudes are assessed by evaluating sentences like: "I use the results from standardized tests to revise my pedagogical practices"; "Discussing the results from standardized tests help me reflect upon my own work" and "The results from external evaluations have been contributing to improve students' performance".

or not the bonus generates no additional information about the probability of winning the bonus the following year and should not cause different reactions from teachers at different sides of the discontinuity.

We argue that those effects can be explained through loss aversion, a central piece of Prospect Theory. According to Khanemann and Tversky (1979), individuals make decisions by framing outcomes as gains or losses relative to a reference point. The utility curve is steeper at the loss side, meaning that losses entail a bigger change in utility than a corresponding gain. Negative marginal utility at the gain side and diminishing sensitivity to losses mean that individuals are risk averse at the gain side and risk takers at the loss side.

The centerpiece of our argument is that losing the bonus in a given year by a small margin increases the salience of the bonus allocation threshold, and that teachers from schools that barely missed the bonus are suddenly placed in the loss side relative to the threshold. The argument does not depend on naïve teachers who believe that the percentage of the target achieved is a pure function of effort.

One possibility is that the threshold becomes salient after teachers are placed near it. Imagine a teacher who expected her school to achieve 100% of the target, for example, but received the message that her school only made 45% of the target. Instead of taking the full target as her reference point, she now evaluates losses and gains relative to 50% of the target. This reasoning is consistent with a model in which the reference point is endogenously chosen based on the expectations the person had in the recent past as in Kószegi and Rabin (2006). By believing that there is some chance of losing the bonus again and by willing to avoid this outcome, teachers change their pedagogical practices, increasing the frequency in which they engage in actions that demand effort, such as assigning and grading homework or using the textbook in class.

It is important to emphasize, though, that different from other studied cases of loss aversion arising from laboratory experiments or from basketball, teachers are not competing in a tournament and are not really behind any scoring. Every year, teachers have a new chance of winning the bonus, and the size of the improvements on the Idepe index that the school has to attain in the next year does not depend neither on winning or not the bonus, neither on the percentage of the

target achieved. [55] In this sense, our setting bares greater similarity to the one studied by Pope and Schweitzer (2011), in which professional golf players change their behavior depending on whether they are ahead or behind a historical reference point on the scoring for each hole[56], but disregarding their relative position to opponents.

It is interesting to notice that the formula of the global Idepe index creates some thresholds that could be taken into account when strategically deciding how to allocate effort. Because target levels for 2010 and 2011 were set based on the 2 previous test scores, schools with relative large gains in the previous year can have a level of target that is lower than their current Idepe index for a given grade. The minimum improvement, relative to their current target, can thus be actually negative (although positive when relative to the average of the 2 previous years). Let us take for example a school that has only grades 5 through 9 and whose Idepe index for 9th grade was 2.24 in 2009 and 6.76 in 2010. Their average Idepe is 4.5 and their target is to improve by 0.17 over 4.5. Teachers from this school receive the full amount of the bonus if their Idepe is greater than 4.67, and receive nothing if their Idepe is lower than 4.58. As their last Idepe was 6.76, teachers from this school can safely reduce effort such that its Idepe falls by 2 points and the school would still make its target.

Coupled with the fact that negative deltas between Idepe indexes from one year to another are disregarded when computing the global Idepe index (and that low Idepe indexes today mean lower targets in the future), we could actually expect schools with no need to improve to make less effort than those who are required to improve in order to achieve the full target. This will be especially true in the presence of loss aversion. If the reference point teachers look at is the point where they need to make no additional effort in order to make the full target (or half of the target), then we would expect teachers at the loss side (those who still have to make effort to improve) to exert more effort than those at the gain side (those who can stay still).

[55] Targets are set based on the average of the 2 previous Idepe indexes for each test subject and grade. The percentage of target achieved depends only on the current global target and the global Idepe index in the current year. Schools whose

[56] In a golf tournament, players

Figures 11 to 16 show students test scores on math and language for year 2011 against the targets of improvements for each grade.[57] On the horizontal axis we plot the amount of improvement (from last Idepe) needed in order to achieved the target. In figures 17 to 22 we replace the horizontal axis by the improvement needed in order to make half of the target. There seems to be no discontinuity in test scores around these thresholds, and table 9 confirms the pattern observed in the figures. We test for the presence of a discontinuity in test scores around the point in which schools need to make no improvement to meet its (full or half) target. We find no evidence that schools take these reference points into account. Instead, schools that barely lost the bonus take only the winning/losing threshold for global Idepe as reference point.

3.7.2.Confusion of bonus rules?

Traditional economic theory would predict no reactions from teachers who fully understand the bonus scheme. It could be possible, though, that teachers react purely because they do not understand the rules of the bonus and believe there could be a possible sanction, or stigma, from not winning the bonus. In a related context, Ahn and Vigdor (2014) present evidence that teachers from schools that failed to qualify for the bonus in North Carolina respond substantially by raising students test scores, but possibly because teachers might have confounded the bonus rules with the No Child Left Behind incentives, which include severe sanctions for schools that fail to make yearly progress.

The SAEPE questionnaire from 2011 allows us to assess the possible role of confusion in explaining our results. Teachers were asked if they fully understand the bonus rules or not. Seventeen percent of teachers reported not knowing about the rules. In Column 1 of table 10, we split the sample of schools by the median proportion of teachers who report knowing the bonus rules. Contrary to the confusion hypothesis, we find the effect to be significant only for schools where the majority of teachers' report knowing about the program rules.

[57] Dynamically, schools that perform worse in a given year have lower targets in the next years, which creates an additional to invest less effort in grades where there is no incentive for improvement.

We also test for confusion by splitting the sample between principals with high and low experience. Experienced teachers should be better informed about the program rules. Ahn and Vigdor (2014) only find a significant effect of almost winning the bonus in North Carolina for schools headed by mid experienced teachers. They argue that inexperienced teachers would know less about how to raise students test scores, while very experienced teachers would be better able to distinguish between luck and performance and would rationally attribute the result to chance. In columns 5 to 7 of table 10 we split our sample by terciles of principal experience. In fact, we find no significant results for schools headed by principals with less than 5 years of experience, but we find similar and significant impacts for schools whose principals have between 6 and 10 years of experience, and for schools whose principal has been managing schools for more than 11 years. Together, these results strongly reject the view that such behavioral biases would arise only for agents who are either uninformed or those who have not enough experience to adequately weight the role of chance.

3.7.3.Long run: Effect of almost winning the bonus for a second time

Whether behavioral biases persist with large stakes and frequent exposure to the phenomenon is a crucial question in the literature of behavioral economics. While Pope and Schweitzer (2014) find that even professional and experienced golf players such as Tiger Woods present loss aversion, John List (2003) finds that the experience with trading, as measured by the number of transactions made, substantially decreases the endowment effect. As List (2003) shows in his sport cards trading market study, what really matters for demonstrating behavioral biases is not the amount of time on activity, but the recent history of exposure to situations in which it is possible for biases to arise. In this sense, teachers first exposed to the loss of bonus should react differently from teachers who have already been in a similar situation. Teachers who have already lost the bonus by a very small margin might already have been exhorted to exert more effort and might be better able to distinguish chance from bad performance. It is also possible that teachers from those schools have already changed their teaching practices and improved students' scores, such that there would be less room for improvement.

To test for these hypotheses, we split our sample from 2011 between schools that had already lost the bonus by a small margin and schools that have never lost by such margin in the 2008/2009 and the 2009/2010 bonuses. We start by defining a short window of schools that have ever been less than 20 percentage points from winning the bonus (i.e. schools that have ever achieved between 30% and 49% of their target). We then enlarge this window to 30 and 40 percentage points. Table 11 shows the estimates. We find significant RD impacts of almost winning the bonus on test scores, in 2011, for schools that just lost the bonus for the first time in 2010/2011. The impact on test scores, in 2011, of losing the 2010/2011 bonus cannot be distinguished from zero for schools that had previously lost by a small margin in 2008/2009 and in 2009/2010. These results suggest that the reactions to losing the bonus might decrease or disappear in the long run when most schools will have missed the bonus by a small margin.

3.8.Conclusion

While previous papers have found that being just behind a reference point can motivate effort, there is little evidence on whether this result can be found outside the lab or sports. This paper provides evidence that not winning a teachers' pay for performance bonus might actually lead to improvements in student learning. By exploiting a discontinuous bonus allocation rule, we find that teachers exert more effort and change their pedagogical practices after knowing that they did not win the bonus, but were very close to winning. These results are perfectly compatible with loss aversion, a central piece of Khaneman and Tversky's (1979) Prospect Theory.

In addition to documenting these basic results, we also show that the same phenomenon is not observed at other thresholds that are relevant for the decision of where to allocate effort, reinforcing the idea that losing the bonus increases the salience of the bonus allocation threshold. We find no evidence that our results are driven by confusion of bonus rules. Finally, we find that schools that had already previously lost the bonus in the past by a short margin do not react to losing the bonus.

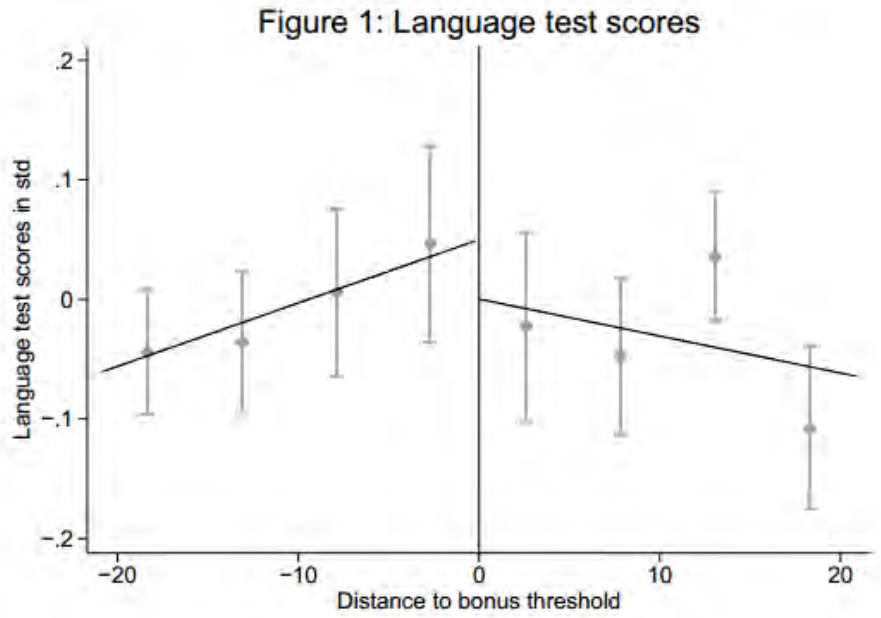
Taken together, these findings corroborate the idea that teachers demonstrate loss aversion when incentivized by bonuses, which has important implications for

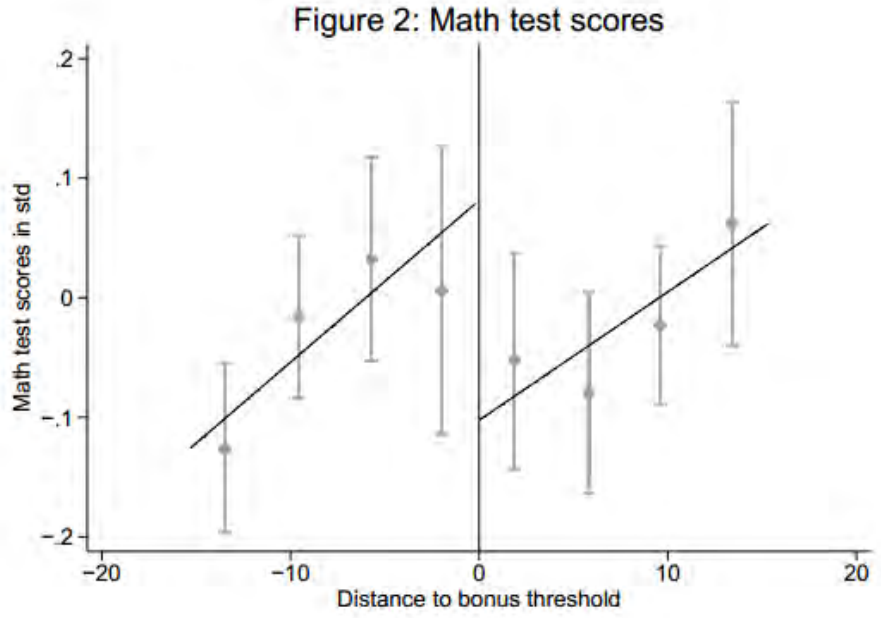
designing bonus schemes. Discontinuous bonus rules in which some schools do not receive a bonus should be superior than continuous rules in which all schools receive some bonus. Grouping schools of similar performance in tournaments in which only some schools receive the bonus should also lead to higher teacher effort and student performance.

3.9.Figures

Figure 1- Language test scores

Figure 2-Math test scores



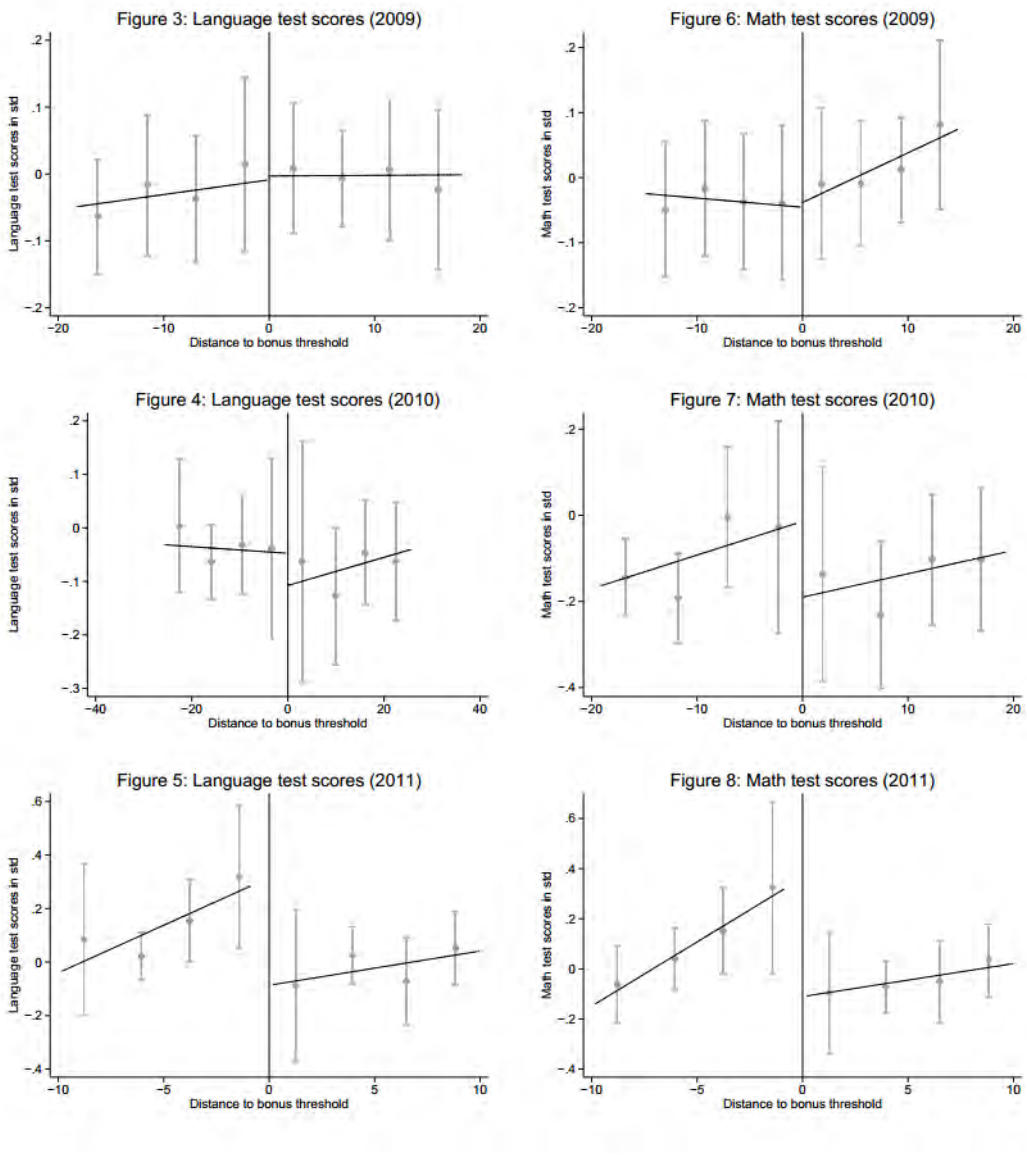


Notes: Figures 1 and 2 present linear estimates of student test scores against the distance to the bonus threshold. Schools at the right side of the threshold receive the bonus.

Figure 3-Language test scores, 2009

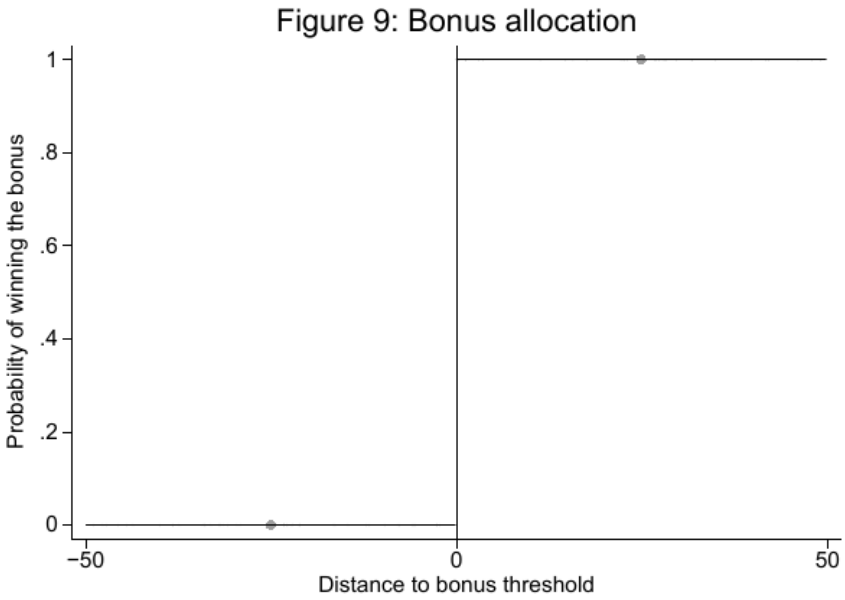
Figure 4-Math test scores, 2009

Figure 5- Language test scores, 2010



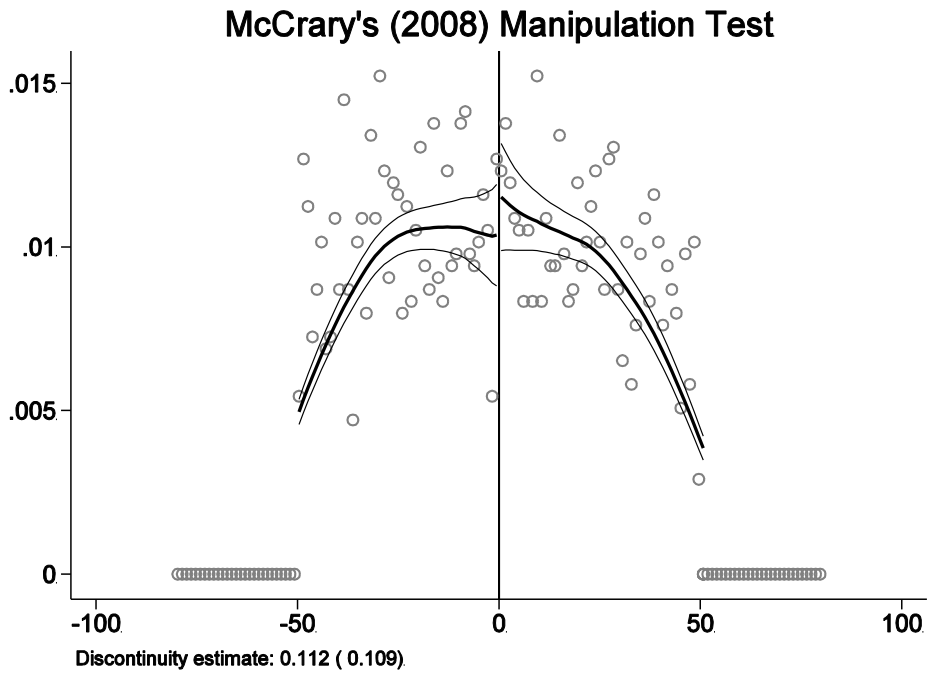
Notes: Figures 1 and 2 present linear estimates of student test scores against the distance to the bonus threshold. Schools at the right side of the threshold receive the bonus.

Figure 9- Bonus allocation

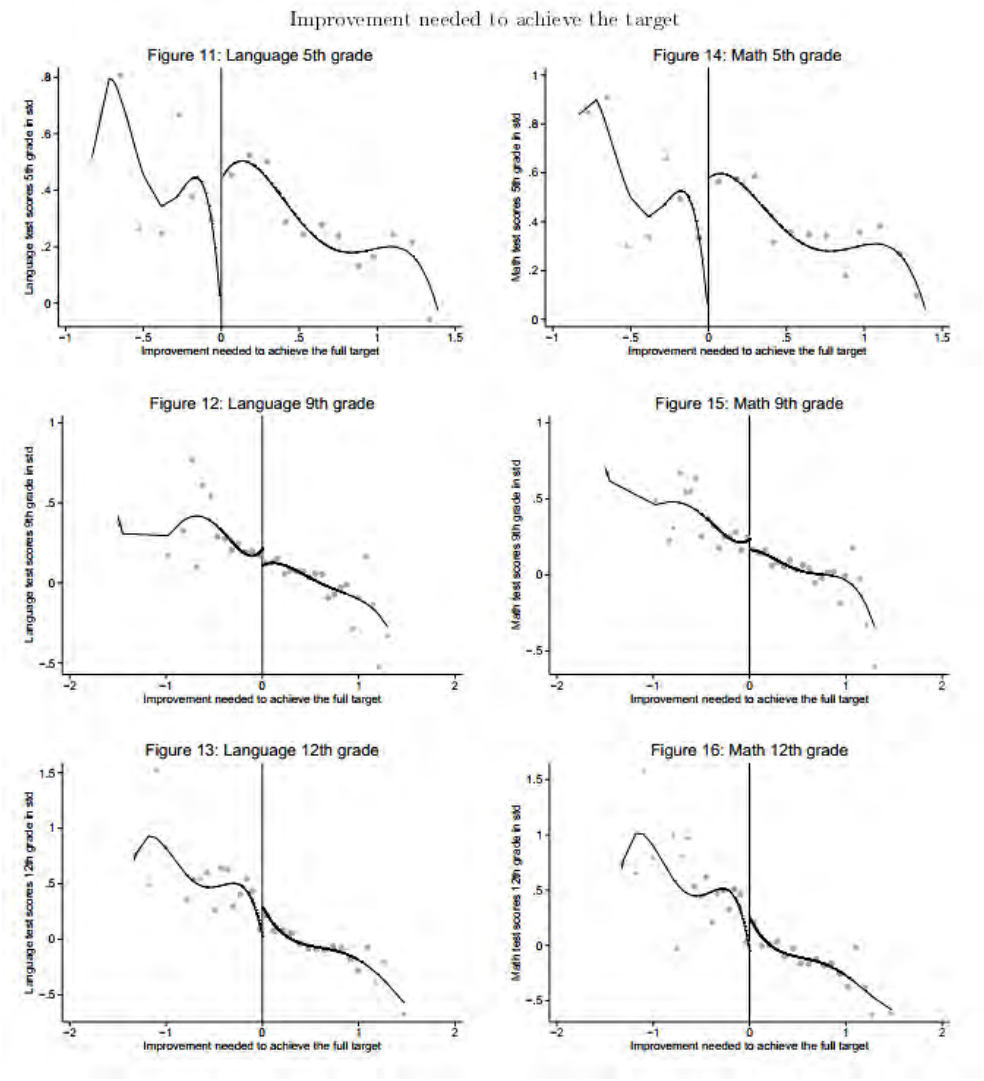


Notes: Figures 9 present of the probability of receiving the bonus against the distance to the bonus threshold. Schools at the right side of the threshold receive the bonus.

Figure 10- McCrary test



Notes: Figure shows the density of schools on the margin of bonus victory, i.e. the percentage of target achieved minus 50%. Discontinuity estimate: 0.112 (0.109)



Notes: Figures 11 to 17 present local linear estimates of student test scores against the distance to the threshold indicating that the school needs to make no additional effort.

Figure 17- Improvement needed to achieve half of target – Language, 5th grade

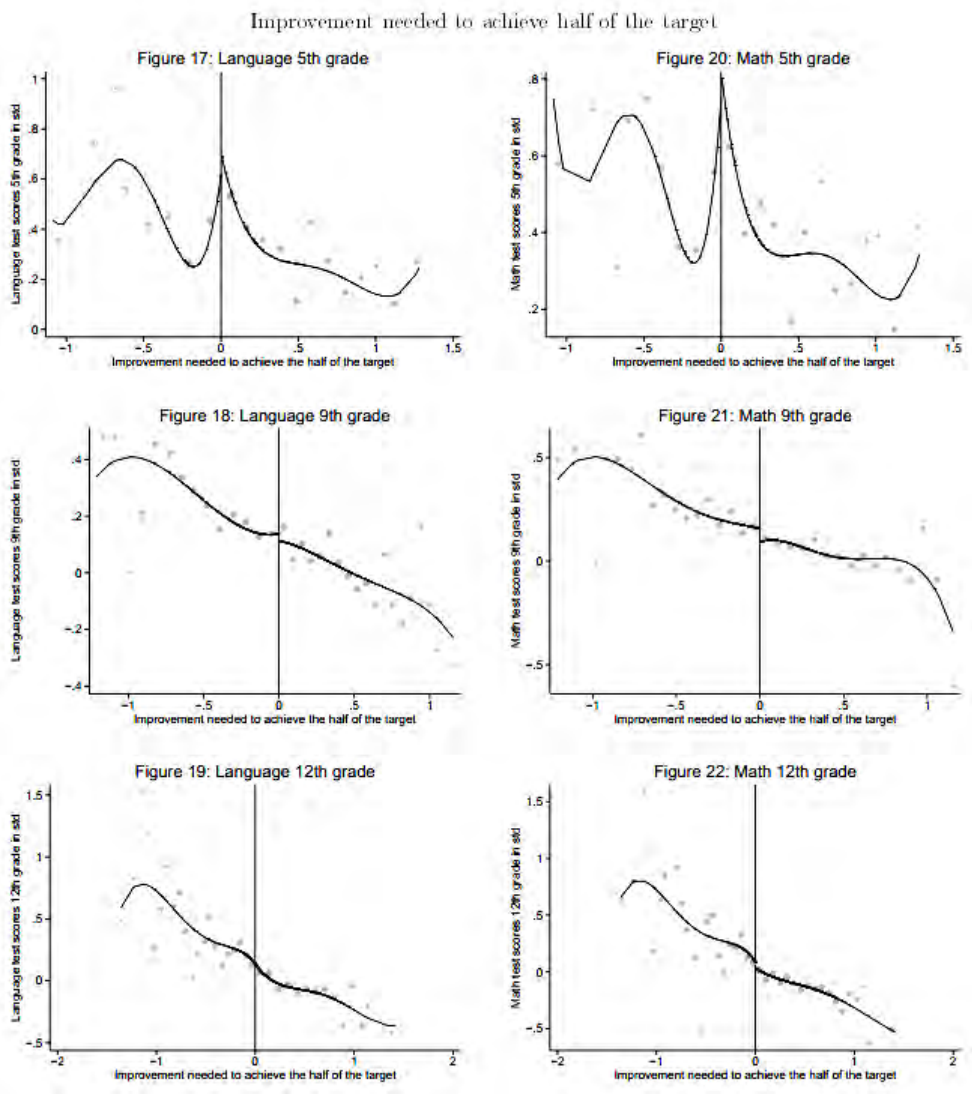
Figure 18- Improvement needed to achieve half of target – Math, 5th grade

Figure 19- Improvement needed to achieve half of target – Language, 9th grade

Figure 20- Improvement needed to achieve half of target – Math, 9th grade

Figure 21- Improvement needed to achieve half of target – Language, 12th grade

Figure 22- Improvement needed to achieve half of target – Math, 9th grade



Notes: Figures 17 to 22 present local linear estimates of student test scores against the distance to the threshold indicating that the school needs to make no additional effort to receive the bonus.

3.10.Tables

Table 1- Bonus value

VALUE OF THE BONUS

Year	Redistribution Factor	Percentage of winners	Total amount Allocated	Paid on
2008/2009	2.31	51%	R$ 28,800,000.00	jul/09
2009/2010	1.78	81%	R$ 41,000,000.00	mai/10
2010/2011	1.88	70%	R$ 44,887,053.00	jul/11
2011/2012	2.30	56%	R$ 49,377,745.00	ago/12

Notes: Source: Secretariat of Education of Pernambuco (SEE-PE) and Diário Oficial do Estado de Pernambuco

Table 2

RD IMPACT OF MISSING THE BONUS ON TEST SCORES

	All grades	5th grade	9th grade	12th grade
PANEL A: Math Test Scores (std)				
Estimate at optimal bandwidth	0.134*	0.137	0.133**	0.096*
	(0.069)	(0.161)	(0.062)	(0.058)
Observations left of discontinuity:	220	98	206	210
Observations right of discontinuity:	210	85	199	204
PANEL B: Language Test Scores (std)				
Estimate at optimal bandwidth	0.119*	0.134	0.146**	0.055
	(0.064)	(0.145)	(0.064)	(0.057)
Observations left of discontinuity:	232	102	213	256
Observations right of discontinuity:	216	91	209	237

Notes: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school in a given year, from 2009 to 2011. Data source for all outcomes is the Secretariat of Education of Pernambuco. Every cell presents the estimates of a local linear regression using the bandwidth proposed by Calonico, Cattaneo and Titiunik (2014). We control for year dummies and for a dummy indicating whether the school is a 'Escola de referência' (full time schools with increased autonomy).

Table 3

RD IMPACT OF MISSING THE BONUS ON TEST SCORES

	Year 2009 (1)	Year 2010 (2)	Year 2011 (3)
PANEL A: Math Test Scores (std)			
Estimate at optimal bandwidth	-0.042	0.149	0.480***
	(0.070)	(0.145)	(0.170)
Observations	283	151	99
PANEL B: Language Test Scores (std)			
Estimate at optimal bandwidth	-0.027	0.063	0.429**
	(0.084)	(0.135)	(0.170)
Observations	193	157	108

Notes: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school in a given year, from 2009 to 2011. Data source for all outcomes is the Secretariat of Education of Pernambuco. Every cell presents the estimates of a local linear regression using the bandwidth proposed by Calonico, Cattaneo and Titiunik (2014). We control for year dummies and for a dummy indicating whether the school is a 'Escola de referência' (full time schools with increased autonomy).

Table 4

VALIDITY CHECKS: DISCONTINUITIES ON COVARIATES AND PREVIOUS TEST SCORES

		RD: Calonico, Cattaneo and Titiurnick (1)	RD: Imbens and Kalyanaraman (2)	RD: Cross validation (3)	OLS: Polynomial 3rd order (4)	OLS: Polynomial 2nd order (5)	OLS: Polynomial 1st order (6)
Students' age	Estimate:	-0.303	-0.343	-0.306	-0.419	-0.412	0.191
	Std error:	(0.400)	(0.283)	(0.263)	(0.284)	(0.278)	(0.188)
	N:	536	1,073	1,228	2,742	2,742	2,742
Proportion of black students	Estimate:	0.012	-0.003	-0.003	-0.028	-0.026	-0.017
	Std error:	(0.040)	(0.018)	(0.019)	(0.020)	(0.019)	(0.012)
	N:	342	1,537	1,228	2,742	2,742	2,742
Number of students	Estimate:	-109.996	-107.116	-75.942	-174.915***	-154.386***	49.800
	Std error:	(88.636)	(68.546)	(59.382)	(56.555)	(55.335)	(36.239)
	N:	584	893	1,115	2,742	2,742	2,742
Teachers years of schooling	Estimate:	-0.040	-0.011	-0.041	-0.043	-0.032	0.025
	Std error:	(0.121)	(0.060)	(0.073)	(0.071)	(0.069)	(0.046)
	N:	439	1,852	1,228	2,742	2,742	2,742
Math test score (t-1)	Estimate:	0.033	-0.045	-0.032	-0.038	-0.009	-0.015
	Std error:	(0.064)	(0.032)	(0.038)	(0.037)	(0.036)	(0.025)
	N:	270	967	680	1,798	1,798	1,798
Language test score (t-1)	Estimate:	0.036	-0.024	-0.024	-0.052	-0.037	-0.021
	Std error:	(0.062)	(0.031)	(0.037)	(0.035)	(0.035)	(0.025)
	N:	280	971	680	1,798	1,798	1,798

Note: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school at a given year. Student and teacher characteristics from INEP's school census. Test scores from SEEDUC-PE. All regressions include year dummies and a dummy for full time schools.

Table 5

SENSITIVITY TO BANDWIDTH CHOICES AND FUNCTIONAL FORM

		RD: Calonico, Cattaneo and Titiurnick (1)	RD: Imbens and Kalyanaraman (2)	RD: Cross validation (3)	OLS: Polynomial 3rd order (4)	OLS: Polynomial 2nd order (5)	OLS: Polynomial 1st order (6)
Language test scores (2011)	Estimate:	0.478**	0.172**	0.147**	0.139*	0.129	0.027
	Std error:	(0.186)	(0.079)	(0.072)	(0.083)	(0.082)	(0.050)
	N:	97	307	354	905	905	905
Math test scores (2011)	Estimate:	0.435**	0.188**	0.174**	0.151*	0.149*	0.044
	Std error:	(0.184)	(0.080)	(0.075)	(0.078)	(0.077)	(0.047)
	N:	103	302	341	905	905	905

Notes: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school in a given year, from 2009 to 2011. Data source for all outcomes is the Secretariat of Education of Pernambuco. Column 1 presents the estimates of a local linear regression using the bandwidth proposed by Calonico, Cattaneo and Titiunik (2014), while column 2 employs the bandwidths proposed by Imbens and Kalyanaraman (2011). Column 3 uses cross validation for choosing the optimal bandwidth. Columns 4, 5 and 6 present estimates from an OLS regression that includes cubic, quadratic and linear terms interacted with the winning/losing dummy. We control for year dummies and for a dummy indicating whether the school is a 'Escola de referência' (full time schools with increased autonomy).

Table 6

FALSIFICATION TESTS USING IRRELEVANT THRESHOLDS

	Threshold as percentage of target achieved:	65%	80%	35%	20%
Language test scores (2011)	Estimate:	-0.118	0.083	-0.082	0.074
	Std error:	(0.086)	(0.088)	(0.094)	(0.098)
	N:	185	192	162	131
Math test scores (2011)	Estimate:	-0.143	0.022	-0.134	0.084
	Std error:	(0.100)	(0.095)	(0.106)	(0.121)
	N:	165	216	166	131

Notes: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school in a given year, in 2011. Data source for all outcomes is the Secretariat of Education of Pernambuco. All cells present the estimates of a local linear regression using the bandwidth proposed by Calonico, Cattaneo and Titiunik (2014). IN column 1, instead of using the threshold of 50% of the target, we use a threshold at 65%, while in column 2 we use a threshold of 80%, 35% in column 3 and 20% in column4. We control for year dummies and for a dummy indicating whether the school is a 'Escola de referência' (full time schools with increased autonomy).

Table 7

MECHANISMS: IMPACT OF NOT WINNING THE BONUS ON TEACHER AND STUDENT PERCEPTIONS

	Pedagogical Practices (reported by teachers) (1)	Pedagogical practices (reported by students) (2)	Student satisfaction (reported by students) (3)	Principal behavior (4)	Teamwork (5)	Trust (6)	Use of standardized tests (7)
Year: 2009	0.224	-	-	0.065	0.110	0.160	0.059
	(0.201)			(0.172)	(0.232)	(0.142)	(0.204)
	170			215	199	269	172
Year: 2010	0.474**	-	-	-0.092	0.109	-0.216	0.320
	(0.241)			(0.181)	(0.215)	(0.240)	(0.253)
	154			155	124	128	91
Year: 2011	0.070	0.221*	0.265*	-0.174	-	-	-0.144
	(0.234)	(0.119)	(0.144)	(0.314)			(0.214)
	152	163	150	125			152

Notes: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school in a given year (2009-2011). In particular, we measure the effect of the revelation of bonus winners in 2009, based on exams taken at the end of 2008, at school contextual factors in 2009. Data source for all outcomes are the SAEPE teacher and student surveys applied to all teachers teaching language and math at 3rd, 5th, 9th and 12th grades. All estimates from a local linear regression using the bandwidth proposed by Calonico, Cattaneo and Titiunik (2014). We control for year dummies and for a dummy indicating whether the school is a 'Escola de referência' (full time schools with increased autonomy).

Table 8

TEACHER TURNOVER

	CCT (1)	IK (2)	Control Average (3)
Teacher turnover	0.039	0.027*	0.336
(proportion of teachers who	(0.037)	(0.016)	(0.006)
leave the school at t+1)	381	1,702	901
Prop of teachers with college degree (t+1)	-0.018	-0.006	0.882
	(0.024)	(0.013)	(0.006)
	430	1,718	901
Number of schools teacher works (t+1)	-0.023	0.005	1.509
	(0.037)	(0.020)	(0.007)
	460	1,713	901
Sex: Male=1 (t+1)	-0.005	-0.002	0.275
	(0.026)	(0.014)	(0.006)
	480	1,630	901
Race:Black or Mixed=1 (t+1)	0.007	-0.002	0.345
	(0.032)	(0.015)	(0.006)
	383	1,721	901
Teachers' average years of schooling (t+1)	-0.088	-0.029	15.399
	(0.121)	(0.062)	(0.028)
	435	1,729	901

Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school in a given year, from 2009 to 2011. Data source for all outcomes is the Secretariat of Education of Pernambuco. Column 1 presents the estimates of a local linear regression using the bandwidth proposed by Calonico, Cattaneo and Titiunik (2014), while column 2 employs the bandwidths proposed by Imbens and Kalyanaraman (2011). Column 3 uses cross validation for choosing the optimal bandwidth. Columns 4, 5 and 6 present estimates from an OLS regression that includes cubic, quadratic and linear terms interacted with the winning/losing dummy. We control for year dummies and for a dummy indicating whether the school is a 'Escola de referência' (full time schools with increased autonomy).

Table 9

SALIENCE OF OTHER THRESHOLDS/
DISCONTINUITIES AT ALTERNATIVE THRESHOLDS

	Language			Math		
	5th grade	9th grade	12th grade	5th grade	9th grade	12th grade
	(1)	(2)	(3)	(4)	(5)	(4)
Panel A: Point of no required improvement to achieve the full target						
Estimate	0.25	-0.051	0.263	0.347	-0.026	0.325*
Std error:	(0.224)	(0.067)	(0.16)	(0.243)	(0.07)	(0.185)
N	108	392	187	107	438	178
Panel B: Point of no required improvement to achieve half of the target						
Estimate	0.121	0.004	-0.058	0.107	-0.041	-0.101
Std error:	(0.214)	(0.059)	(0.107)	(0.216)	(0.055)	(0.118)
N	139	443	364	139	511	389

Notes: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school in a given year, at 2011. Panel A presents estimates of the RD impact of being required to making additional effort in order to achieve the target. Panel B presents estimates of the RD effect of being required to make additional effort in order to achieve half of the target. Data source for all outcomes is the Secretariat of Education of Pernambuco. Every cell presents the estimates of a local linear regression using the bandwidth proposed by Calonico, Cattaneo and Titiunik (2014). We control for year dummies and for a dummy indicating whether the school is a 'Escola de referência' (full time schools with increased autonomy).

Table 10

CONFUSION AND EXPERIENCE: IMPACT OF NOT WINNING THE BONUS

	Teachers know bonus rules		Teacher experience higher than 6 years		Principal experience		
	< 50%	>=50%	<60%	>=60%	<=5 years	>=6 and <=10 years	>=11 years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Language test scores							
Estimate	0.021	0.249***	0.057	0.233**	-0.022	0.373***	0.373**
Std error:	(0.124)	(0.091)	(0.083)	(0.112)	(0.068)	(0.144)	(0.165)
N	91	236	214	166	204	86	89
Panel B: Math test scores							
Estimate	0.03	0.218**	0.025	0.206**	-0.04	0.300**	0.372***
Std error:	(0.124)	(0.091)	(0.090)	(0.082)	(0.074)	(0.129)	(0.144)
N	87	241	215	187	240	90	98

Notes: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school in a given year, at 2011. Panel A presents estimates of the RD impact of missing the bonus on language test scores, while panel B presents the RD impact on math test scores. Columns 1 and 2 split the sample between schools where less than half of teachers report knowing the bonus rules and schools where the majority of teachers declare knowing about the bonus. Columns 3 and 4 split the sample between schools in which more than 60% of teachers have more than 6 years of experience. Columns 5, 6 and 7 split the sample according to experience of the principal. Data source for all outcomes is the Secretariat of Education of Pernambuco. Every cell presents the estimates of a local linear regression using the bandwidth proposed by Calonico, Cattaneo and Titiunik (2014). We control for year dummies and for a dummy indicating whether the school is a 'Escola de referência' (full time schools with increased autonomy).

Table 11

IMPACT ON TEST SCORES- BY PAST EXPERIENCE IN ALMOST WINING THE BONUS

	Window: 30% to 49%		Window: 20% to 49%		Window: 10% to 49%	
	Never been	Ever been	Never been	Ever been	Never been	Ever been
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Language test scores						
Estimate	0.241***	-0.019	0.267**	-0.01	0.300***	0.005
Std error:	(0.092)	(0.097)	(0.104)	(0.078)	(0.106)	(0.070)
N	235	84	204	145	186	175
Panel B: Math test scores						
Estimate	0.216**	0.039	0.232**	-0.038	0.299***	-0.06
Std error:	(0.092)	(0.095)	(0.099)	(0.082)	(0.100)	(0.070)
N	235	84	216	154	191	187

Notes: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school in a given year, at 2011. Panel A presents estimates of the RD impact of missing the bonus on language test scores, while panel B presents the RD impact on math test scores. We split the sample based on past experience on almost winning the bonus. In columns 1 and 2, we define that schools whose percentage of target achieved ever lied between 30% and 50% is an experienced school. In columns 3 and 4 we change the definition including schools that ever achieved between 20% and 30% of the global target. Finally, columns 5 and 6 widens the window to include all schools that have achieved 10% to 20%. Data source for all outcomes is the Secretariat of Education of Pernambuco. Every cell presents the estimates of a local linear regression using the bandwidth proposed by Calonico, Cattaneo and Titiunik (2014). We control for year dummies and for a dummy indicating whether the school is a 'Escola de referência' (full time schools with increased autonomy).

4
References

AHN, Thomas; VIGDOR, Jacob L. **When Incentives Matter Too Much: Explaining Significant Responses to Irrelevant Information**, NBER Working Papers 20321, 2014.

ANGRIST, J., BETTINGER, E., BLOOM, E., KING, E., KREMER, M.. **Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment**. American Economic Review 92 (5), 15351558, 2002.

ANGRIST, J., BETTINGER, E., KREMER, M. **Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia.** The American Economic Review 96 (3), pp. 847/862, 2006.

ANGRIST, J., LANG, D., OREOPOULOS, P.. **Incentives and services for college achievement: Evidence from a randomized trial.** American Economic Journal: Applied Economics 1 (1), 136/63, 2009.

ANGRIST, J., LAVY, V.. **The effects of high stakes high school achievement awards: Evidence from a randomized trial.** American Economic Review 99 (4), 1384/1414, 2009.

ATTANASIO, O., DI MARO, V. and VERA-HERNANDEZ, M.. **Community Nurseries and the Nutritional Status of Poor Children. Evidence from Colombia** The Economic Journal. Vol 123, Issue 571. pp 1025-1058, 2013.

BAIRD, S., MCINTOSH, C., ÖZLER, B.. **Cash or condition? evidence from a cash transfer experiment**. The Quarterly Journal of Economics 126 (4), 1709/1753, 2011.

BANDIERA, O., BARANKAY, I. and RASUL, I.. **Team incentives: Evidence from a firm level experiment.** Journal of the European Economic Association, 11: 1079–1114, 2013.

BANERJEE, Abhijit V.; GLENNERSTER, Rachel; DUFLO, Esther.. **Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System.** Journal of the European Economic Association 6 (2/3): 487-500, 2008.

BARNETT W.S.. **Effectiveness of Early Education Intervention**. Science vol 333 no. 6045 pp.975-978, 2011.

BARRERA-OSORIO, F., BERTRAND, M., LINDEN, L. L., PEREZ-CALLE, F., **Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia.** American Economic Journal: Applied Economics 3 (2), 167/95, 2011.

BARRERA-OSORIO, F., FILMER, D., **Incentivizing schooling for learning: evidence on the impact of alternative targeting approaches.** Policy Research Working Paper WPS 6541; Impact Evaluation Series 98.37/52, 2013.

BARRERA-OSORIO, F., LINDEN, L. L., SAAVEDRA, J. E.. **Medium term educational consequences of alternative conditional cash transfer designs: Experimental evidence from Colombia**, mimeo, 2015.

BECKER, Gary. **Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education.** New York: Columbia University Press, 1964

BECKER, Gary, TOMES, Nigel. **An Equilibrium Theory of The Distribution of Income and Intergenerational Mobility.** Journal of Political Economy, 87(6):1153-1189, 1979.

BEHRMAN, J., CHENG, Y. and TODD, P.. **Evaluating Pre-school Programs when Length of Exposure to the Program Varies: A Nonparametric Approach**. Review of Economics and Statistic 86(1): 108-32, 2004.

BENHASSINE, N., DEVOTO, F., DUFLO, E., DUPAS, P., POULIQUEN, V.. **Turning a shove into a nudge? a "labeled cash transfer" for education.** American Economic Journal: Economic Policy 7 (3), 86/125, 2015.

BEN-PORATH, Yoram. **The Production of Human Capital and the Life Cycle Earnings** Journal of Political Economy, 75(4):352-365, 1967.

BERGER, J.; POPE, D. **Can Losing Lead to Winning?** Management Science, Vol. 57, No. 5, May 2011, pp. 817–827, 2011.

BERLINSKI, S., GALIANI, S. & MANACORDA, M. **Giving children a better start: Preschool attendance and school-age profiles**. Journal of Public Economics, 92, Issues 5-6, pp 1416-1440, 2008.

BERLINSKI, S., GALIANI, S., & GERTLER, P. **The effect of pre-primary education on primary school performance.** Journal of Public Economics, 93, 219–234, 2009.

BERNAL, R. and FERNANDEZ, C. **Subsidized childcare and child development in Colombia: effects of Hogares Comunitarios de Bienestar as a function of timing and length of exposure.** Social Science and Medicine, Vol. 97 pp 241-249, 2013.

BETTINGER, E. P., August **Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores.** The Review of Economics and Statistics 94 (3), 686/698, 2012.

BETTINGER, E., KREMER, M., SAAVEDRA, J. E. **Are educational vouchers only redistributive?** The Economic Journal 120 (546), F204/F228, 2010.

BLACK, S. E., DEVEREAUX, P. J., SALVANES, K. **Fast Times at Ridgemont High? The Effect of Compulsory Schooling Laws on Teenage Births.** NBER Working Papers10911, National Bureau of Economic Research, Inc, 2004.

BOTELHO, F., MADEIRA, R. A., RANGEL, M. A. **Racial discrimination in grading: Evidence from Brazil.** American Economic Journal: Applied Economics 7 (4), 37-52, 2015.

BOUGUEN, A, FILMER, D., MACOURS, K,. NAUDEAU, S. **Impact Evaluation of Three Types of Early Childhood Development Interventions in Cambodia.** World Bank, Washington, DC, 2013.

BRITTO, Pia Rabello; ENGLE, Patrice; ALDERMAN, Harold. **Early Intervention and Caregiving: Evidence from the Uganda Nutrition and Early Childhood Development Program**. Mimeo, 2007.

BRUNS, B. E., LUQUE, D., JAVIER. **Achieving World-Class Education in Brazil: The Next Agenda.** World Bank, Washington, D.C, 2013.

BRUNS, B., MARTINEZ, S., NAUDEAU, S. and PEREIRA, V. **Impact Evaluation of Save the Children Early Childhood Development Program in Mozambique: Baseline Results**. The World Bank. Washington, D.C, 2010.

CALONICO, S.; CATTANEO, M. D.; TITIUNIK, R.. **Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs**, Econometrica 82(6), 2295—2326, 2014.

CALONICO, S.; CATTANEO, M. D.; TITIUNIK, R.. **Optimal Data-Driven Regression Discontinuity Plots**, Journal of the American Statistical Association 110(512), 1753-1769, 2015.

CAMERON, Judy. **Evidence For An Early Sensitive Period For The Development Of Brain Systems Underlying Social Affiliative Behavior**, mimeo, 2004.

CARNEIRO, P., HECKMAN, J. **Human capital policy**. In: HECKMAN, J., KRUEGER, A. (Eds.), **Inequality in America: What Role for Human Capital Policies**. MIT Press, Boston, MA, 2003.

CARVALHO, Leandro S.; SOARES, Rodrigo R.. **Living on the edge: Youth entry, career and exit in drug-selling gangs** Journal of Economic Behavior & Organization, Elsevier, vol. 121, pages 77-98, 2016.

CESAR, V; de ONIS, Mercedes; CURI HALLAL, Pedro; BLÖSSNER, Monika; SHRIMPTON, Roger. **Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions.** Pediatrics 125(3):473-480, 2010.

CHAY, K. Y.; McEWAN, P. J.; URQUIOLA, M. **The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools**, American Economic Review 95(4), 1237- 1258, 2005.

CHETTY, R.; FRIEDMAN, J. N.; ROCKOFF, J. E. **Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood**, American Economic Review 104(9), 2633-79, 2014.

CLOTFELTER, C. T.; GLENNIE, E. J.; LADD, H. F.; VIGDOR, J. L. **Teacher Bonuses and Teacher Retention in Low-Performing Schools: Evidence from the North Carolina $1,800 Teacher Bonus Program**, Public Finance Review 36(1), 63-87, 2008.

CONTRERAS, D.: RAU, T. **Tournament Incentives for Teachers: Evidence from a ScaledUp Intervention in Chile**, Economic Development and Cultural Change 61(1), 219-246, 2012.

COOK, P., DODGE, K., FARKAS, G., FRYER, R. J., GURYAN, J., LUDWIG, J., MAYER, S., POLLACK, H., STEINBERG, L. **Not too late: Improving academic outcomes for disadvantaged youth**. Tech. rep, 2015.

COUNCIL FOR EARLY CHILD DEVELOPMENT. **The Science of Early Child Development** CECD, Vancouver, Canada, 2010

CRUSH, Jonathan; RAIMUNDO, Ines; SIMELANE, Hamilton; CAU, Boaventura and DOREY, David. **Migration-Induced HIV and AIDS in Rural Mozambique and Swaziland**, Southern African Migration Programme and International Organization for Migration, Idasa, Cape Town, 2010.

CUNHA, F., HECKMAN, J. J., LOCHNER, L. **Interpreting the Evidence on Life Cycle Skill Formation**. Vol. 1 of **Handbook of the Economics of Education**. Elsevier, Ch. 12, pp. 69/812. 2006.

CUNHA, F., J. HECKMAN, L. LOCHNER, D. MASTEROV. "**Interpreting the Evidence on Life Cycle Skill Formation.**" 2005. NBER Working Paper 11331, Cambridge, MA, 2005.

CUNHA, Flávio. **Investments in Children When Markets Are Incomplete**, mimeo, 2013.

CUNHA, Flávio; ELO, Irma; CULHANE, Jennifer. **Eliciting Maternal Expectations about the Technology of Skill Formation**. NBER Working Paper No. 19144, 2013.

CUNNINGHAM, W. M., VERDÚ, L. G., TESLIUC, R., VERNER, C., DORTE. **Youth at Risk in Latin America and the Caribbean: Understanding the Causes, Realizing the Potential.** World Bank, Washington, DC, 2008.

DEATON, A.; ZAIDI, S. **Guidelines for Constructing Consumption Aggregates.** LSMS Working Paper No. 135. World Bank, Washington, DC, 2002.

DEE, T.S., DOBBIE, W., JACOB, B.A.; ROCKOFF, J.. **The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations**, CEPA Working Paper No.16-08, 2016.

DYNARSKI, M., CLARKE, L., COBB, B., FINN, J., RUMBERGER, R., SMINK, J.. **Dropout prevention: A practice guide.** ZEW Discussion Papers 2008-4025, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, 2008.

DYNARSKI, M., GLEASON, P. **How can we help? what we have learned from recent federal dropout prevention evaluations.** Journal of Education for Students Placed at Risk (JESPAR) 7 (1), 43/69, 2002.

ENGLE, P; BLACK, M.; BEHRMAN, J.; CABRAL DE MELLO, M ; GERTLER, P; KAPIRIRI, L; MARTORELL, R; YOUNG, M and the INTERNATIONAL CHILD DEVELOPMENT STEERING GROUP. **Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world** Lancet 2007; 369: 229–42, 2007.

ENGLE, P., FERNALD, P., ALDERMAN, H., BERHAM, J., O'GARA, C., YOUSAFZAI, A., CABRAL DE MELLO, M., HIDROBO, M., ULKUER, N., ERTEM, I., ILTUS, S., and the GLOBAL CHILD DEVELOPMENT STEERING GROUP. **Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries**. Lancet 2011 (11): 1-15, 2011.

FERNALD, L., KARIGER, P., ENGLE, P., and RAIKES, A. **Examining early child development in low-income countries: A toolkit for the assessment of children in the first five years of life**. The World Bank. Washington, D.C, 2009.

FERSHTMAN, Chaim; GNEEZY, Uri. **The tradeoff between performance and quitting in high power tournaments** Journal of the European Economic Association April 2011 9(2):318–336, 2011.

FILMER, D.; PRITCHETT, L.. **Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India**. Demography 38(1):115-132, 2001.

FILMER, D.; PRITCHETT, L.. **The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries.** Population and Development Review 25(1): 85–120, 1999.

FRYER, R. **Teacher Incentives and Student Achievement: Evidence from New York City Public Schools**, Journal of Labor Economics 31(2), 373-427, 2013.

FRYER, R. G. **Financial incentives and student achievement: Evidence from randomized trials.** The Quarterly Journal of Economics 126 (4), 1755/1798, 2011.

GARCIA, Marito ; VIRATA, Gilian; DUNKELBERG, Erika**The state of Young Children in Sub Saharan Africa** in GARCIA, Marito, PENCE, Alan, EVENS, Judith (org) **Africa's Future, Africa's Challenge: Early Childhood Development in Sub-Saharan Africa**, World Bank, Washington, DC, 2008.

GELMAN, A.; IMBENS, G. **Why High-order Polynomials Should not be Used in Regression Discontinuity Designs**, Technical report 20405, National Bureau of Economic Research, 2014.

GELMAN, A.; ZELIZER, A. **Evidence on the deleterious impact of sustained use of polynomial regression on causal inference**, Research & Politics 2(1), 1-7, 2015.

GERTLER, P. J.; FERNALD, L.C. **The Medium Term Impact of Oportunidades on Child Development in Rural Areas**. Mimeo, 2004.

GERTLER, P.; HECKMAN, J; PINTO, R; ZANOLINI, A; VERMEERSCH, C; WALKER, S., CHANG, S. and GRANTHAM-MCGREGOR, S. **Labor market returns to an early childhood stimulation intervention in Jamaica** Science vol 344 no 6187 pp 998-1001, 2014.

GERTLER, Paul; LEVINE, David ; MARTINEZ, Sebastian. **The Presence and Presents of Parents: Do Parents Matter for More than Money?** ,mimeo, 2003.

GINTHER, Donna; POLLAK, Robert. **Family Structure and Children's Educational Outcomes: Blended Families, Stylized Facts, and Descriptive Regressions** Demography 41(4),671-96, 2004.

GLAESER, E. L., SACERDOTE, B. I., SCHEINKMAN, J. A. **The Social Multiplier**. Journal of the European Economic Association 1 (2-3), 345/353, 2003.

GLEWWE, P.; ILIAS, N.; KREMER, M. **Teacher Incentives**, American Economic Journal:Applied Economics 2(3), 205-27, 2010.

GORMLEY, Willian; PHILLIPS, Deborah ; GEYER, Ted. **Pre-School Programs Can Boost School Readness** Science, vol 320, 2008

GRANTHAM-MCGREGOR S; CHEUNG YB; CUETO S; GLEWWE P, RICHTER L; STRUPP B.; INTERNATIONAL CHILD DEVELOPMENT STEERING GROUP **Developmental potential in the first 5 years for children in developing countries**. Lancet 2007; 369 (9555): 60–70, 2007

HECKMAN, J.J.; MOON, S.H.; PINTO, R.; SAVELYEV, P. A.; YAVITZ, A. **The rate of return to the High Scope Perry Preschool Program** Journal of Public Economics, Elsevier, vol. 94(1-2), pages 114-128, February, 2010.

HECKMAN, J. **The economics, technology, and neuroscience of human capability formation** PNAS August 14, 2007 vol. 104 no. 33, 2007.

HECKMAN, J. J. **Schools, Skills, and Synapses.** Economic Inquiry 46 (3): 289–324, 2008.

HECKMAN, James, STIXRUD, Jora; URZUA, Sergio. **The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior**, Journal of Labor Economics, 24, issue 3, p. 411-482, 2006.

HOPKINS, Kenneth D., BRACHT, Glenn H.. **Ten-Year Stability of Verbal and Nonverbal IQ Scores.** American Educational Research Journal, 12(4): 469–77, 1975.

HOXBY, C. **Peer Effects in the Classroom: Learning from Gender and Race Variation**. NBER Working Papers 7867, National Bureau of Economic Research, Inc, 2000.

IMBERMAN, S. A., KUGLER, A. D., SACERDOTE, B. I.. **Katrina's children: Evidence on the structure of peer effects from hurricane evacuees.** American Economic Review 102 (5), 2048/82, 2012.

INSTITUTO NACIONAL DE ESTATÍSTICA; MINISTÉRIO DA SAÚDE; MEASURE DHS, **Mozambique DHS, Inquérito Demográfico e de Saúde 2011-Final Report (Portuguese)**, ICF International, Calverton, Maryland, USA. March, 2013

INSTITUTO NACIONAL DE SAÚDE; INSTITUTO NACIONAL DE ESTATÍSTICA; ICF MACRO. **Avaliação Nacional de Prevalência, Riscos Comportamentais e Informação sobre o HIV e SIDA em Moçambique**. Calverton, Maryland, 2009.

INSTITUTO NACIONAL DE SAÚDE; INSTITUTO NACIONAL DE ESTATÍSTICA; ICF MACRO. **Inquérito Nacional de Prevalência, Riscos Comportamentais e Informação sobre o HIV e SIDA em Moçambique**. Calverton, Maryland, 2010.

JACKSON, C. K. **A little now for a lot later: A look at a Texas advanced placement incentive program.** Journal of Human Resources 45 (3), 591/639, 2010.

JACOB, B. A., **Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools**, Journal of Public Economics 89(5-6), 761-796. 23, 2005.

JACOB, B.; A., LEFGREN, L.. **The effect of grade retention on high school completion.** American Economic Journal: Applied Economics 1 (3), 3358, 2009

JANUS, Magdalena; OFFORD, David. **Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness.** Canadian Journal of Behavioral Science 39(1), 1-22, 2007.

JANUS, Magdalena; BRINKMAN, Sally; DUKU, Eric; HERTZMAN, Clyde; SANTOS, Robert; SAYERS, M.; and SHROEDDER, Joanne. **The Early Development Instrument: A Population-based Measure for Communities**. Offord Center for Child Studies. McMaster University, 2007.

JARAMILLO, Adriana; MINGAT, Alain. **Early childhood care and education in Sub-Saharan Africa: What would it take to meet the Millennium Development Goals?** in GARCIA, Marito, PENCE, Alan, EVENS, Judith (org) **Africa's Future, Africa's Challenge: Early Childhood Development in Sub-Saharan Africa**, World Bank, Washington, DC, 2008.

KAHNEMAN, D., KNETSCH, J. L., THALER, R. H. **Anomalies: The endowment effect, loss aversion, and status quo bias.** Journal of Economic Perspectives 5 (1), 193 206, 1991.

KAHNEMAN, Daniel; TVERSKY, Amos. **Prospect Theory: An Analysis of Decision under Risk.** Econometrica, 47(2): 263–91, 1979

KAHNEMAN, Daniel; KNETSCH, Jack L; THALER, Richard. **Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias.** Journal of Economic Perspectives, 5(1): 193-206, 1991.

KANE, T. J.; STAIGER, D. O., **The Promise and Pitfalls of Using Imprecise School Accountability Measures**, Journal of Economic Perspectives 16(4), 91-114, 2002.

KAZIANGA, H., de WALQUE, D., ALDERMAN, H.. **Educational and child labour impacts of two food-for-education schemes: Evidence from a randomized trial in rural Burkina Faso.** Journal of African Economies 21 (5), 723760, 2012.

KHAN, Adnan Q.; KHWAJA, Asim I.; OLKEN,Benjamin A.. **Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors.** NBER Working Paper 20627, 2014.

KOCH, Alexander; NAFZIGER, Julia; NIELSEN, Helena Skyt. **Behavioral economics of education** , Journal of Economic Behavior & Organization, Volume 115, 2015, Pages 3-17, 2015.

KOSZEGI, B; RABIN, M.. **A Model of Reference-Dependent Preferences.** Quarterly Journal of Economics, 121(4): 1133–65, 2006.

LAVY, V. **Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement**, Journal of Political Economy 110(6), 1286-1317, 2002.

LAVY, V. **Performance Pay and Teachers' Effort, Productivity, and Grading Ethics**, American Economic Review 99(5), 1979-2011, 2009.

LEROY, J., GADSDEN, P., and GUIJARRO, M.. **The impact of day-care programs on child health, nutrition, and development in developing countries: a systematic review**. Journal of Development Effectivenss. Vol 4 Iss 3, 2012, 2012.

LEVITT, S.; LIST, J.. **Homo Economicus Evolves.** Science, 319(5865): 909–10, 2008.

LIST, J. **Does Market Experience Eliminate Market Anomalies?** Quarterly Journal of Economics, (2003), 118(1), pp. 41-71, 2003.

LLEWELLYN, Deborah. **Program Operations Manual**. Unpublished manuscript, 2008.

LOCHNER, L., MORETTI, E.. **The effect of education on crime: Evidence from prison inmates, arrests, and self-reports.** American Economic Review 94 (1), 155189, 2004.

MANACORDA, M.. **The cost of grade retention.** Review of Economics and Statistics 94 (2), 596606, 2010.

MARTORELL R., KHAN L. K., SCHROEDER, D. G.. **Reversibility of stunting: epidemiological findings in children from developing countries.** European Journal of Clinical Nutrition 48(1): 45-57, 1994.

MURALIDHARAN, Karthik; SUNDARARAMAN, Venkatesh. **Teacher opinions on performance pay: Evidence from India**, Economics of Education Review 30(3), 394 – 403, 2011

MURALIDHARAN, Karthik; SUNDARARAMAN, Venkatesh. **Teacher Performance Pay: Experimental Evidence from India.** Journal of Political Economy 119 (1): 39-77, 2011.

MWAURA, Peter A. M.; SYLVA, Kathy; MALMBERG, Lars-Erik **Evaluating the Madrasa Preschool Programme in East Africa: A Quasi-Experimental Study**, International Journal of Early Years Education, Volume 16, Number 3, October 2008, pp. 237-255(19), 2008.

NAUDEAU, Sophie; KATAOKA, Naoko; VALERIO Alexandra; NEUMAN, Michelle J.; ELDER, Leslie K.. **Investing in Young Children: An Early Childhood Development Guide for Policy Dialogue and Project Preparation.** Washington, DC: World Bank., 2010.

NAUDEAU, Sophie; MARTINEZ, Sebastian; PREMAND, Patrick; FILMER, Deon **Cognitive Development among Young Children in Low-income countries**. In ALDERMAN, Harold, ed., **No Small Matter: The Impact Of Poverty, Shocks, And Human Capital Investments In Early Childhood Development**, Washington, DC: World Bank, pp. 9-50, 2011

NELSON, C. A.; de HAAN M., THOMAS. K. M. **Neuroscience and Cognitive Development: The Role of Experience and the Developing Brain.** New York: John Wiley, 2006.

NEWPORT, Elissa L. "**Critical Periods in Language Development**". in NADEL,Lynn, ed., **Encyclopedia of Cognitive Science**. London: Macmillan Publishers Ltd./Nature Publishing Group, pp.737-740, 2002.

NORES, M.; BARNETT, W.S. **Benefits of early childhood interventions across the world: (Under) Investing in the very young.** Economics of Education Review, 29(2), pp. 271-282, 2010.

OLKEN, Benjamin A.; ONISHI Junko; WONG Susan. **Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia.** American Economic Journal: Applied Economics 6 (4): 1-34, 2014.

OREOPOULOS, P. **Do dropouts drop out too soon? wealth, health and happiness from compulsory schooling.** Journal of Public Economics 91 (2007), 2213 /2229, 2007.

OREOPOULOS, P. **Estimating average and local average treatment effects of education when compulsory schooling laws really matter.** American Economic Review 96 (1), 152175, 2006.

PAXSON, Christina; SCHADY, Norbert. **Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health, and Parenting,** Journal of Human Resources, 42(1) 49-84, 2007.

POPE, D.; SCHWEITZERIS, M. **Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes**, American Economic Review 101 (February 2011): 129–157, 2011.

RAO N; SUN J; PEARSON V; LIU H; CONSTAS MA; ENGLE PL. **Is something better than nothing? An evaluation of early childhood programs in Cambodia.** Child Dev. 2012 May-Jun;83(3):864-76, 2012.

RIZZO, John; ZECKHAUSER, Richard **Reference Incomes, Loss Aversion, and Physician Behavior.** The Review of Economics and Statistics LXXXV, 909-922, 2003.

ROLNICK, A.J.; GRUNEWALD, R.. **The Economics of Early Childhood Development as Seen by Two Fed Economists.** Community Investments 19 (2): 13-14, 30, 2007.

RUTSTEIN, Shea O.; JOHNSON, Kiersten. **The DHS Wealth Index**. DHS Comparative Reports No. 6. Calverton, Maryland: ORC Macro, 2004.

SAAVEDRA, Juan Esteban. **The effects of conditional cash transfer programs on poverty reduction, human capital accumulation and wellbeing**, mimeo, 2016

SAAVEDRA, Juan Esteban; GARCIA, Sandra **Educational Impacts and Cost-Effectiveness of Conditional Cash Transfer Programs in Developing Countries: A Meta-Analysis**. Working paper EGOB No. 3, 2013.

SANCHEZ, Manolo; NORDANG, Hakon. **Follow up Study of Primary Education and Baseline Study of Secondary Education, Mozambique- Final report**, mimeo, 2007.

SCHADY, Norbert. **Early Childhood Development in Latin America and the Carrbean**. World Bank Policy Research Working Paper, 2006.

SHONKOFF, Jack; PHILLIPS, Deborah (eds). **From Neurons to Neighbourhoods: The Science of Early Childhood Development**, National Academy Press, Washington, DC, 2000.

SOLON, Gary; HAIDER, Steven J.; WOOLDRIDGE, Jeffrey M.. **What Are We Weighting For?**", mimeo, 2013.

UNICEF. **The State of the World's Children 2004.** New York, 2004

UNICEF. **The State of the World's Children 2006.** New York, 2006

SPEAR, L. **The adolescent brain and age-related behavioral manifestations.** Neuroscience Biobehavioral Review 24, 417/463, 2000.

SUDFELD C.R.; McCOY D.C.; DANAEI G.; FINK G.; EZZATI M.; ANDREWS K.G., FAWZI W.W.. **Linear growth and child development in low- and middle-income countries: a meta-analysis.** Pediatrics. May;135(5):e1266-75, 2015.

THALER, Richard H.; TVERSKY, Amos; KAHNEMAN, Daniel; SCHWARTZ, Alan. **The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test**. Quarterly Journal of Economics, 112(2): 647–61, 1997.

THAPA, A.; COHEN, J.; GUFFEY, S.; HIGGINS-D'ALESSANDRO, A., **A Review of School Climate Research**, Review of Educational Research 83(3), 357-385, 2013.

UNITED NATIONS DEVELOPMENT PROGRAMME. **Human Development Report 2013: The Rise of the South: Human Progress in a Diverse World.** Undp, 2013.

VEGAS, E.; SANTIBANEZ, L.. **The Promise of Early Childhood Development in Latin America and the Caribbean** The World Bank, Washington DC, 2010.

VERSOSA, Cecilia Cabañero. **Counting on Communication: The Uganda Nutrition and Early Childhood Development Project** World Bank Working Paper no 59, 2005.

VERSOSA, Cecilia Cabañero **Uganda's Nutrition and Early Child Development Project -Counting on Communication** Findings Good Practice Infobrief; Africa Region, The World Bank, no 11, 2005.

WALKER S; WACHS T; GRANTHAM-MCGREGOR S.; BLACK, M.; NELSON, C.; HUFFMAN, S.; BAKER-HENNINGHAM, H.; CHANG, S.; HAMADANI, J.D.; LOZOFF, B.; MEEKS GARDNER, J.; POWELL, C.A.; RAHMAN, A.; AND RICHTER, L. **Inequality in Early Childhood: Risk and Protective Factors for Early Child Development.** Lancet 2011;(11):1-14, 2011.

WORLD BANK **Project Appraisal Document for the Mozambique Education Sector Support Project**. http://go.worldbank.org/MJSPTFYM70. Accessed on November 8th, 2011.

WORLD BANK. **Mozambique - Education Sector Support Project.** Washington, DC: World Bank, 2011

WORLD BANK. **World Development Indicators**. World Bank, Washington, DC, 2013.

5
Appendix Chapter 1

5.1.Save the Children's preschool program in Gaza Province

5.1.1.Program description

Save the children's preschool program main goal is to make all participating children (aged 3 to 5) grow to their full potential. Main actions include the community mobilization, the formation of a local preschool committee, the construction of preschool infrastructure, the training of the animadoras, regular parenting meetings and daily activities to support positive social and emotional development of children.

The first step for the implementation of the program is the initial contact between Save the Children and the community. Following the initial contacts, a series of meetings between Save's representatives, community leaders, caregivers and other community members take place, at an attempt to learn about the local culture and to discuss the best format of the program. At this phase, community members commit to arrange a space for the construction the escolinhas, to mobilize parents and caregivers to enroll their children and to form a management committee.

Preschool classrooms are built by local members of the community, who not only provide labor, but who can also provide local materials, while Save the Children provides financial and logistical support. Classrooms have a thatched or tin roof depending on each community's context, with cement floor. Each escolinha counts with latrines and hand washing stations, and an outside playground built with local materials. Each community has up to 3 classrooms, with a maximum capacity of 105 children, aged 3 to 6 years, from the local community.

The committee is composed by 8 to 12 members appointed by the community. Committee members' tasks include mobilizing local laborers to construct the

escolinhas, managing parents' contributions and monitoring the health status of children.

Each class counts with a team of two animadores (pre-school teachers), resulting in an average pupil-teacher ratio of around 15 children per animadora. The animadores are selected by the community and they are required to have completed at least primary school (5 years of education). Animadoras are usually females and receive 10 USD per month from Save the Children. Community members are incentivized to provide monetary and in-kind contributions, as feasible. Animadores are prepared to have a positive and warm interaction with children, creating a welcome environment for learning. The animadores training is facilitated by education specialists. Training takes one week and uses experiential learning methodologies in which new knowledge is introduced, modeled, practiced, and reflected upon. Animadores learn about children's holistic development and activities to support children's development. A daily routine is introduced, modeled, and practiced mimicking a real escolinhas environment.

After the foundational training, animadores receive support on the escolinha's opening day and mentoring visits by the community development agents at least once a month. Animadores participate in learning circles facilitated by the community development agents with the objective to consolidate foundational skills and learn additional activities.

Children spent 3 to 4 hours per day at the escolinhas. Children are not clustered by age; instead assignment to classes is made at random. Classes take about 3 to 4 hours per day, but specific time of operation is left up to community. Usually classes are taken by the morning. Escolinhas don't provide food to children, as Save the Children's previous experience in the region has shown that food supplementation may could cause parents to view the program as a feeding service rather than learning program.

The daily routine and the classroom structure are prepared to support positive social and emotional development. Classes are taken in the local language, Changana, but children are introduced to Portuguese to make a successful transition to primary school.

Children begin each class by greeting each other, showing his/her attendance, washing hands and singing a song. During class, they are continuously stimulated to listen to and discuss stories, tell about personal experiences, draw pictures, play games with alphabet letters, label, align and group objects, among other activities.

The physical development is enhanced by outdoor play and activities to develop hands and fingers muscles. When playing outside, children can dance, walk, hop or jump, thus enhancing gross motor, social and emotional skills.

Thinking skills are the foundation upon which children learn to make decisions, regulate their own behavior, meet complex challenges and take responsibility for their actions. Thinking is developed in the escolinhas by planning and reflecting about news or about what they did learn during the class.

Activities like story telling, news sharing and alphabet learning are also practiced. These activities are are thought as tools to develop not only language and communicating skills, but also the capacity of thinking and reasoning.

Hand washing with soap, safe drinking water and access to latrines are absolute and non-negotiable components of the program. Toys are adapted to the communities' environments using local materials and are made by community members.

Children are daily exposed to math through everyday play activities. They have materials to count, sort, compare, match, put together and take apart. Rhythm and sequence games, calendar activities, and measuring sand and water with cups also provide informal exposure to math concepts.

Save the Children facilitates meetings in which positive parenting practices are discussed between parents and caregivers of participating children. Parents are motivated to participate in monthly meetings, creating an opportunity for them to discuss childrearing and to learn from one another.

Meetings are facilitated by the animadores, the community development agents and a parent of the day, who is chosen by the community. Some of the topics discussed at parent meetings include breastfeeding, de worming, nutrition, child development, literacy, playing with children, attendance and demand driven components defined by the community. Topics are discussed using an appreciative

inquiry approach in which knowledge is built from existing positive parenting practices. Harmful practices are also brought to light, and strategies to change those practices are discussed.

Community development agents also advice parents how to make simple toys at home with cheap materials. They also show parents example of cognitive stimulating games, like putting together pieces of wood and asking their children to reorder them from the smallest to the bigger one.

One of the main short outcomes of the escolinhas is making a successful transition to primary school. Many children who begin schooling without going to pre-school tend to fear teachers, cry at class and have lower performance in the first years of school. The language spoken at primary school imposes an additional complication to children in Gaza. As it occurs in many other African countries, the language in which classes are taken is not the mother tongue of children. Although children in Gaza learn Changana at home, according to the law, they must be instructed in Portuguese in primary school. The lack of bilingual instruction in primary school can delay gains in verbal and communicating skills in the first schooling years.

Although classes at escolinhas are taken in Changana, children are early introduced to Portuguese and learn reading and writing simple words in Portuguese. Primary school welcome days are scheduled for children from the escolinhas to visit their future primary school, while primary teachers are invited to visit the escolinhas and early get in contact with children and their families.

5.1.2. Preschool daily routine

Greetings (15 minutes):

1- At the beginning of the day, each child must turn on a card with their own drawing to her name to show her attendance.
2- Children wash hands before entering the classroom.
3- Teacher greets each child.
4- The class reviews the attendance chart.
5- Teacher identifies the "Child of the Day" and invites him/her to help lead a song or game.

Literacy Circle (50 minutes)

1- News Sharing (Mon/Wed); Journals (Tue/Thu); Theme Journal (Fri) (20 minutes)
2- Story time (storybook or oral story telling (15 minutes)
3- Rhymes or Song (5 minutes)
4- Alphabet Activity – one letters per week (10 minutes)

Corner Play (1 hour)

1- Children play in the 5 corners (Games & Puzzles; Imagination; Construction; Books and Pictures; and Sand and Water Play (outside of the classroom)
2- Teacher observes the children and talks with them (non-instructional talk)

Math Circle (4 days)/Cultural Day (1 day) (25 minutes)

1- Calendar activity, Days of the Week (5 minutes)
2- Lesson using Math bags (20 minutes)
3- Counting Song/Rhyme (as time allows)
4- On Fridays, Math Circle and Outdoor Play are replaced for one hour of Cultural Day

Outdoor Play (30 minutes)

Children play outside freely or with a game organized by the animador.

Children wash their hands before re-entering the classroom.

Closing/Review (15 minutes)

Clean-up (about 10 minutes)

5.1.3. Example of math activities

Table 1-Sixteen preschool math lessons

Material	Activity description	Goal
1.Math bag – soda bottle caps	Take out the small work mat. Empty the soda bottle caps on the mat and sort them into groups so that all the ones of the same color are together. Ask the children to talk about how they sorted the caps. Can we name the colors of the groups?	Sorting – noticing color differences
2. Math bag- seeds	Children use 10 seeds to make a design anyway they like. Next, they work in pairs. Child # 1 makes a design with the seeds. Then Child #2 tries to make the same design. Next Child #2 makes a design for Child # 1 to copy.	Noticing color and design differences
3. Math bag – soda bottle caps	Pick two colors of tops. Lay them in a straight row. Make a pattern such as orange green orange green…	Patterns & colors
4. Math bag – seeds	Children make a pattern using only seeds of 1 color. Now they can turn add 5 seeds of another color to with two-colors. Finally, they can play with the seeds anyway they like.	Following directions; noticing colors; making patterns of two colors
5.Cubes from block corner	Give each child 5 cubes. Let them play with them for a few minutes, building anything they like. Now ask them to work with a partner to make two towers that are equal or the same size. Now ask them can they make two towers one that has more and one that has less. They should be able to point out which is less and which is more.	Concepts: same/equal; more, and less
6. Cubes from block corner	Each child has 5 cubes. Count them together. Let them play as they like. Notice if any make a pattern as they did with the caps. Point this out to the others. Then give them a challenge. The teacher uses a child and an object to show *above* and *below.* Work with a partner and find some other ways to show above and below.	Review patterns, Concept: Above and below
7. Math bag- Soda bottle caps	We sorted the bottle caps into groups by colors. Can you find another way that the caps are alike or different and put them in groups to show ones that are the same? Teacher gives time to explore. Then she asks student to explain what they did. She encourages others to find more ways to sort the caps. The students explain.	Sorting & concepts large and small, more or less

	Examples might be caps with words; caps that are bent; caps that are new and caps that are old.	
8. Cubes from blocks corner	Children work in pairs with 3 blocks each of 3 shapes. One child stacks blocks, to make a pattern, and then asks his friend what comes next? The other child makes a pattern stacking blocks and asks which comes next? They continue playing this game until time is up. Teacher then asks them how many blocks of each shape do they have.	Patterns; and understanding the number 3
9. Math bag- soda bottle caps	The children again work in pairs to solve a problem using 3 colors of 3 soda bottle caps. The teacher writes the number 3 on the board. The children draw number 3 in the air. The teacher poses a problem to solve. Show all the different ways you can combine the caps to make 3. Example 1 red cap and 2 blue caps. After they are finished, the teacher asks for someone to tell one way they solved the problem. The student shows the others. The teacher asks the students to raise their hands if they made the same combination. The teacher then asks another student to different way. Again, she asks all students to look at their caps and see if they solved the problem the same way. The teacher says something like, “Oh look we have 1+1+1 or we have 1 + 2 = 3. She uses the language of addition but she does not write it as such at this early stage.	First experience with the meaning of addition and how things add up to make 3. The teacher does not call it addition at this early stage. Children also notice things that are same and different.
10. Math bag – Toothpicks	Today take out your toothpicks. Yesterday we worked with 3 cubes. Today let’s try something different with 3. Show all the different ways you can make a design with 3 toothpicks. The sticks have to touch each other in the design. When the students are finished, asked them to work in pairs. They are given this problem: Look at your friend’s toothpicks. Make sure that every design is different and no two are the same. If they are the same, the friend has to put the toothpicks back in the box. The second student does the same thing, looking carefully at his friend’s toothpicks. When everyone is finished, the teacher asks. How many had more than one design? How many had more than two? Three? Four? Five?	Solving problems with the number 3. Creative design, developing concept of “3ness”, noticing things that are different; practicing identifying “more than”; practice counting together sets of

		things up to five.
11. Math bag- string + soda bottle caps	The teacher shows the circle shape. The children say the name and make a circle with the string. Next, they take out the caps and lay the caps on the string to make a cap circle.	Shape - circle
12. Math bag- Soda bottle caps	The children remember the shape called circle. They now use two colors of caps and make a pattern with the two colors and form the pattern in the shape of a circle. They can make other cap pattern circles with other colors.	Patterns and circle concept
13. Math bag – string, small rocks and toothpicks	Teacher draws a straight line on the board. She asks students to make a straight line with the string. Next, she shows them a circle shape and asks them to make a circle. Another way to make a curved line (waves). You can make a curve line that looks like ocean waves. Try it with a string. Students now take out seeds and toothpicks. Can you make a long straight-line pattern with seeds and toothpicks? Now can you make the same pattern in a circle shape? Now make a different straight-line pattern. Now make the same pattern in a circle shape.	Straight-line curves and circle.
14. Walk	Today we are going on a circle hunt. I will take a piece of paper. Every time we find a circle I will make a tally mark. The teacher asks the children to hold hands on the walk and follow her. They are to stay in a line. If they see a circle they are to call out to her. If everyone agrees it is a circle, the teacher will make a tally mark. When they return to the room, the teacher draws a large circle on the blackboard. She asks the children to use their journals and draw one circle they remember finding during the walk. Each child gets to show the circle. If this activity is done on paper rather than a slate, it can be made into a book about “Circles”.	Noticing shapes (circle) in the environment. Extending math out of the classroom. Remembering things they have seen. Circle poster on Wall labeled with name circle. Teacher writes number of objects seen on the walk.
15. Math bag Toothpicks	Now take out your toothpicks. Teacher draws a triangle on the board. Asks students if they know the name of the shape. How many sides does the triangle have? Next she poses a problem. See how many triangles you can make using your toothpicks. When everyone is finished, the teacher says, “Let’s all count together.” Now see	Review of 3, introduce triangle, solving a problem with triangles, there are different

	if you can make triangles that connect to each other so that a new triangle shares one side of the first triangle. Use your toothpicks and keep making more and more triangles that connect to each other. When everyone is finished, the teacher admires the different ways students solved the problem. The students look at each other's work.	ways to solve a problem.
16. Math bags – Toothpicks	Independent. What is the smallest triangle you can make with your toothpicks? What is the largest triangle you can make with your toothpicks? Pairs. Together with all your toothpicks, what is the largest triangle you can make? Now separate the toothpicks so that you each have the same number. How can you be sure that each has the same number? Can anyone suggest a method? Children listen to each other's suggestions and then divide the toothpicks into two equal groups. After the teacher checks each one, they put them back in the box.	"3ness", different sizes of triangles, different ways to solve a problem. Creative exploration

5.2. Survey and Tracking

From the full sample, nly 1.2% of children were not located. For remaining children, interviews were either rejected (1.4%), or households moved outside the tracking area, with 1.8% of children moving to South Africa and 0.9% moving outside the province to another part of Mozambique. A total of 18 children were reported as deceased over the period and in those cases the caregiver and household members were interviewed when located.

Table 2

Endline survey household tracking

Survey completed	1897	94,9%
Household not located	23	1,2%
Household located but survey not completed (refusal or other)	27	1,4%
Household moved to South Africa and not tracked	35	1,8%
Household moved outside Maputo or Gaza and not tracked	18	0,9%
TOTAL	2000	100,0%

Table 3

ATTRITION RATE

	Surveys not completed	Attrition rate
Treatment	53	.0494
Control	45	.0488
Difference		.0012
P-value		0.896

Note: Difference and P-Value calculated through simple OLS regression, with dummies for geographical location and randomization blocks, clustered at the community level. A simple t-test yields a p-value of 0.952.

Table 4

SURVEY CONTENT

Module	Description	Baseline Sample	Endline Sample
	Instrument: Household Survey		
Demographic	All Household (HH) members: education, marital status, health conditions	13,608	14,902
Preschool Participation	Children < 12 years old: preschool participation	6,092	5,699
Labor *	Members > 11 years old: Labor market participation (formal, informal, business)	5,759	8,825
Time Use	All household members: Time spent in different activities in the last week	13,608	14,902
Consumption and Transfers	Food and non-food consumption; inter-household transfers	2,000	1,897
Housing Characteristics	Housing materials, access to services (water, sanitation, electricity)	2,000	1,897
Farm Characteristics	Land ownership and use	2,000	1,897
Assets	Durables, production goods, animals	2,000	1,897
Child Health	Target child: Health, vaccination records	2,000	1,897
Anthropometrics	Target child and caregiver height and weight (and youngest sibling in endline)	4,000	4,357
Child Development Tests	Target child: ASQ, TVIP, SDQ (enline only)	2,000	1,897

Caregiver Practices	Caregiver: Parenting practices, activities with target child	2,000	1,897
Satisfaction and Expectations	Caregiver: Satisfaction with child development and expectations about target child future education	2,000	1,897
Health Practices	Caregiver: Health related knowledge and practices	2,000	1,897
Preschool Involvement	Caregiver: Participation in preschool activities (maintenance, management, etc)	2,000	1,897
Social Capital	Caregiver: Participation in meetings, local organizations and relationship with neighbors	2,000	1,897
Time Preferences	Caregiver: Time preferences	2,000	1,897
Missing Mother and Father	Characteristics of missing parents	2,000	1,897
	Instrument: Community Leader Survey		
Personal Information	Individual characteristics (Education, age, tenure as leader)	76	76
Facilities	Community infrastructure and access to services	76	76
Distances	Distances and costs to/from different facilities (school, bank, health center)	76	76
Crops	Information about farms and agricultural activities	76	76
Shocks	Community level shocks in the last 10 years and consequences for community members	76	76
Prices	Cost of basic goods and services (food, education, fuel)	76	76
Satisfaction	Community leader satisfaction with the community's development	76	76
Social Capital	Community leader participation in the community groups/ associations/ meeting and the interaction with the neighbors	76	76
Inheritance	Inheritance common practices in the community, especially with children as beneficiaries	76	76
	Instrument: School Survey		
Principal	Principals information about the Primary School infrastructure, routines, and students	51	55
Teachers	First-grade teachers' information about school routines and students	51	55
EDI	EDI Surveys for sample first graders	1020	919

*Labor module was applied to household members 18 and older at baseline and 12 and older at endline

5.3. Sampling and Randomization

To identify the effect of preschool on children and their families we use an experimental evaluation framework with random assignment of preschools to treatment and control communities. The evaluation sites were selected using operational and logistical requirements determined by Save the Children, which had resources available to build and support preschools in a total of 30 communities. First, three districts in Gaza province (Manjacaze, Xai Xai and Bilene) were selected given Save the Children's operational presence in the area. Based on the organization's capacity for community mobilization, only communities with between 500 to 8000 residents were eligible for the program. Population size and the community listing were drawn from the National Census data of 2007. Additionally, communities needed to be grouped within sufficient geographic proximity so that Save the Children field teams could travel between communities within the same day.

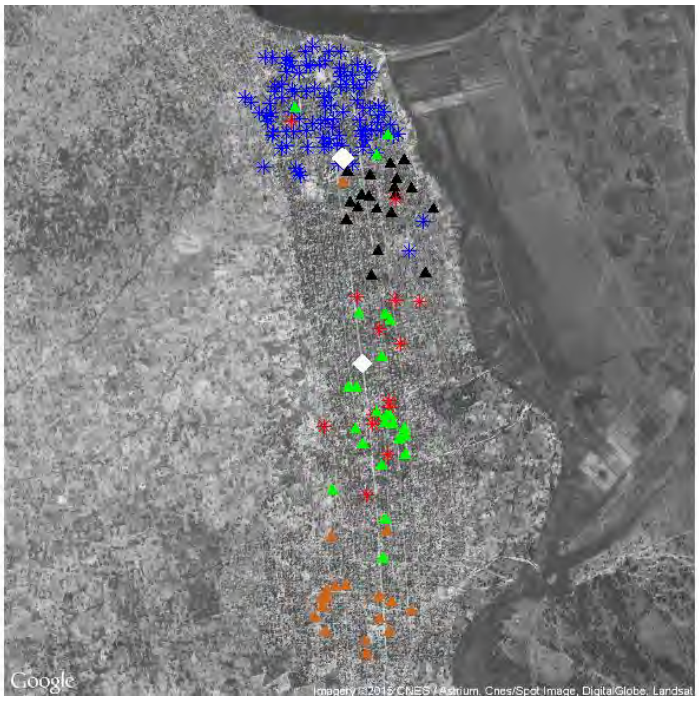
A total of 252 communities were identified in the three intervention districts. After applying eligibility criteria, the number was reduced to 167 communities concentrated in 11 distinct areas. To maximize the number of communities available for the evaluation and ensure the presence of the project in all three districts, the program selected the two areas with the largest number of communities in Manjacaze and Xai Xai, and the single largest area in Bilene, for a total of 5 intervention areas containing 98 villages. For operational reasons, the program required that each area include the same number of treatment communities, which meant assigning 6 treatment communities to each of the 5 areas.

We stratified communities into 37 "blocks" based on population size and then randomly assigned one community to the treatment group within each block. Block randomization was performed to improve balance amongst treatment and comparison groups and increase statistical power. The number of communities per area ranged from 15 to 24. In the two areas with fewer than 18 communities, communities were blocked into pairs while in the three areas with 18 or more communities, communities were blocked into triplets. The two smallest communities that did not form part of a block were dropped from the sample.

Of the 37 blocks, 30 were randomly selected to be offered the program first and 7 blocks were held as replacement in case one or more of the original 30 treatment communities did not accept the program. Once all 30 initially selected treatment communities signaled their interest to participate in the program, the 7 replacement blocks were dropped from the sample, for a total of 76 communities with 30 randomly assigned to treatment and 46 to control.

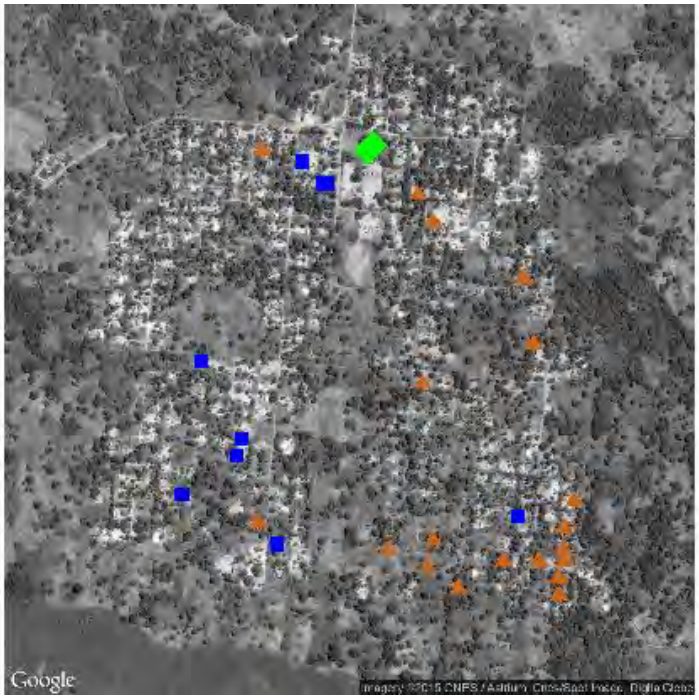
After the randomization took place, however, we noticed that some control communities listed at the National Census are actually just neighborhoods of a bigger community, and although those are treated as separate communities by the official statistics and have different community leaders, escolinhas treated children from all the neighborhoods equally. Consequently, children from all the neighborhoods were offered the treatment. In order to deal with that, we considered those as just one community, and we collapsed their community id and reassigned their treatment dummy to one. In order to preserve the randomization structure, we also merged their randomization block dummy in case 2 or more neighborhoods fell in different blocks. In addition, the boundaries of Nhamavilla Sede and Tetene were indistinguishable. They were merged into one community, as well as their community id and randomization block, and their sampling weight has been recalculated.

Figure 1- Communites of Muzingane B1/ Muzingane B2/ Muzingane B3/ Muzingane B5



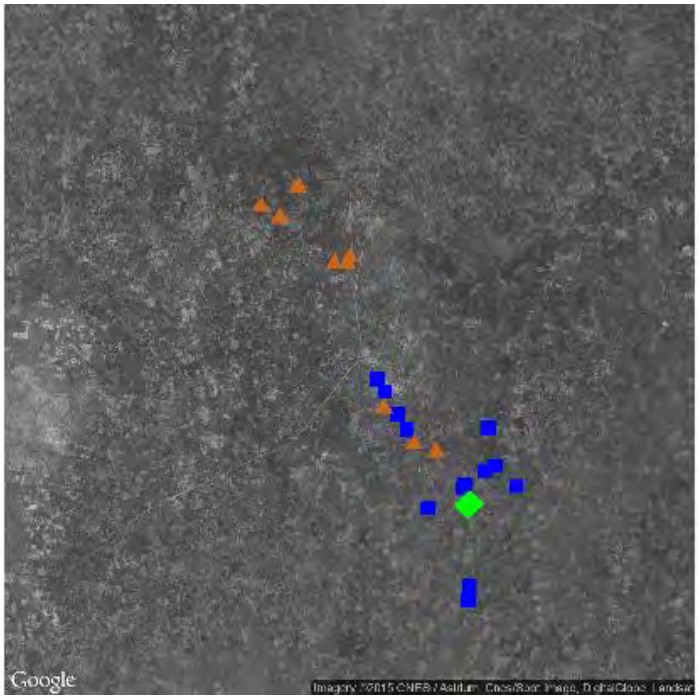
Notes: White diamond: Save the Children's Preschool. Households by color: Orange: Muzingane Bairro 1, Green: Muzingane Bairro 2, Red: Muzingane Bairro 3, Black: Muzingane Bairro 4, Blue: Muzingane Bairro 5. Stars denote original community assignment as a treatment. Triangles denote original community assignment as control.

Figure 2- Communities of Tlacula B1 and Tlacula B2



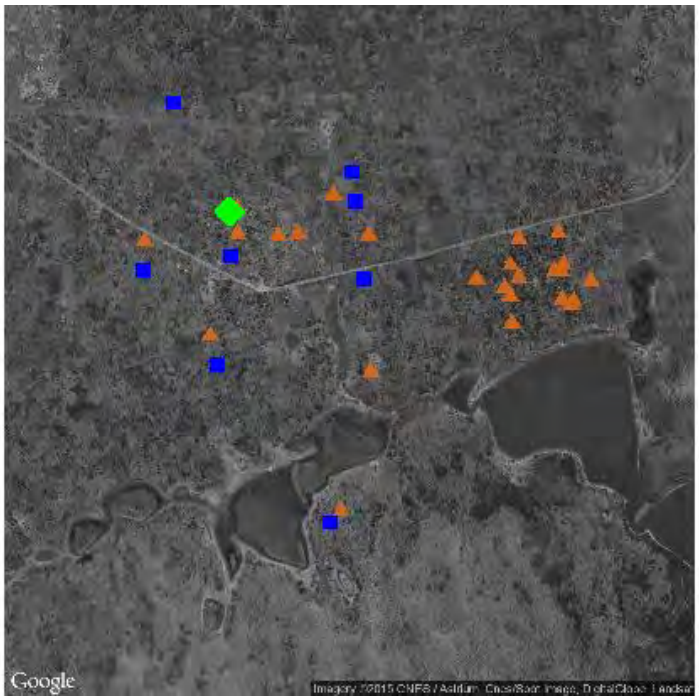
Notes: Green diamonds: Save the Children's preschool. Households: Blue: Tlacula Bairro 1, Orange: Tlacula Bairro 2. Squares denote original community assignment as treatment. Triangles denote original community assignment as control.

Figure 3- Communities of Chitsembe Bairro 1 and Bairro2



Notes: Green diamonds: Save the Children's preschools. Households: Orange: Chtsembe Bairro A, Blue: Chitsembe Bairro B. Squares denote original community assignment as treatment. Triangles denote original community assignment as control.

Figure 4- Communities of Tetene and Nhamavila



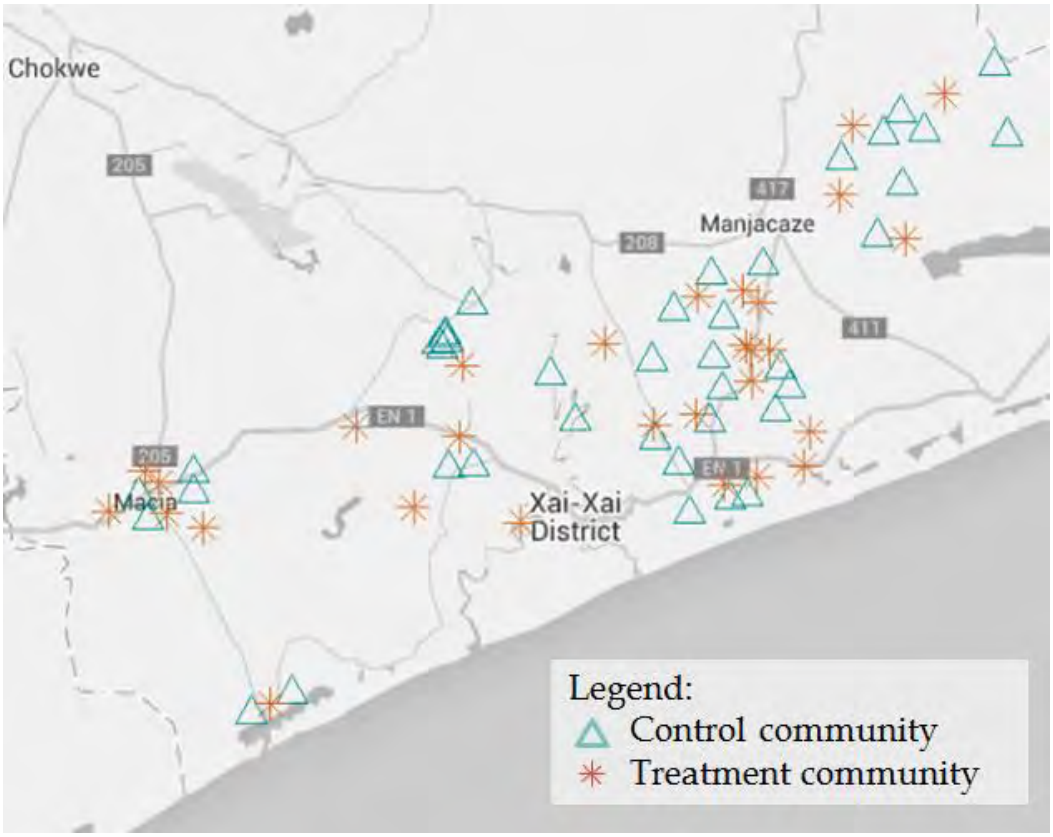
Notes: Green diamonds: Save the Children's preschools. Households: Orange: Tetene, Blue: Nhamavila-Sede. Squares denote original community assignment as treatment. Triangles denote original community assignment as control.

5.3.1. Location of communities

Figure 5- Gaza province in Mozambique

Figure 6- Districts of Bilene, Xai Xai and Mandlakazi in Gaza Province

Figure 7-Location of treatment and control communities



5.3.2.Community treatment assignment

Table 5

COMMUNITY TREATMENT ASSIGNMENT

Community Name	Treatment Status	Other preschool managed by	Community Name	Treatment Status	Other preschool managed by
Aldeia O.M.M.	0	NGO (ADCR)	Tsoveca	0	-
Banhine-Sede	0	Church	Uampaco	0	-
Bocodane	0	-	Armando Tivane	1	-
Chibonzane-Sede	0	-	Bungane	1	-
Chicomo A	0	-	Chibielene	1	-
Chiconela	0	-	Chicavane-Sede	1	-
Chiguitine	0	Catholic Church	Chilumbele B	1	-
Chimutane	0	-	Chimungo	1	-
Chipendane	0	-	Chitsembe A *	1	-
Conjoene	0	-	Chitsembe B	1	-
Dzimba	0	-	Cumbene	1	-
Emília Daússe	0	-	Lipanga	1	-
Gangalene	0	-	Machachuvane	1	-
Gombane	0	-	Machalucuane B/3	1	-
Helane	0	-	Machingane	1	-
Macachene	0	-	Macuane	1	-
Machulane-Sede	0	-	Madjele	1	-
Maciene-Sede	0	Anglican Church	Mahungo	1	-
Mafangue	0	-	Mahuntsane	1	-
Magaíza	0	-	Malene	1	-
Magula	0	-	Mangunze A	1	-
Maniquinique	0	-	Mangunze B	1	-
Mavengane	0	-	Marramine	1	-
Mucindo	0	NGO (World Vision)	Mazucane-Sede	1	-
Muwawasse 2000	0	-	Menguelene	1	-
Muwawasse B/1	0	-	Muchabje	1	-

Muwawasse B/2	0	-	Muzingane B/1 *	1	-
Muwawasse B/3	0	-	Muzingane B/2 *	1	-
Muwawasse B/4	0	-	Muzingane B/3	1	-
Muxaxane	0	NGO (Arpache)	Muzingane B/4 *	1	-
Ncane	0	-	Muzingane B/5	1	-
Nhafoco	0	-	Nhamavila-Sede	1	-
Nhampfuine	0	-	Tetene*	1	-
Nhiuane	0	-	Nhenguene	1	-
Nwachihissa	0	-	Nuvunguene	1	-
Pomulene	0	-	Tlacula B/1	1	-
Tavane	0	-	Tlacula B/2 *	1	-
Vamangue	0	-	Totoe	1	-

Notes: * Originally a control community

5.4. Baseline balance

Table 6

BASELINE BALANCE OF COMMUNITY CHARACTERISTICS

	Treatment mean N=36	Control mean N=40	Means difference	T-stat
Community leader				
Age	64.861	66.750	-1.889	-0.659
Sex: male=1	0.889	0.900	-0.011	-0.156
Speaks Portuguese	0.444	0.475	-0.031	-0.263
Reads and writes	0.861	0.950	-0.089	-1.336
Years of education	3.545	3.686	-0.140	-0.547
Existence of facilities at the community				
Piped water	0.250	0.200	0.050	0.516
Water fountain	0.444	0.475	-0.031	-0.263
Electricity	0.306	0.225	0.081	0.789
Primary school	0.611	0.575	0.036	0.316
Secondary school	0.278	0.350	-0.072	-0.669
Health post	0.611	0.650	-0.039	-0.347
Traditional doctor (curandeiro)	0.889	0.875	0.014	0.185
Bus stop	0.528	0.625	-0.097	-0.850
Paved road	0.167	0.225	-0.058	-0.631
Market-place/fair	0.333	0.205	0.128	1.251
ATM or Bank agency	0.083	0.175	-0.092	-1.176
Existence of organized groups at the community				
Microcredit group	0.167	0.100	0.067	0.851
Orphaned and vulnerable children group	0.333	0.350	-0.017	-0.151
Education council	0.389	0.350	0.039	0.347
Health council	0.444	0.350	0.094	0.834
Economic activity				
Agriculture is the main activity	0.972	0.925	0.047	0.913
Cash crop is the main activity	0.056	0.075	-0.019	-0.337
Facilities distances from leader's house (Kms)				
Closest bus stop	5.139	5.900	-0.761	-0.448
Closest paved road	12.485	27.263	-14.777	-1.843
District center	32.588	37.813	-5.224	-0.549
Closest market place	14.757	18.418	-3.661	-0.788
Closest ATM or banking agency	26.714	34.546	-7.832	-1.022
Closest preschool	8.089	6.631	1.458	0.496
Closest primary school	6.200	5.533	0.668	0.230

Closest secondary school	10.629	14.510	-3.881	-0.923

Notes: Author's calculation from baseline interview with each community leader.

Table 7

BASELINE CHARACTERISTICS BY PRESCHOOL ATTENDENCE OF TARGET CHILD

	Attended preschool N=527	Did not attend preschool N=458	Means Difference	T-Stat
Household characteristics				
Number of household members	4.925	5.179	-0.255	-1.675
Asset index	-0.247	-0.191	-0.056	-0.54
Number of rooms at home	2.075	2.091	-0.016	-0.144
Improved latrine at home	0.113	0.172	-0.059	-1.76
Adobe walls at home	0.668	0.656	0.012	0.275
Dirty floor at home	0.775	0.814	-0.039	-1.196
Target child characteristics				
Female	0.545	0.498	0.047	0.838
Age (years)	3.466	3.459	0.007	0.325
Speaks Portuguese	0.2	0.104	0.096	2.319
Orphaned	0.13	0.091	0.038	1.594
ASQ Total Score	210.25	195.446	14.804	2.967
TVIP raw score	5.883	5.791	0.091	0.224
TVIP score-within sample standardized score	-0.035	0.115	-0.15	-0.871
TVIP score-standardized by developers table	78.564	78.976	-0.412	-0.567
Child had skin problems in the last 4 weeks	0.049	0.203	-0.154	-1.698
Child had swallowing difficulties in the last 4 weeks	0.064	0.027	0.037	1.587
Respiratory illness (flu, pneumonia, asthma) in the last 4 weeks	0.172	0.125	0.047	1.58
Child had diarrhea in the last 4 weeks	0.07	0.059	0.011	0.414
Child slept in mosquito net the night before	0.208	0.12	0.089	2.972
Child has been dewormed in the last 12 months	0.07	0.133	-0.063	-2.672
Child received vitamin A	0.477	0.402	0.075	1.565
Child was diagnosed with malaria in the last 4 weeks	0.095	0.066	0.029	1.399
Weight for age z-score	-0.203	-0.375	0.172	2.236
Height for age z-score	-1.411	-1.611	0.2	1.358
Average weight for height Z-score	1.348	1.245	0.103	0.597
Caregiver characteristics				

Age (years)	35.775	35.688	0.087	0.092
Female	0.928	0.841	0.087	4.126
Speaks Portuguese	0.561	0.436	0.125	3.578
Read and write	0.68	0.556	0.124	2.909
Years of education	3.745	2.884	0.861	3.221
Married or partnered	0.606	0.661	-0.055	-0.962
Reads/skims through books with child	0.581	0.481	0.1	2.607
Plays with child in the garden	0.493	0.411	0.082	1.246
Spends time naming and drawing objects with child	0.378	0.377	0.001	0.04
Plays games with child	0.51	0.384	0.126	3.442
Practices self-sufficiency activities with child	0.616	0.549	0.067	1.664

Notes: T-stats computed through simple linear regression with standard errors clustered at community level. Asset index calculated by principal components using a list of assets at the household. Dirty floor includes mud, sand, and adobe. Within sample standardized TVIP score calculating by subtracting the age in months controls average and dividing the age in months standard deviation.

Table 8

PRESCHOOL PARTICIPATION

	Treatment mean	Control mean	Means difference	T-stat	N
Enrollment (children 3-9)	0.41	0.11	0.31	8.98	3706
Enrollment (target children 3-5)	0.54	0.12	0.42	9.76	1839
Enrollment Age =3	0.32	0.08	0.23	3.81	417
Enrollment Age =4	0.52	0.06	0.46	9.40	398
Enrollment Age =5	0.55	0.12	0.43	8.60	1011
Enrollment Age =6	0.51	0.13	0.39	8.34	875
Enrollment Age =7	0.31	0.13	0.18	2.47	265
Enrollment Age =8	0.16	0.09	0.07	1.59	352
Enrollment Age =9	0.14	0.09	0.05	1.10	388
Enrollment Age =10	0.04	0.08	-0.04	-1.43	440
Enrollment Age =11	0.05	0.06	0.00	-0.16	354
Only target children (3-5):	N=1020	N=887			
Access to Preschool	0.71	0.24	0.48	9.17	1830
Preschool source of funding: Save the Children	0.53	0.09	0.44	8.37	603
Preschool source of funding: Church	0.01	0.29	-0.28	-2.27	603
Preschool source of funding: Government	0.06	0.05	0.01	0.17	603
Preschool source of funding: Community	0.05	0.00	0.05	4.30	603
Preschool source of funding: Other	0.04	0.08	-0.04	-1.23	603
Conditional on enrolling into preschool:	N=527	N=76			
Days per week	4.91	4.75	0.16	1.35	565
Hours per day	3.75	3.99	-0.24	-0.42	522
Travel time (minutes)	21.87	20.64	1.23	0.34	567
Amount paid for preschool (MTN per month)	4.97	23.11	-18.14	-1.50	556
Reasons for not enrolling the target child if had access to preschool:	N=187	N=94			
Child too young	0.09	0.12	-0.03	-0.60	281
Attempted to enroll but child not accepted	0.05	0.13	-0.08	-2.13	281
Distance	0.22	0.15	0.07	1.08	281
Too expensive	0.06	0.08	-0.02	-0.36	281

Note: T-stats computed through simple linear regression with standard errors clustered at community level.

5.5. Sampling weights

One of the man motivations for weighting is to correct for heteroskedastic error terms and improve the precision of estimates. In order to test for heteroskedasticity, we start by regressing preschool participation on the treatment community indicator. We include the contaminated community dummy (indicating that another NGO or a church has built a preschool in the community), a series of controls, location dummies (district and administrative post) and the dummies for randomization blocks. We also cluster at the community level.

We perform one separate regression for the target children and one separate regression for children aged 5 to 9. We take the square of the residuals and plot against the inverse of probability of child selection:

Figure 8 -Square of residuals and probability of selection

Target children

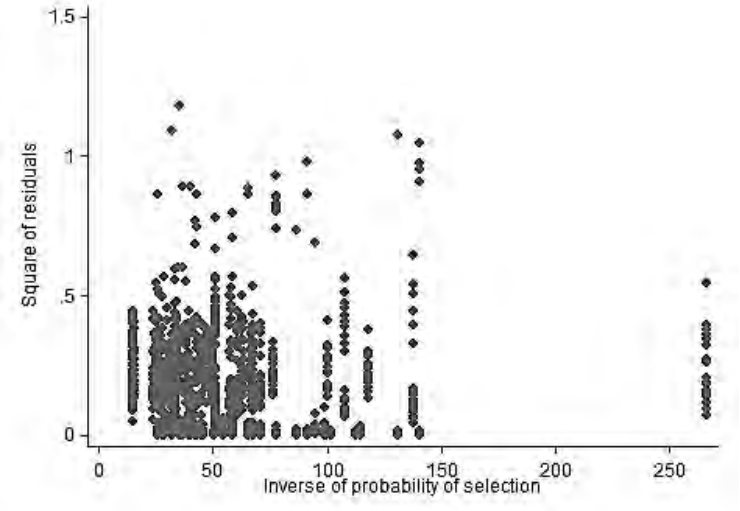
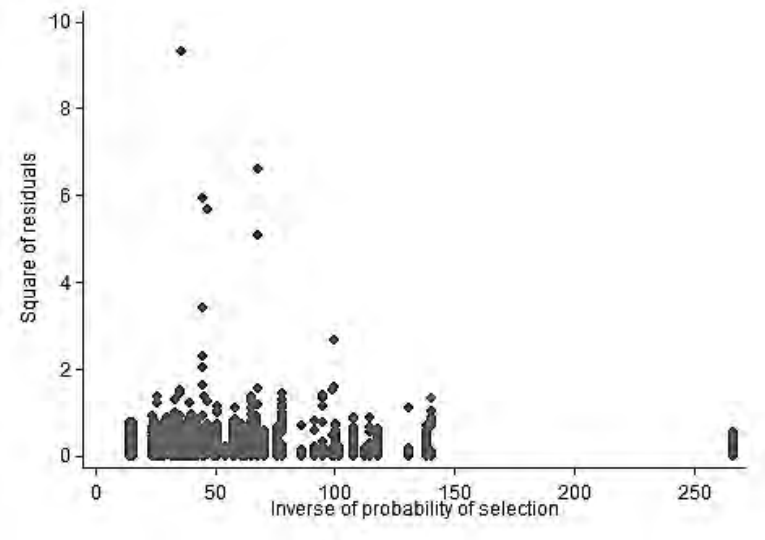


Figure 9- Square of residuals and probability of selection
Children aged 5 to 9



We then regress the square of the residuals on the inverse of the probability of selection, including all the controls and dummies as before. The next table shows the coefficient on the population weights, the constant and the resulting R2 of the regression. As by inspecting the graphs we could suspect that heteroskedasticity could arise from the extreme vale of population weight above 250, we also run a separate regression dropping the community with such weight. For both the 4 regressions, the null hypothesis of homoscedastic errors is strongly rejected.

Table 9

TESTS OF HETEROCEDASTICITY

	Total ASQ Score	TVIP Changana (std)	Time spent studying	Ever enrolled in primary school
Pagan-Hall general test statistic	85.342	111.624	98.904	448.126
P-Value	0.022	0.000	0.000	0.000
Pagan-Hall test with assumed normality	332.747	472.818	89.027	203.515
P-Value	0.000	0.000	0.001	0.000
White/Koenker nR2 test Statistic	78.959	93.942	99.062	388.034
P-Value	0.061	0.001	0.000	0.000
Breusch-Pagan/Godfrey/Cook-Weisberg Statatistic	357.938	522.656	93.765	220.212
P-Value	0.000	0.000	0.000	0.000

5.6.Cognitive index

We perform Exploratory Factor Analysis to uncover the underlying structure between the tests that were applied to target children. Our underlying assumption is that there is a latent cognitive skill that is imperfectly measured by one of the four sections of the Ages and Stages Questionnaire and the TVIP (Teste de Vocabulário por Imagens Peabody).

Estimation is performed with endline data. The 5 variables used are described in the table below, in their raw version.

Table 10

DESCRIPTIVE STATISTICS OF COGNITIVE VARIABLES

Variable	N	Mean	Std. Dev.	Min	Max
ASQ Motor coordination section	1886	56.214	10.524	0	60
ASQ Communication section	1886	83.579	19.498	0	100
ASQ Precise motor coordination section	1886	60.284	21.037	0	80
ASQ Problem solving section	1886	84.679	25.720	0	120
TVIP Changana - raw score	1897	9.315	7.352	0	50

Notes: Author's calculation from endline survey.

We standardize each variable by subtracting the control mean and dividing by the control standard deviation. For the TVIP variable, we perform this procedure for each month for child age, as in Schady et al (2015).

We search for the cognitive factor by using the method of principal factors, with oblique quartimin rotation. Neither rotation nor the choice of rotation method substantially alters the constructed index, as seen by figure X. We only extract the first factor, as indicated by the Keiser's criteria (Keiser, 1960), Horn's test (Horn, 1965) and by the Scree Test (Cattel, 1966), shown in figure X.

The pairwise Pearson correlations of the standardized variables are shown in table X, while table X shows the factor loadings after quartimin rotation.

Finally, we present the Keiser- Meyer-Olkin (KMO) measure of sampling adequacy in table X. The KMO is a summary of how small are the partial correlations

(after partialling out the influence of the other variables), relative to the original correlations. It takes a value between 0 and 1. Values between 0.7 and 0.79 are historically (Keiser, 1974) considered middling, while values between 0.8 and 0.89 are considered meritorious. Our overall KMO is 0.76.

Table 11

COGNITIVE FACTOR: POLYCHORIC CORRELATION

	ASQ Motor Coordination Section	ASQ Communication Section	ASQ Precise Motor Coordination Section	ASQ Problem Solving Section	TVIP Changana (standardized Within sample)
ASQ motor coordination section	1				
ASQ communication section	0.568	1			
ASQ precise motor coordination section	0.441	0.566	1		
ASQ problem solving section	0.467	0.622	0.729	1	
TVIP (standardized within sample)	0.066	0.162	0.198	0.217	1

Note: Polychoric correlation. Star (*) denote significance at 1%

Table 12

FACTOR LOADINGS AFTER OBLIQUE ROTATION

	Loadings	Uniqueness
ASQ Motor Coordination	0.472	0.575
ASQ Communication	0.655	0.413
ASQ Precise Motor Coordination	0.817	0.372
ASQ Problem Solving	0.848	0.311
Within sample standardized TVIP Changana	0.303	0.917

Notes: Factor loadings based on the exploratory factor analysis with oblique quartimin rotation.

Table 13

COGNITIVE FACTOR - ADEQUACY	
Kaiser-Meyer-Olkin measure of sampling adequacy	
ASQ Motor Coordination	0.814
ASQ Communication	0.801
ASQ Precise Motor Coordination	0.755
ASQ Problem Solving	0.736
Within sample standardized TVIP changana	0.857
Overall	0.773

Figure 10- Screeplot of eigenvalues- Cognitive skills

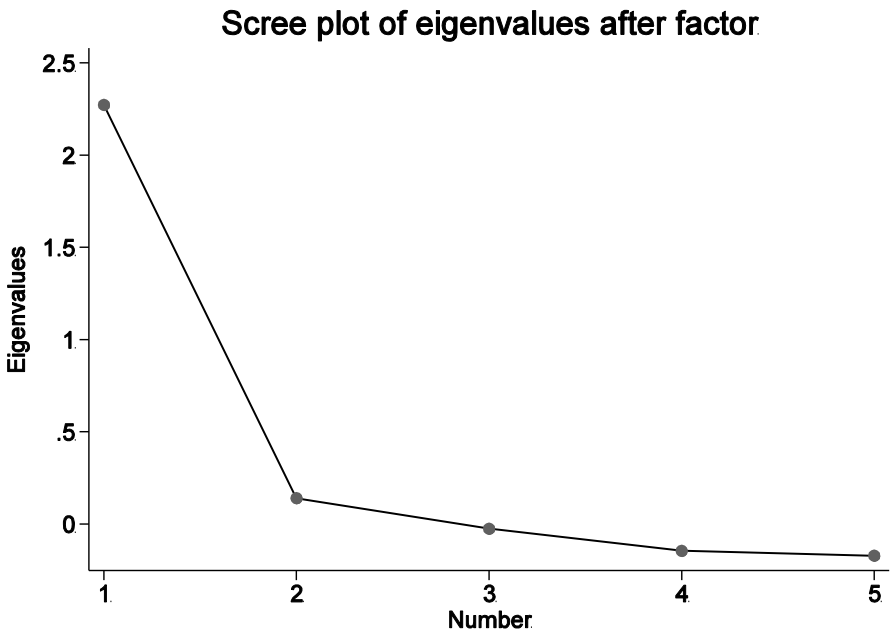
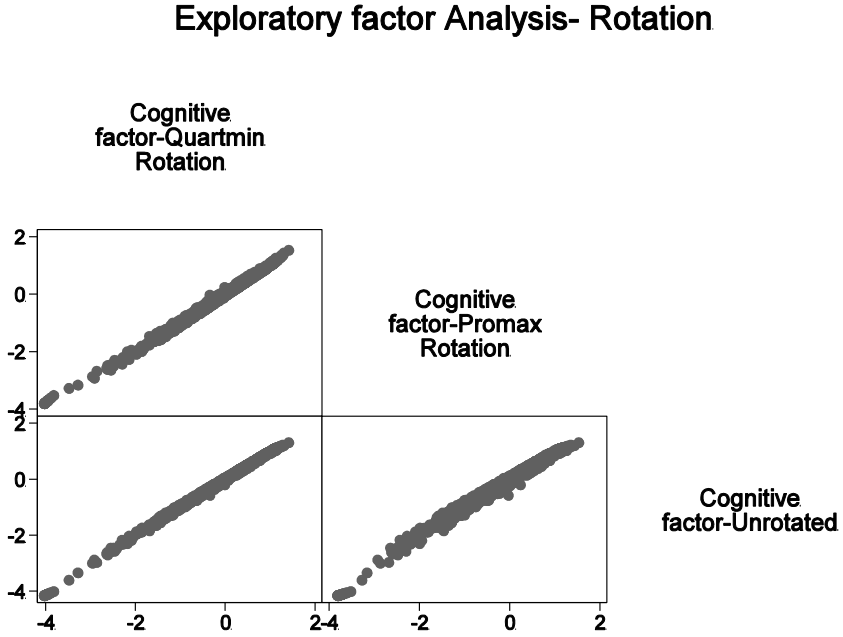


Figure 11- Robustnees to rotation method



5.7. Parenting index (by factor analysis)

Our household questionnaire contains a battery of questions aimed to capture parental investments on children by asking about caregiver practices with the child, and materials children use to play with. [58]All questions refer only to the target child and the caregiver, who is the mother at the most of the cases, but can include grandparents, the father or older siblings.

The first set of questions asks about objects children play with, and all items are binaries. The second set of questions refers to activities involving the caregiver and the target child. Those questions are recorded in a Likert Scale, where one means the caregiver performs the activity many times, and four means the activity is never

[58] See Household Questionnaire, Section "Aprendizagem Pré-Escolar", questions s08 to s26.

performed. The table below presents the summary statistics of the variables included in the index.

We perform Exploratory Factor Analysis searching for the latent factor of parent practices. As we are dealing with binary and categorical data, we first compute the matrix of Tetrachoric and Polychoric correlations. We use the method of principal factors, with oblique quartimin rotation. Next tables present the polychoric correlations between each pair of variables, the factor loadings and the Keiser-Meyer-Olkin measure of sampling adequacy. The overall KMO is 0.81, historically considered as meritorious.

For the subsequent analysis using the predicted index, we standardize it by subtracting the control mean and dividing by the control standard deviation.

Table 14

DESCRIPTIVE STATISTICS- PARENTING PRACTICES

	N	Mean	Std. Dev.	Min	Max
Does your child plays with toys bought in the market?	874	0.452	0.498	0	1
Does your child plays with toys made at home?	878	0.767	0.423	0	1
Does your child plays with toys for rolling?	876	0.638	0.481	0	1
Does your child plays with manipulative objects?	880	0.593	0.491	0	1
Does your child plays with ball?	877	0.693	0.461	0	1
Caregiver-child activities : 1= A Lot (Muito), 2= A few (Pouco), 3= Rarely (Raro) , 4-Never (Nunca)					
Do you read books to your child?	1880	2.641	1.235	1	4
Do you tell stories to your child?	1882	2.617	1.177	1	4
Do you sing songs with your child?	1885	2.259	1.148	1	4
Do you take your kid to the garden/yard?	1882	3.075	1.128	1	4
Do you give names to objects or drawings with your child?	1880	3.268	1.039	1	4
Do you play games of ordering object sizes?	1862	3.308	1.036	1	4
Do you play active games with your child (like throwing a ball, jumping or climbing)?	1882	2.323	1.251	1	4
Do you practice daily routines with your child?	1884	2.254	1.264	1	4
Do you practice self-sufficient activities with your child?	1884	2.132	1.190	1	4
Do you play games that show how things are different?	1871	3.149	1.117	1	4

Notes: Author's calculation from endline survey.

Table 15

TETRACHORIC AND POLYCHORIC CORRELATION OF PARENTING PRACTICES VARIABLES

	Child plays With toys bought in the market	Child plays with toys made at home	Child plays with toys for rolling	Child plays with manipulative Objects	Child plays with ball	Reads Books with Child	Tell stories to your child	Sing songs with your child
Does your child plays with toys bought in the market?	1							
Does your child plays with toys made at home?	0.463	1						
Does your child plays with toys for rolling?	0.279	0.621	1					
Does your child plays with manipulative objects?	0.625	0.614	0.550	1				
Does your child plays with ball?	0.342	0.533	0.562	0.448	1			
Do you read books to your child?	-0.366	-0.240	-0.176	-0.224	-0.180	1		
Do you tell stories to your child?	-0.275	-0.195	-0.136	-0.182	-0.072	0.609	1	
Do you sing songs with your child?	-0.306	-0.295	-0.224	-0.269	-0.216	0.529	0.570	1
Do you take your kid to the garden/yard?	-0.162	-0.163	-0.237	-0.209	-0.281	0.188	0.159	0.345
Do you give names to objects or drawings with your child?	-0.169	-0.207	-0.184	-0.261	-0.239	0.426	0.300	0.291
Do you play games of ordering object sizes?	-0.090	-0.241	-0.132	-0.108	-0.233	0.388	0.264	0.193
Do you play active games with your child?	-0.229	-0.476	-0.417	-0.392	-0.399	0.363	0.198	0.436
Do you practice daily routines with your child?	-0.145	-0.361	-0.295	-0.319	-0.312	0.256	0.162	0.234
Do you practice self-sufficient activities with your child?	-0.152	-0.315	-0.275	-0.314	-0.327	0.118	0.113	0.262
Do you play games that show how things are different?	-0.077	-0.142	-0.120	-0.072	-0.148	0.397	0.322	0.253

TABLE 15 (CONT...)

TETRACHORIC AND POLYCHORIC CORRELATION OF PARENTING PRACTICES VARIABLES

	Take your child to garden/ yard	Give names to objects or drawings	Play Games of ordering object sizes	Play Active Games with child	Practice daily routines with child	Practice self-sufficient activities with child	Play games that show how things are different
Does your child plays with toys bought in the market?							
Does your child plays with toys made at home?							
Does your child plays with toys for rolling?							
Does your child plays with manipulative objects?							
Does your child plays with ball?							
Do you read books to your child?							
Do you tell stories to your child?							
Do you sing songs with your child?							
Do you take your kid to the garden/yard?	1						
Do you give names to objects or drawings with your child?	0.351	1					
Do you play games of ordering object sizes?	0.170	0.564	1				
Do you play active games with your child?	0.308	0.363	0.401	1			
Do you practice daily routines with your child?	0.232	0.289	0.286	0.449	1		
Do you practice self-sufficient activities with your child?	0.229	0.283	0.135	0.394	0.684	1	
Do you play games that show how things are different?	0.233	0.437	0.515	0.288	0.332	0.317	1

Table 16

FACTOR LOADINGS AFTER OBLIQUE ROTATION

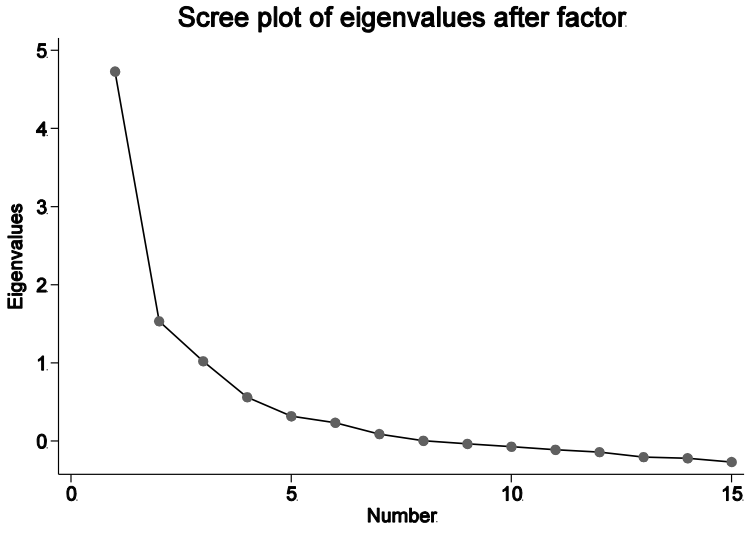
	Loadings	Uniqueness
Does your child plays with toys bought in the market?	-0.032	0.432
Does your child plays with toys made at home?	0.522	0.350
Does your child plays with toys for rolling?	0.782	0.409
Does your child plays with manipulative objects?	0.188	0.317
Does your child plays with ball?	0.604	0.510
Do you read books to your child?	0.033	0.376
Do you tell stories to your child?	-0.023	0.429
Do you sing songs with your child?	-0.040	0.406
Do you take your kid to the garden/yard?	-0.145	0.680
Do you give names to objects or drawings with your child?	0.037	0.463
Do you play games of ordering object sizes?	-0.050	0.404
Do you play active games with your child?	-0.170	0.457
Do you practice daily routines with your child?	0.001	0.377
Do you practice self-sufficient activities with your child?	0.006	0.364
Do you play games that show how things are different?	-0.039	0.543

Notes: Factor loadings based on the exploratory factor analysis with oblique quartimin rotation.

Table 17

PARENTING FACTOR - ADEQUACY	
Kaiser-Meyer-Olkin measure of sampling adequacy	
Does your child plays with toys bought in the market?	0.738
Does your child plays with toys made at home?	0.879
Does your child plays with toys for rolling?	0.820
Does your child plays with manipulative objects?	0.807
Does your child plays with ball?	0.887
Do you read books to your child?	0.804
Do you tell stories to your child?	0.780
Do you sing songs with your child?	0.811
Do you take your kid to the garden/yard?	0.813
Do you give names to objects or drawings with your child?	0.813
Do you play games of ordering object sizes?	0.731
Do you play active games with your child?	0.887
Do you practice daily routines with your child?	0.774
Do you practice self-sufficient activities with your child?	0.708
Do you play games that show how things are different?	0.849
Overall	0.808

Figure 12



5.8. Household assets index (by principal component analysis)

We asked the caregiver about how many of each asset, from a list of items commonly found at the region, are owned by household members. Items include goats, pigs, cows, chicken, trees, furniture, utensils, appliance and tools. We build a household asset index by Principal Component Analysis, as our only purpose is to reduce the dimensionality of several items about ownership of assets. As our purpose is to provide a control in the regressions, we build the index only for the baseline. The next tables provide the descriptive statistics of variables included in the index, the KMO measure of sampling adequacy, and the averages of items owned for each quintile of the index distribution.

Table 18

DESCRIPTIVE STATISTICS- HOUSEHOLD ASSETS

	N	Mean	Std. Dev.	Min	Max
Hoes	13581	3.111	2.353	0	32
Axes	13581	0.634	1.424	0	31
Cashew trees	13581	8.378	24.655	0	500
Coconut trees	13581	2.328	6.140	0	90
Goats	13581	1.365	2.562	0	62
Pigs	13581	0.876	1.779	0	23
Cows	13581	1.131	4.060	0	82
Chicken/Ducks	13581	6.027	7.399	0	60
Mobile phones	13581	0.126	0.757	0	42
Bicycles	13581	0.404	1.063	0	25
Sound system	13581	0.444	0.922	0	15
Tables	13581	1.434	2.294	0	33
Chairs	13581	5.752	5.365	0	56
Radio	13581	0.736	1.206	0	18
Clock	13581	0.473	1.186	0	28
Fishing net	13581	0.107	0.828	0	18
Canoes	13581	0.099	0.513	0	10
Beds	13581	1.411	1.556	0	22
Stoves (gas or electric)	13581	0.118	0.440	0	6
Stove (wood)	13581	0.540	0.879	0	12
Fridge	13581	0.162	0.526	0	11
TV	13581	0.261	0.654	0	12
Motorbike	13581	0.043	0.329	0	11
Solar panel	13581	0.199	0.498	0	5

Notes: Author's calculation from baseline data

Table 19

TABLE 19
HOUSEHOLD ASSETS INDEX
KAISER-MEYER-OLKIN MEASURE OF SAMPLING ADEQUACY

Hoes	0.796
Axes	0.863
Cashew trees	0.876
Coconut trees	0.841
Goats	0.810
Pigs	0.814
Cows	0.889
Chicken/Ducks	0.862
Mobile phones	0.678
Bicycles	0.884
Sound system	0.898
Tables	0.780
Chairs	0.850
Radio	0.863
Clock	0.819
Fishing net	0.796
Canoes	0.698
Beds	0.856
Stoves (gas or electric)	0.584
Stove (wood)	0.823
Fridge	0.807
TV	0.821
Motorbike	0.674
Solar panel	0.902
Overal	0.806

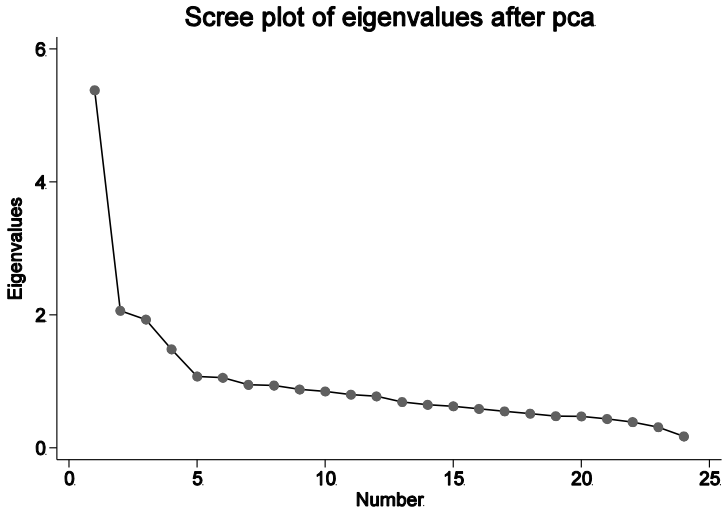
Table 20

ASSET INDEX

	Quintile of asset index distribution				
Asset	1st (poorest)	2nd	3rd	4th	5th (richest)
Hoes	1.824	2.357	2.953	3.529	4.889
Axes	0.228	0.431	0.619	0.738	1.152
Cashew trees	4.412	6.602	8.994	8.668	13.204
Coconut trees	0.716	1.358	1.928	2.814	4.822
Goats	0.188	0.556	1.106	1.922	3.052
Pigs	0.191	0.420	0.728	1.105	1.936
Cows	0.076	0.189	0.519	1.200	3.668
Chicken/Ducks	1.815	4.332	5.893	7.017	11.070
Mobile phones	0.000	0.013	0.037	0.104	0.474
Bicycles	0.025	0.109	0.256	0.476	1.152
Sound system	0.010	0.125	0.315	0.518	1.250
Tables	0.334	0.879	1.121	1.596	3.239
Chairs	2.003	3.959	5.228	7.094	10.467
Radio	0.121	0.509	0.651	0.881	1.516
Clock	0.050	0.201	0.375	0.525	1.214
Fishing net	0.008	0.012	0.075	0.090	0.349
Canoes	0.004	0.024	0.051	0.077	0.341
Beds	0.332	0.798	1.237	1.763	2.923
Stoves (gas or electric)	0.000	0.013	0.037	0.104	0.434
Stove (wood)	0.237	0.465	0.479	0.658	0.863
Fridge	0.003	0.006	0.046	0.173	0.584
TV	0.006	0.029	0.102	0.304	0.862
Motorbike	0.000	0.000	0.018	0.050	0.148
Solar panel	0.019	0.036	0.152	0.288	0.497

Notes: Author's calculation from baseline data

Figure 13- Screeplot-asset index



5.9. Program impacts and alternative pathways

Table 21

ROBUSTNESS TO INCLUSION OF CONTROLS

Dep var:	Total ASQ Score		Ever been to preschool		Total ASQ Score	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Community	0.214***	0.184***				
	(0.045)	(0.043)				
Treatment Community			0.501***	0.496***		
			(0.033)	(0.033)		
Ever been to preschool					0.427***	0.370***
					(0.102)	(0.096)
Controls		X		X		X
Observations	1,831	1,831	1,839	1,839	1,831	1,831
R-squared	0.058	0.213	0.271	0.286	0.033	0.195
Control Mean:	0.000	0.000	0.089	0.089	0.000	0.000
Control Standard Deviation:	1.000	1.000	0.285	0.285	1.000	1.000
Control Complier Mean:					-0.315	-0.257

Notes: This table reports estimates of the effects of the provision of preschool centers at community and, the effect of the provision of preschool centers on preschool enrollment and the estimates of preschool attendance. Only the target children were tested. The first line reports the estimates of an OLS regression of total score from Ages and Stages Questionnaire on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance. Preschool attendance is instrumented by the community treatment status. Total ASQ score standardized by subtracting the control mean and dividing by control standard deviation. Control complier mean calculated as in Kling et al (2001). All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in months, sex, height for age at baseline, weight for age at baseline, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, stunted at baseline, child with risks of communication deficits at baseline, child with risks of motor coordination deficits at baseline, child with risks of precise motor coordination at baseline, child with threat of problem resolution deficits at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size. See online appendix for alternative specifications and alternative samples.

Table 22

IMPACT ON SUMMARY INDECES

Dep var:	Cognitive Factor (1)	Schooling index (2)	Parenting index (3)	Health index (4)
OLS: Treatment community	0.171***	0.136***	0.119**	0.064
	(0.050)	(0.044)	(0.052)	(0.062)
IV: Ever been to preschool	0.337***	0.360***	0.232**	0.126
	(0.107)	(0.117)	(0.104)	(0.125)
Observations	1,686	2,891	1,630	1,697
Control Mean:	0.075	-0.039	-0.011	-0.002
Control Standard Deviation:	0.956	1.010	0.994	1.017
Control Complier Mean:	-0.183	-0.402	-0.141	-0.099

Notes: This table reports estimates of the effects of the provision of preschool centers at community and the estimates of the effects of preschool attendance. For columns (1), (3) and (4), sample includes only target children. For column 2, sample are all children aged 5 to 9 at endline survey. The first line reports the estimates of an OLS regression of each outcome on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance. Preschool attendance is instrumented by the community treatment status. See on line appendix for the construction of each index. .All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Control complier mean calculated as in Kling et al (2001). Estimates weighted by community population size. Standard errors clustered at community level. Controls at column (2) include child age in years, sex, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size. For columns (1), (3) and (4), child age in months, height for age at baseline, weight for age at baseline, dummy for being under median of asset index at baseline, stunted at baseline, child with risks of communication deficits at baseline, child with risks of motor coordination deficits at baseline, child with risks of precise motor coordination at baseline and dummy for child with risks of problem resolution deficits at baseline are included in controls, in addition to controls from column (2). See online appendix for alternative specifications and alternative samples.

Table 23

LABOR SUPPLY OF OLDER HOUSEHOLD MEMBERS

VARIABLES	Member older than 15 worked in the last 30 days	Caregiver worked in the last 30 days	Caregiver (male) worked in the last 30 days	Caregiver (female) worked in the last 30 days	Father worked in the last 30 days	Mother worked in the last 30 days
	(1)	(2)	(3)	(4)	(5)	(6)
OLS: Treatment community	0.010	0.037*	0.018	0.024	0.028	0.015
	(0.015)	(0.021)	(0.059)	(0.025)	(0.029)	(0.020)
IV: A household member went to PS	0.018	0.071*	0.044	0.045	0.051	0.028
	(0.029)	(0.040)	(0.151)	(0.046)	(0.055)	(0.037)
Observations	5,678	1,726	230	1,496	1,114	1,542
Control Mean:	0.287	0.240	0.550	0.190	0.574	0.198
Control Standard Deviation:	0.452	0.428	0.500	0.393	0.495	0.399
Control Complier Mean:	0.288	0.184	0.469	0.172	0.571	0.181

Notes: This table reports estimates of the effects of the provision of preschool centers at community and the estimates of spillover effects of preschool attendance of at least one household member. Work in the last 30 days includes any paid work, thus excluding self-sufficient cropping activities. First column includes all household members older than 15. Second column includes only household members identified as the primary caregiver of the target child. Column (3) and (4) splits the sample between male and female caregivers. Sample from column (5) is comprised by fathers of target children and (6) by mothers of target children. First line reports the estimates of an OLS regression of the probability of working on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of having had a household member who went to preschool. The endogenous variable is a dummy that is equal to one if any younger household member has been to preschool. Instrument is the community treatment status. All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Control complier mean calculated as in Kling et al (2001). Estimates weighted by community population size. Standard errors clustered at community level. For all regressions, controls are: years of education (of father, mother, caregiver of member older than 15), age in years, dummy for household under median of asset index at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size and flags for missing data. Columns (1) includes sex, parents of target child speak Portuguese at baseline as controls. Column (2) includes sex as control. See online appendix for alternative specifications and alternative samples.

Table 24

LABOR SUPPLY AND CHILD DEVELOPMENT

VARIABLES	Cognitive factor	Cognitive factor	Ever been to primary school	Ever been to primary school
Communities	Only control	All	Only control	All
	(1)	(2)	(3)	(4)
Caregiver worked in the last 30 days	0.183**	0.039	0.017	0.016
	(0.071)	(0.055)	(0.034)	(0.034)
Treatment community		0.209***		0.057**
		(0.046)		(0.025)
Observations	786	1,712	1,485	1,485

Notes: This table reports estimates of caregiver's labor supply on child cognitive factor and child preschool attendance. In columns 1 and 3, only control communities are included, while columns 2 and 4 include all communities. All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in years, sex, whether parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size, child age in months, height for age at baseline, weight for age at baseline, dummy for being under median of asset index at baseline, stunted at baseline, child with risks of communication deficits at baseline, child with risks of motor coordination deficits at baseline, child with risks of precise motor coordination at baseline and dummy for child with risks of problem resolution deficits at baseline are included in controls.

Table 25

ALTERNATIVE PATHWAYS: CLEANNESS AND HANDWASHING

Dep var:	Child urinated During Interview Urinated	Child's face was clean during interview	Child's hair was clean during interview	Child's hands were clean during interview	Caregiver thinks that washing hands is important	Child washed hands after Necessities Yesterday	Child washed hands after Meal Yesterday	Child washed hands before meal yesterday	Caregiver Knows when Washing hands is important
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
OLS: Treatment community	-0.036	-0.027	-0.010	0.014	-0.022	-0.008	-0.000	-0.005	-0.022
	(0.034)	(0.036)	(0.035)	(0.024)	(0.014)	(0.016)	(0.025)	(0.017)	(0.026)
IV:Ever been to preschool	-0.072	-0.054	-0.020	0.029	-0.043	-0.017	-0.001	-0.011	-0.044
	(0.067)	(0.072)	(0.071)	(0.048)	(0.029)	(0.034)	(0.051)	(0.034)	(0.053)
Observations	1,839	1,839	1,838	1,839	1,839	1,699	1,731	1,768	1,838
Control Mean:	0.365	0.481	0.457	0.375	0.933	0.876	0.859	0.914	0.343
Control Standard Deviation:	0.482	0.500	0.498	0.484	0.250	0.329	0.348	0.280	0.475
Control Complier Mean:	0.508	0.545	0.490	0.329	0.985	0.925	0.869	0.969	0.407

Notes: This table reports estimates of the effects of the provision of preschool centers on child cleanness and handwashing and the effect of preschool enrollment on cleanness and handwashing. All variables refer only to the target child. Columns (1) to (4) present perceptions from interviewer about child cleanness, filled by the end of the interview. Dependent variables from columns (5) to (9) are perceptions and practices reported by the caregiver about handwashing. The first line reports the estimates of an OLS regression of each dependent variable on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance. Preschool attendance is instrumented by the community treatment status. All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in months, sex, height for age at baseline, weight for age at baseline, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, stunted at baseline, child with risks of communication deficits at baseline, child with risks of motor coordination deficits at baseline, child with risks of precise motor coordination at baseline, child with threat of problem resolution deficits at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size. See online appendix for alternative specifications and alternative samples.

Table 26

ALTERNATIVE PATHWAYS: CHILD VACCINATION

Dep var:	Child received vitamin A supplement in last 6 months? (1)	Child was dewormed in last 12 months (2)	Child vaccinated: BCG vaccine (3)	Child vaccinated: DPT vaccine (4)	Child vaccinated: Measles Vaccine (5)	Child vaccinated: yellow fever vaccine (6)
OLS: Treatment community	-0.034	0.041	0.013	0.024	0.010	0.006
	(0.029)	(0.026)	(0.014)	(0.020)	(0.024)	(0.018)
IV:Ever been to preschool	-0.068	0.083	0.025	0.045	0.018	0.011
	(0.060)	(0.051)	(0.028)	(0.038)	(0.045)	(0.034)
Observations	1,745	1,753	967	901	875	856
Control Mean:	0.367	0.266	0.957	0.927	0.882	0.916
Control Standard Deviation:	0.482	0.442	0.203	0.260	0.322	0.278
Control Complier Mean:	0.517	0.209	0.941	0.872	0.859	0.901

Notes: This table reports estimates of the effects of the provision of preschool centers on child vaccination and deworming, and the effect of preschool enrollment on vaccination and deworming. All variables refer only to the target child. The first line reports the estimates of an OLS regression of each dependent variable on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance. Preschool attendance is instrumented by the community treatment status. All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in months, sex, height for age at baseline, weight for age at baseline, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, stunted at baseline, child with risks of communication deficits at baseline, child with risks of motor coordination deficits at baseline, child with risks of precise motor coordination at baseline, child with threat of problem resolution deficits at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size. See online appendix for alternative specifications and alternative samples

Table 27

CHILD ANTHROPOMETRICS

Dep var:	Weight for age z score (1)	Child is wasted (2)	Weight for age z score (3)	Child is stunted (4)
OLS: Treatment community	0.031	0.005	0.036	0.015
	(0.045)	(0.014)	(0.076)	(0.031)
IV: Ever been to preschool	0.063	0.009	0.073	0.029
	(0.091)	(0.029)	(0.158)	(0.063)
Observations	1,803	1,803	1,729	1,729
Control Mean:	-0.734	0.091	-1.500	0.324
Control Standard Deviation:	0.990	0.287	1.133	0.468
Control Complier Mean:	-0.805	0.085	-1.547	0.293

Notes: This table reports estimates of the effects of the provision of preschool centers and the effects of preschool attendance on child anthropometric measures. All variables refer only to the target child. Z-Scores calculated according to new Child Growth Standards from the World Health Organization, using Stata WHO 2007Anthro Package. All scores specific for child age in months and child sex. Wasting is defined as weight for age below -2 standard deviations from the reference. Stunting is defined as height for age below -2 standard deviations from reference. Implausible scores below -4 std, as well as scores above 4 std were recoded to missing. The first line reports the estimates of an OLS regression of each dependent variable on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance. Preschool attendance is instrumented by the community treatment status. All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in months, sex, height for age at baseline, weight for age at baseline, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, stunted at baseline, child with risks of communication deficits at baseline, child with risks of motor coordination deficits at baseline, child with risks of precise motor coordination at baseline, child with threat of problem resolution deficits at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size. See online appendix for alternative specifications and alternative samples

Table 28

YOUNGER SIBLING ANTHROPOMETRICS

Dep var:	Weight for age z score (1)	Child is wasted (2)	Weight for age z score (3)	Child is stunted (4)
OLS: Treatment community	-0.029	0.045	0.024	-0.187
	(0.029)	(0.085)	(0.045)	(0.117)
IV:Ever been to preschool	-0.051	0.078	0.042	-0.330
	(0.051)	(0.147)	(0.081)	(0.210)
Observations	526	526	453	453
Control Mean:	0.110	-0.641	0.395	-1.499
Control Standard Deviation:	0.314	1.092	0.490	1.367
Control Complier Mean:	0.153	-0.725	0.336	-1.247

Notes: This table reports estimates of the effects of the provision of preschool centers and the effects of preschool attendance on child anthropometric measures. Sample is comprised by the next (in birth order) younger sibling of target children who is older than 36 moths. Z-Scores calculated according to new Child Growth Standards (2006) from the World Health Organization, using Stata WHO 2007Anthro Package. All scores specific for child age in months and child sex. Wasting is defined as weight for age below -2 standard deviations from the reference. Stunting is defined as height for age below -2 standard deviations from reference. Implausible scores below -4 std, as well as scores above 4 std were recoded to missing. The first line reports the estimates of an OLS regression of each dependent variable on the dummy that indicates the treatment status of the community. The second line reports IV estimates of the effect of preschool attendance. Preschool attendance is instrumented by the community treatment status. All regressions include dummies of randomization blocks, local district and local administrative post, as well the presence of other than Save the Children preschools at the community. Estimates weighted by community population size. Standard errors clustered at community level. Controls include child age in months, sex, height for age at baseline, weight for age at baseline, parents speak Portuguese at baseline, mother dead at baseline, father dead at baseline, mother's education, father's education, mother's age, father's age, dummy for being under median of asset index at baseline, orphan at baseline, stunted at baseline, child with risks of communication deficits at baseline, child with risks of motor coordination deficits at baseline, child with risks of precise motor coordination at baseline, child with threat of problem resolution deficits at baseline, number of male household members under 1, 2, 3, 4 and 5 years old, number of female household members under 1, 2, 3, 4 and 5 years old, household age equivalent size. See online appendix for alternative specifications and alternative samples

5.10.Early development index

Table 29

EARLY DEVELOPMENT INDEX ITEMS

Item (1)	OLS: Treatment community Estimate (2)	Std (3)	Coding (4)	Obs (5)	Control Mean (6)	Control Std (7)
A2. Over- or underdressed for school-related activities	0.076	(0.059)	A*	917	0.193	0.395
A3. Too tired/sick to do school work	0.032	(0.056)	A*	911	0.145	0.352
A4. Late	0.095*	(0.055)	A*	914	0.403	0.491
A5. Hungry	-0.186***	(0.057)	A*	810	0.213	0.410
A6. Is independent in washroom habits most of the time	-0.031	(0.055)	A*	905	0.743	0.437
A7. Shows an established hand preference (right vs. left or vice versa)	0.025	(0.033)	A*	915	0.909	0.288
A8. Is well coordinated	0.046	(0.048)	B*	916	0.880	0.325
A9. Proficiency at holding a pen, crayons, or a brush	-0.093	(0.069)	B*	919	1.479	0.588
A10. Ability to manipulate objects	-0.238**	(0.101)	B*	917	1.585	0.633
A11. Level of energy throughout the school day	-0.194**	(0.092)	B*	919	1.659	0.682
A12. Overall physical development	-0.060	(0.113)	B*	918	1.441	0.599
B1. Ability to use language effectively in Portuguese	-0.079	(0.123)	B*	918	2.234	0.673
B2. Ability to listen in Portuguese	-0.128	(0.092)	B*	918	2.095	0.669
B3. Ability to tell a story	-0.025	(0.117)	B*	884	1.748	0.715
B4. Ability to take part in imaginative play	-0.321***	(0.115)	B*	886	2.006	0.690
B5. Ability to communicate own needs in a way understandable to	-0.062	(0.106)	B*	913	1.563	0.619
B6. Ability to understand on first try what is being said to him/her	-0.132	(0.101)	B*	913	1.723	0.696
B7. Ability to articulate clearly, without sound substitutions	-0.0018	(0.115)	B*	913	1.637	0.685
B8. Knows how to handle a book	0.022	(0.051)	A*	917	0.880	0.325
B9. Is generally interested in books	0.130*	(0.071)	A*	904	0.793	0.405
B10. Is interested in reading	0.176*	(0.101)	A*	899	0.587	0.493
B11. Is able to identify at least 10 letters of the alphabet	0.123**	(0.059)	A*	909	0.522	0.500
B14. Is able to participate in group reading activities	0.225***	(0.064)	A*	901	0.620	0.486
B15. Is able to read simple words	-0.033	(0.067)	A*	911	0.593	0.492
B16. Is able to read complex words	-0.010	(0.039)	A*	900	0.149	0.357
B17. Is able to read simple sentences	0.072	(0.094)	A*	909	0.434	0.496

(Cont….) TABLE 29
EARLY DEVELOPMENT INDEX ITEMS

Item	OLS: Treatment community Estimate	Std	Coding	Obs	Control Mean	Control Std
(1)	(2)	(3)	(4)	(5)	(6)	(7)
B19. is aware of writing directions in English	-0.001	(0.035)	A*	915	0.876	0.330
B20. is interested in writing voluntarily	0.141	(0.090)	A*	904	0.659	0.475
B21. is able to write his/her own name in English	-0.006	(0.049)	A*	911	0.380	0.486
B22. is able to write simple words	-0.114**	(0.051)	A*	907	0.531	0.500
B23. is able to write simple sentences	-0.098	(0.072)	A*	905	0.356	0.479
B24. is able to remember things easily	0.102	(0.068)	A*	888	0.550	0.498
B25. is interested in mathematics	0.159***	(0.058)	A*	889	0.669	0.471
B26. is interested in games involving numbers	0.303***	(0.048)	A*	838	0.592	0.492
B27. is able to sort and classify objects by a common characteristic	0.275***	(0.080)	A*	883	0.611	0.488
B28. is able to use one-to-one correspondence	0.239***	(0.079)	A*	906	0.619	0.486
B29. is able to count to 20	0.208***	(0.048)	A*	914	0.568	0.496
B30. is able to recognize numbers 1 – 10	0.115***	(0.038)	A*	911	0.705	0.457
B31. is able to say which number is bigger of the two	0.364***	(0.092)	A*	901	0.595	0.491
B32. is able to recognize geometric shapes (e.g., triangle, circle, square)	0.241***	(0.088)	A*	900	0.355	0.479
B33. understands simple time concepts (e.g., today, summer, bedtime)	0.151***	(0.047)	A*	903	0.727	0.446
B34. demonstrates special numeracy skills or talents	0.118	(0.102)	A*	824	0.419	0.494
B35. demonstrates special literacy skills or talents	0.012	(0.072)	A*	857	0.403	0.491
B36. demonstrates special skills or talents in arts	-0.043	(0.083)	A*	770	0.377	0.485
B37. demonstrates special skills or talents in music	-0.124	(0.110)	A*	800	0.516	0.500
B38. demonstrates special skills or talents in athletics/dance	0.062	(0.087)	A*	841	0.509	0.500
B40. demonstrates special skills or talents in other areas	0.043	(0.054)	A*	663	0.0634	0.244
C1. overall social/emotional development	-0.275***	(0.091)	B*	912	1.694	0.563
C2. ability to get along with peers	-0.227***	(0.077)	B*	917	1.538	0.558
C3. plays and works cooperatively with other children at the level	-0.050	(0.071)	C*	917	1.337	0.503
C4. is able to play with various children	-0.085	(0.128)	C*	911	1.424	0.554
C5. follows rules and instructions	0.055	(0.109)	C*	893	1.594	0.603

(Cont….) TABLE 29
EARLY DEVELOPMENT INDEX ITEMS

Item (1)	OLS: Treatment community Estimate (2)	Std (3)	Coding (4)	Obs (5)	Control Mean (6)	Control Std (7)
C6. respects the property of others	-0.099	(0.104)	C*	899	1.447	0.579
C7. demonstrates self-control	0.010	(0.134)	C*	906	1.578	0.628
C8. shows self-confidence	-0.039	(0.164)	C*	892	1.640	0.654
C9. demonstrates respect for adults	-0.140	(0.103)	C*	911	1.363	0.554
C10. demonstrates respect for other children	-0.134	(0.087)	C*	918	1.376	0.541
C11. accepts responsibility for actions	-0.168	(0.128)	C*	903	1.709	0.691
C12. listens attentively	-0.035	(0.093)	C*	918	1.448	0.596
C13. follows directions	-0.005	(0.085)	C*	918	1.591	0.622
C14. completes work on time	-0.201	(0.122)	C*	919	1.789	0.688
C15. works independently	0.053	(0.107)	C*	916	1.572	0.651
C16. takes care of school materials	-0.054	(0.104)	C*	918	1.556	0.643
C17. works neatly and carefully	0.056	(0.113)	C*	911	1.668	0.668
C18. is curious about the world	-0.056	(0.155)	C*	898	1.916	0.726
C19. is eager to play with a new toy	-0.374**	(0.174)	C*	863	1.747	0.718
C20. is eager to play a new game	-0.512***	(0.126)	C*	843	1.658	0.762
C21. is eager to play with/read a new book	-0.542***	(0.186)	C*	847	1.953	0.802
C22. is able to solve day-to-day problems by him/herself	-0.291	(0.189)	C*	825	2.009	0.781
C24. is able to follow class routines without reminders	-0.102	(0.0941)	C*	905	1.906	0.726
C25. is able to adjust to changes in routines	-0.177	(0.130)	C*	867	1.686	0.664
C27. shows tolerance to someone who made a mistake	-0.020	(0.152)	C*	901	1.731	0.665
C28. will try to help someone who has been hurt	-0.283	(0.177)	C*	885	1.831	0.753
C29. volunteers to help clear up a mess someone else has made	-0.278	(0.224)	C*	878	2.030	0.785
C30. if there is a quarrel or dispute will try to stop it	-0.305	(0.212)	C*	876	2.054	0.775
C31. offers to help other children who have difficulty with a task	-0.326*	(0.176)	C*	890	1.930	0.783
C32. comforts a child who is crying or upset	-0.358*	(0.189)	C*	889	1.970	0.778
C34. will invite bystanders to join in a game	-0.314	(0.228)	C*	853	2.087	0.794

(Cont....) TABLE 29
EARLY DEVELOPMENT INDEX ITEMS

Item	OLS: Treatment community Estimate	Std	Coding	Obs	Control Mean	Control Std
(1)	(2)	(3)	(4)	(5)	(6)	(7)
C35. helps other children who are feeling sick	-0.223	(0.184)	C*	877	1.908	0.780
C37. gets into physical fights	-0.310**	(0.127)	C*	890	2.548	0.705
C38. bullies or is mean to others	-0.168***	(0.048)	C*	890	2.701	0.536
C39. kicks, bites, hits other children or adults	-0.064*	(0.035)	C*	897	2.710	0.551
C42. can't sit still, is restless	-0.293***	(0.089)	C*	909	2.560	0.627
C43. is distractible, has trouble sticking to any activity	-0.208**	(0.087)	C*	906	2.301	0.688
C45. is disobedient	-0.209***	(0.056)	C*	903	2.573	0.629
C46. has temper tantrums	-0.109*	(0.055)	C*	854	2.637	0.572
C47. is impulsive, acts without thinking	-0.213**	(0.090)	C*	889	2.556	0.612
C48. has difficulty awaiting turn in games or groups	-0.285*	(0.159)	C*	910	2.439	0.665
C49. cannot settle to anything for more than a few moments	-0.049	(0.159)	C*	908	2.301	0.664
C50. is inattentive	-0.087	(0.197)	C*	915	2.178	0.692
C51. seems to be unhappy, sad, or depressed	-0.121	(0.104)	C*	905	2.485	0.655
C52. appears fearful or anxious	-0.082	(0.121)	C*	909	2.468	0.681
C53. appears worried	-0.061	(0.105)	C*	908	2.469	0.639
C54. cries a lot	-0.015	(0.072)	C*	909	2.679	0.583
C55. is nervous, high-strung, or tense	-0.059	(0.076)	C*	902	2.620	0.609
C56. is incapable of making decisions	-0.351***	(0.103)	C*	834	2.455	0.682
C57. is shy	-0.456***	(0.113)	C*	898	2.557	0.622
C58. sucks a thumb/finger	-0.074	(0.054)	C*	910	2.762	0.543

Notes: : Coding : A*: 0-No 1-Yes; B*: B *; C*: 1-Regularly 2-Sometimes 3-Never. This table reports estimates of the effects of the provision of preschool centers at development domains of first graders, as measured by the Early Development Index. Sample consists of first graders of primary schools, randomly chosen from the list of first graders from each primary school operating at the sampling area. Column 1 shows each item exactly as in EDI instrument. Column 2 presents the OLS estimates for the dummy that indicates that the community where the primary school is located received a preschool. Column 4 shows the coding as in the Edi instrument. All regressions include dummies of randomization blocks, local district and local administrative post. Standard errors clustered at class level. Controls include child age in years, sex, time elapsed since the start of school year and the date of the interview, flag for date of start of classes not reported, date of interview, number of students at class, teacher's sex, teacher's highest grade completed, flag for highest grade not reported, teacher's subjective familiarity with students.

Table 30

EARLY DEVELOPMENT INDEX- BY DOMAINS

Dep var:	Physical Health and Well-being (1)	Communication and General Knowledge (2)	Cognitive Development and Language (3)	Social Competence (4)	Emotional Maturity (5)
OLS: Treatment community	0.0629	-0.0454	0.208*	0.0287	0.0611
	(0.133)	(0.147)	(0.123)	(0.189)	(0.153)
Observations	919	919	919	919	919
Control Mean:	-0.054	-0.030	-0.094	-0.070	-0.057
Control Standard Deviation:	0.986	1,013	1,051	1,031	0.890

Notes: This table reports estimates of the effects of the provision of preschool centers at development domains of first graders, as measured by the Early Development Index. Sample consists of first graders of primary schools, randomly chosen from the list of first graders from each primary school operating at the sampling are. The first line reports the estimates of an OLS regression of each development domain on the dummy that indicates that a preschool was built at the community where the primary school is located. All regressions include dummies of randomization blocks, local district and local administrative post. Standard errors clustered at class level. No additional controls are included.

5.11.Robustness to alternative specifications

In our main specification, we reclassify 6 control communities as treatment. Those communities were originally coded by the Mozambican National Institute of Statistics (INE), as different and separate communities. Those communities, however, are neighborhoods or larger communities that have been treated by Save the Children. This fact was not known at the moment of randomization, although some of those communities share the same name and are differentiated by numbers or letters, as Muzingane B/1 to B/5 or Chitsembe A and B. This fact was only noticed after the randomization took place, when data was collected in the field and later when a high number of kids from control communities were reported as having been enrolled into preschool. The plot of the GPS locations of households from those communities confirmed our suspicion.

Our second issue is the existence of other preschools in control areas that were not built by Save the Children. We managed to identify 6 preschools in control communities, which were built and operated wither by churches or other NGOs. This section presents a series of alternative specification that deal both with the question of contamination and reclassification of control communities. Our main goal here is to better characterize the effect of Save the Children' preschool model on the enrolled child.

In our main specification, we included 69 communities, out of which 6 were reclassified from a control status to treatment. Other 6 communities had the presence of a preschool that was built from some organization other than Save the Children. Our main endogenous variable is a dummy indicating that the child has ever been enrolled into preschool. We do not differentiate who runs the preschool, so the dummy is also equal to one for children who went to a Save the Children preschool or to a child that went to a preschool run by the church. In particular, we estimate a two stage least squares model:

$$Y_{ijt} = \eta + \gamma_1 \hat{D}_{ijt} + \sum_{n=1}^{N} \gamma_n X_{nit-1} + \xi_{it} \qquad (1)$$

$$D_{ijt} = \alpha + \theta_1 T_j + \sum_{n=1}^{N} \theta_n X_{nit-1} + \varepsilon_{it} \qquad (2)$$

Where Y_{ijt} is our outcome of interest and D_{ijt} is dummy equal to one if a child i, from community j, attended preschool. T_j is a dummy that indicates that the community received a Save the Children preschool. In the main specification, T_j =1 for the control merged communities that were merged. X_{nit-1} is a vector of predetermined individual and household characteristics as taken from baseline data, and also include geographical dummies, randomization block dummies and dummies indicating that a church of other NGO built a preschool inside the community. In case merged communities belong to two different original randomization blocks, we merged the randomization block into a single one.

Column (1) presents the estimates of our main specification. We interpret it as the effect for the child of going to ANY preschool, induced by the construction of Save the Children preschools, controlled for the effect of preschools run by other institution. All those estimates are contained in the main tables of the paper.

In column 2, we deal with the problem of contamination caused by those preschools run by other institutions. In order to preserve the randomization structure, we drop all randomization blocks that contain a preschool run by an organization other than Save the Children. We also exclude the randomization triplet containing Machalucuane. In this community, Save the Children built a preschool in an area of very difficult access and no children from Machalucuane were enrolled, resulting in a treatment community with zero take up. This exercise ended up excluding eighteen communities, leaving 51 communities in total. Relative to the main specification, the only difference is the exclusion of those 18 communities.

The proportion of children from control communities ever enrolled into preschool falls to 4.2 pp, while the proportion of children enrolled in treatment communities increase to 57pp. While we still cannot formally claim these are the estimates of the treatment on the treated, this specification alleviates the problem of contamination. The impact on communication section of the ASQ becomes significant at 10%. The estimates for the impact on receptive language as measured by the TVIP are stronger, and we estimate of almost two extra words recognized by the child on the TVIP, significant at 5%. Additionally, we observe a significant reduction of child labor. The time children spend working or accompanying the

mother in the field diminishes by 1.662 hours per week. The time spent on community meetings also falls by 1 hour per week.

Next, we take a different approach and reclassify the endogenous variable, using all the communities as in our main specification. However, instead of taking the value of one if a child has ever been enrolled in ANY preschool, the participation dummy only turns on when the child has been to a preschool AND the child lives in a treatment community. As we did not identify any preschool run by other institution but save the Children in treatment areas, we are pretty confident that those children were enrolled in a preschool run by Save the Children. Mechanically, this will yield estimates that are smaller than those from the first specification. By doing so, we are intuitively estimating a lower bound for the LATE effect of going to a Save the Children preschool, instead as estimating the effect of going to ANY preschool (controlled for the effect of non- Save the Children preschools) as we did in the main specification. As expected, the results are slightly smaller when compared to the ones from column 1. Nonetheless, qualitatively the results are robust and the conclusions about the impact of the program remain unchanged.

Another way of dealing with the problem of the communities that were originally assigned as controls but that were found to be mere neighborhoods of treatment communities is to replace the binary definition of treatment and control and use a continuous definition given by the distance from the household to the closest preschool from Save the Children. GPS location from households and from each Save the Children classroom (For some communities, classrooms are spread over distant places) allow us to calculate this distance. In this case, we can be completely agnostic about which communities were treated or not. The "intensity of treatment" will be given by the distance to the closest Save the Children preschool. If a household is located in a (originally assigned) "control" community that is a neighborhood of a treated community, the distance to the Save the Children preschool will be similar to the distance from households located inside the treatment community. It is still also possible that children commute from other control communities to attend Save the Children preschools even if the communities are physically separated but located not too far from each other. By using the distance to the closest Save the Children school we can characterize the situation from those children in a much finer grain than using a binary instrument.

Column 4 shows the IV results of this exercise using the original participation dummy (=1 if the child has been to any preschool) but replacing the binary instrument by the distance from the household to the closest Save the Children preschool. Although not reported in the table for lack of space[59], the number of observations from the regression of the ASQ total score falls from 1842 to 1440, due to missing GPS data. Despite that, results are robust not only qualitatively, but also (surprisingly) quantitatively in respect to our main specification.

In column (5), we deal again with the robustness of our results to the reclassification of schools. Instead of reclassifying communities, we use the original community assignment. The results are very similar to the ones from our main specification. The estimate of the impact on the TVIP, when rescaled in standard deviations from the control average for each child age in months, cannot longer be statistically distinguished from zero. At the same time, the point estimate for the TVIP score (normed according to the developers table) increases relative from the main specification, but is only significant at 10%. All other results are very close the main specification, indicating that the reclassification of treatment assignment only marginally changes the estimates and does not interfere on the main conclusions of our study.

Column 6 tests the robustness of our results weighting. The specification is similar to our main one, but we don't weight the observations by the inverse of the probability of selection. Although in general the results are smaller than in our preferred specification, qualitatively the results remain almost all unchanged. The only qualitative difference is the impact of parenting practices, which is no longer significant.

Finally, in column 7, we present estimates for the impact of spending an additional month on preschool. Children aged 5 to 9 who were ever enrolled attended preschool by 11 months on average, while target children who have ever been to preschool attended for 13 months on average.[60] Relative to our preferred specification, the only difference is that the participation dummy is replaced by the continuous variable that indicates the reported length of stay on preschool. The

[59] The number of observations vary for each outcome. This would make it infeasible to report all the number of observations, as well as the control average and the control complier mean in the table.

[60] Save the children administrative data indicates that each enrolled child spent on average 16 months on preschool.

instrument is still the binary variable indicating community treatment status after reclassification.

The results presented not only confirm that preschool enrollment is important on accumulating human capital, but the "intensity of the treatment" also matters. Each additional month spent on preschool significantly increases children total ASQ score, as well as their scores on communication, problem solving and precise motor coordination and language skills as measured by the TVIP (normed by developers table) (significant at 10%). Additionally, staying longer on preschool increases the probability of making the transition to primary school. Each additional month on preschool increases the probability of having ever been to primary school, to be currently enrolled in primary school, and to be enrolled in the correct grade for age. This is interesting since the time spent on preschool could in principle mechanically count against being enrolled in primary school, as children cannot be enrolled in preschool and primary school at the same time. Finally, we estimate a positive causal impact of the time spent on preschool on child cognitive factor and on the index of parenting practices.

5.12. Program cost estimates

This appendix presents the methodology used to compute program costs per child per year. We try to detail the costs as much as possible. We believe this can help policy makers understand the structure of the program, which by its turn could clarify how costs could change under different circumstances. Additionally, it also enhances comparability of this to other studies. We included the costs of designing the program, building the preschools, as well as the costs of running the preschools in a daily basis. The project relies heavily on community participation, voluntary labor and in kind contribution. We monetized those costs based on local wages and included them in the budget.

In order to show an example on how we monetized those costs, we show in the next tables the costs of building one simple classroom, of reed walls, tin roof and cement floor, and one latrine. The detailed table is also useful for policy makers who might be interested on estimating the costs of implementing the program in other places where input prices might differ. In this example, the classroom costed USD$ 1343, of which USD$ 311 were donated by the community on labor hours and local materials, as sand. On average, each classroom costed USD$ 1422, of which USD$ 946 were paid by Save the Children, USD$ 226 on materials donated by the community and USD$250 on community labor. Usually each escolinha is equipped with 3 classrooms and one to 3 latrines depending on the location of the classrooms. In this case, the latrine costed about USD$ 255, of which USD$75 were donated by the community. Each escolinha also contains a playground, which costed USD$ 50 on average.

Table 31 shows the total budget for the first 3 years of program implementation, with expenditures from 2007 to June of 2010, including all local labor costs and in kind donations. We constructed the table based on Save the Children expenditures, but as many items were monetized, it doesn't correspond to Save's actual outlays. In order to produce this table, we also had to make several assumptions:

Local materials donated by the community for classroom construction are priced at 226 USD per classroom. The total cost of local labor for classroom construction is priced at 250 USD per classroom.

The total costs of labor for the construction of latrines is $50 USD per latrine. The costs of materials donated by the community is USD$ 10 per latrine.

The total cost of local labor for playground construction is priced at 50 USD per school

Each animadora receive 10 USD per month. Each classroom has 2 animadoras.

School management committee is voluntary. Caregivers' time spent on ECD meetings is priced at zero.

Inflation rate is 12% per year

Exchange rate is 29 MTn per USD

Based on those assumptions, we were able to form columns 1 to 4 of table 33. Each single expenditure was classified and aggregated into broad categories labeled at the left of table 33. The description of expenses contained on each label can be seen at table XX. Next, we calculate the present value of the costs of running the program. The rationale is that the amount allocated for the program would yield the benchmark Mozambican interest rate (assumed here as the standing lending facility rate determined by the Central Bank of Mozambique) and that, over time, both the principal and the interests are fully spent on the program. Therefore, the program is not only financed by the initial funding, but also by the interests generated. To calculate the present value of the costs, we need to make some additional assumptions:

1. Program lasts for 30 years
2. Real interest rate is 5% per year and remains constant
3. Exchange rate remains constant at 29 Mtn per USD
4. Initial expenses with consultants for program design are not repeated.
5. Classrooms last for 15 years (and are reconstructed at every 15 years)
6. Cars last for 8 years (and are bought again every 8 years)
7. Motorbikes last for 5 years (and are bought again every 5 years)
8. Durable learning kits from experimental libraries last for 3 years (and are bought again every 3 years)

Once the flow of expenditures is constructed, everything is brought to present value according to this simple formula:

$$\text{PVTC= Present Value Total Cost} = \sum_{t=1}^{30} \frac{\sum_{n=1}^{N} COST_{nt}}{(1+\pi)^t * (1+i)^t}$$

In which i corresponds to the interest rate, π to inflation, n to each expenditure and t is the time subscript. With that in hand, we only need to know how many children will benefit from the program in 30 years.

According to Save the Children monitoring sheets, the program served 4500 children in the first two years and each child spent approximately 16 months on the program. That means the program benefited 4500*16 children-months in 2 years, where a children-month means one child enrolled for one month. Consequently, in 30 years the project would produce 30*(4500*16)/2 children-months. So the cost per child per month is simply:

$$\text{Cost per child per month=} \frac{PVTC}{30 * \frac{(4500*16)}{2}}$$

Finally, assuming each school will be open for 10 months per year, we just need to multiply the cost per child-month by 10 to compute the annual cost per child. Column 8 of table 32 shows the annual costs per child, broken by each expenditure category. The total cost per child is USD$ 30 96 per year, or USD$ 3.09 per child per month.

Next, we make some additional simulations to check how sensitive the cost per child is to departures of our assumptions. Because most of the costs are fixed per classroom, the more children per classroom, the lower is the cost per child. So, for each different scenario, we present the costs assuming the program benefits 3000, 3600 and 4000 children per year.

Column 1 shows our basic scenario, under an interest rate of 5%, animadoras' wage of USD$ 10 per month, no redesign of curriculum, classrooms lasting for 15 years, cars lasting for 8 years and motorbikes lasting for 5 years. Under this scenario, the cost per child also varies by the number of children-months benefited. Our estimate of USD$ 30.96 can decrease to USD$ 27 if 4000 kids are attended per year, or increase to USD$ 37 if only 3000 kids are attended.

In columns 2 and 3, we vary the interest rate. If the discount is 10% per year, by keeping constant the number of children attended, our estimate decreases to USD$ 19.88, while if the interest rate is 3% per year, costs increase to USD$ 39.89. A discount rate of 3%, however, seems way too low for a developing country. According to Zhuang et al (2007), developing countries tend to apply higher social discount rates between 8% to 15%. In this sense, we are being conservative when using a discount of 5% as our benchmark.

In column 4, we change our assumptions about the program fixed costs. Instead of never being redesigned, we incur in the same costs of designing the curriculum after 15 years. We also shorten the lifetime of classrooms, from 15 to 10 years, as well as cars' lifetime from 8 to 5 years. Interest rate is kept at 5%. In this case, the annual cost per child increases to USD$ 33.57.

In column 5, we reduce even more the time between redesigning the curriculum, rebuilding classrooms and buying new cars. All those fixed costs are now paid every 5 years. Under this scenario, our costs when benefiting 3600 children per year increases to USD$ 33.7, while enrolling 3000 children yields a cost of USD$ 40.45 and enrolling 4000 children yields a cost USD$ 30.33.

Finally, in column 6 we keep our basic scenario, but we increase the wage of the animadoras from USD$ 10 to USD$ 100 per month. Each classroom has 2 animadoras. This salary would be closer to what primary school teachers receive to work at Mozambique public schools. This seems more realistic to a scenario in which the government of Mozambique scales up the program. In this case, the cost per student jumps from USD$ 30.96 to USD$ 50.63. This reveals that the low costs of the program rely heavily on the voluntary nature of the animadoras' work, although an annual cost per child in the range of USD$ 45 to USD$ 60 seems still low.

Table 31 - COSTS OF CONTRUCTION OF ONE CLASSROOM WITH TIN ROOF (IN USD)

Material	Unity	Community donation (unit)	Save The Children donation (unit)	Price per unit (USD)	Total Save the Children (USD)	Total community material (USD)	Total community labor (USD)
Stones for riprap	m3	-	2.00	5.17	100.34	-	-
Blocks for the foundation	Unit	-	90.00	0.43	38.79	-	-
Cement mortar for pavement	Unit	-	10.00	9.24	92.41	-	-
Piles	Unit	-	24.00	3.10	74.48	-	-
Reed (caniço) for walls	sheaf	-	45.00	1.55	69.83	-	-
Battens for wall	sheaf	-	5.00	5.17	25.86	-	-
Pine beams for roof structure measuring 75x50x5m	Unit	-	22.00	7.24	159.31	-	-
Eucalyptus poles for structure	Unit	-	-	1.90	-	-	-
Nails -2 inches	Kg	-	2.00	1.59	3.17	-	-
Nails -3 inches	Kg	-	4.00	1.86	7.45	-	-
Nails -5 inches	Kg	-	3.00	1.69	5.07	-	-
Nails -6 inches	Kg	-	-	1.55	-	-	-
Bonding wire	Kg	-	5.00	1.98	9.90	-	-
Reed (caniço) for roof	sheaf	-		1.55	-	-	-
3mm galvanized wire	Kg	-	10.00	5.00	50.00	-	-
Corrugated zinc plate 12 feet (Tin roof)	Unit	-	25.00	15.00	375.00	-	-
Nails for tin roof	packages	-	3.00	6.72	20.17	-	-
Labor (One construction worker and four helpers)	Days	15.00		12.07	-	-	181.03
Community support (water supply)	25 liters	95.00		0.10	-	9.83	-
Community support (sand supply)	5 m3	2.00		34.48	-	68.97	-
Community support (unloading material)	USD	1.00		51.72	-	-	51.72
TOTAL					1,031.79	78.79	232.76

Note: Exchange rate: 1 USD= 29 MTn

Table 32

COSTS OF CONTRUCTION OF ONE LATRINE (IN USD)

Material	Unity	Community donation (unit)	Save the Children donation (unit)	Price per unit (USD)	Total Save the Children (USD)	Total community material	Total community labor
Concrete slabs	Unit	-	1.00	15.52	15.52	-	-
Blocks for tank coating	m3	-	90.00	0.43	38.79	-	-
Stakes	Unit	-	17.00	3.10	52.76	-	-
2.5mm galvanized wire	Unit	-	10.00	4.14	41.38	-	-
Burnt wire	Kg	-	5.00	1.98	9.90	-	-
Nails 2 inches	Kg	-	2.00	1.59	3.17	-	-
Reed	Sheaf	20.00	-	1.38	-	27.59	-
Concrete for pavement		-	2.00	9.24	18.48	-	-
Labor (builder and helpers)	Days	4.00		12.07	-	-	48.28
TOTAL					180.00	27.59	48.28

Note: Exchange rate: 1 USD= 29 MTn

Table 33

SAVE THE CHILDREN PRESCHOOL BUDGET (IN USD)

	Year				Simulation	Annual cost per child
	2007	2008	2009	2010 (Jan-June)	Repeats	(simulated)
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Costs						
Consultants	12798.92	86398.79	41544.94	48977.08	Never	1.36
Construction of infrastructure	0.00	20713.34	123060.85	2445.65	In 15 years	1.49
Acquisition of cars	56000.00	0.00	0.00	0.00	Every 8 years	1.31
Acquisition of Motorcycles	0.00	17500.00	0.00	0.00	Every 5 years	0.49
Running Costs						0.00
Program staff	19357.00	92972.20	81858.96	36047.82	Every year	7.62
International support staff	0.00	16086.94	17300.52	8182.19	Every year	1.37
National support staff	0.00	14373.27	14545.42	7358.39	Every year	1.20
Teacher incentives	0.00	12902.02	18218.00	14077.00	Every year	1.56
Trainings	636.00	37213.92	39765.13	39708.93	Every year	4.15
Monitoring visits	630.00	16983.06	16548.07	5921.36	Every year	1.30
Production of learning kits	0.00	1646.10	3307.45	2748.60	Every year	0.27
Durable learning kits	0.00	36367.33	16580.54	477.40	Every 3 years	0.81
Children rights intervention	0.00	1444.30	1796.71	1688.92	Every year	0.17
Health interventions	0.00	2315.94	2926.35	2834.48	Every year	0.29
Travel and transportation	249.90	2999.43	6975.29	1573.83	Every year	0.37
Office supplies and fees	0.00	21596.00	21596.00	10350.00	Every year	1.76

Indirect administrative costs (total)	190000.00	Every year	5.45
TOTAL budget (2007-June 2010):			1,249,600.33
Present value of budget (30 years)			3,343,424.55
Children attended (30 years)			108,000
Annual cost per child (30 years):			30.96

Notes: Table shows Save the Children's budget for the ECD Program in Gaza province. All values expressed in dollars. Exchange rate: 1USD=29 Mtn. Expenditures start in 2007 before the program operation. Budget from 2010 corresponds to expenditures made from January until June. The simulation exercise repeats expenditures in determined intervals of time. For example, the program staff wage bill is repeated every year, while school infrastructure is only rebuilt at every 15 years. We assume the program, with 69 built classrooms, benefit about 3600 children in total. The last column (6) shows the annual cost per student for each item of the budget. Last column is calculated by dividing the present value of the budget by 108000 children.

Table 34

DESCRIPTION OF EXPENSES

Consultants	Consultancy and expenses for drafting program design, situation analysis, foundation guides, checklists, fieldworker's guide, animador foundation training and storybook production guide. Design of: Curriculum, learning materials, production plan, games, cards, database. Artist production and design of children's books. Translations of materials. Advocacy and capacity building.
Construction of infrastructure	Materials and labor for construction of classrooms, playgrounds, latrines and water tanks.
Acquisition of cars	Cars
Acquisition of Motorcycles	Motorcycles
Program staff	Program manager, education program coordinator, operations official, construction official, drivers, other staff
International support staff	International staff
National support staff	Finance manager, accountant, personnel manager, service manager, logistician, transport chief, receptionist, assistant accountant, other national staff
Teacher incentives	Seed funding to provide incentives for animadoras, shirts and capulanas
Trainings	Animadoras training, training of preschool management committees, training of community development agents, training on community mobilization, training on monitoring and evaluation, training of provincial and district officials on ECD approach, training of primary school staff on ECD approach, learning circles with animadoras, meetings with leaders and preschool management committee, parenting meetings
Monitoring visits	Fuel, maintenance costs, delivery of preschool kits
Production of learning kits	Annual Replacement Materials: Soap, crayons, pencils, notebooks, ream of paper, copies of M&E tools and curriculum, in kind labor and materials from teachers and community
Durable learning kits	Materials for experimental libraries, laminating machines, Library box, slates, books, soap, crayons, pencils, notebooks, copies and M&E tools
Day of African child	Activities for the day of the African child
Health interventions	Deworming tablets, vitamin A supplements, assistance with child registration and vaccination
Travel and transportation	National and international travel
Office supplies and fees	Office supplies, phone, fax, office rental, utilities, building maintenance, building repair, building security, equipment maintenance, legal fees, bank fees, insurance, computer supplies
Indirect administrative costs (total)	Management and administration indirect costs

Table 35

PRESCHOOL ANNUAL COST PER STUDENT UNDER DIFFERENT SCENARIOS

	Basic Scenario (1)	Scenario 2 (2)	Scenario 3 (3)	Scenario 4 (4)	Scenario 5 (5)	Scenario 6 (6)
If 3600 children are attended per year	USD 30.96	USD 19.88	USD 39.89	USD 33.57	USD 33.70	USD 50.63
If 4000 children are attended per year	USD 27.86	USD 17.90	USD 35.90	USD 30.21	USD 30.33	USD 45.57
If 3000 children are attended per year	USD 37.15	USD 23.86	USD 47.87	USD 40.28	USD 40.45	USD 60.76
Assumptions:						
Curriculum redesigned	Never	Never	Never	15 years	5 years	Never
Classrooms rebuilt	15 years	15 years	15 years	10 years	5 years	15 years
Cars bought	8 years	8 years	8 years	5 years	5 years	8 years
Motorbikes	5 years	5 years	5 years	5 years	5 years	5 years
Animadoras' wage (USD)	USD 10.00	USD 10.00	USD 10.00	USD 10.00	USD 10.00	USD 100.00
Exchange rate: Mtn per USD	29 Mtn	29 Mtn	29 Mtn	29 Mtn	29 Mtn	29 Mtn
Interest rate	5%	10%	3%	5%	5%	5%

Note: Table shows simulations of the cost per student under different scenarios. For each scenario, we calculate the cost per student if 3600 children are attended per year, if 4000 children are attended or if 3000 children are attended. The assumptions underlying each scenario are described in the lower half of the table. The basic scenario assumes that the program lasts for 30 years, the curriculum is never redesigned, classrooms are rebuilt every 15 years, cars are bought every 8 years, motorcycles are replaced every 5 years, each classroom has 2 animadoras who receive USD 10 per month each one, inflation rate is 12%, real interest rate is 5% per year, exchange rate is 1USD= 29 Mtn.

5.13. Cost effectiveness analysis

Once we calculate the cost per student, it seems natural to compare the costs to the benefits of the program. A cost-benefit analysis would allow us to put both the benefits and costs in the same monetary scale, and allow us to compare the rate of return of this program to returns of several alternative programs. However, it would require us to make a series of assumptions that would be very hard to make at this point. How many extra years of education will each child enrolled in preschool get because of Save's program? What is the return to each year of education in the context of rural Mozambique? For the same level of educational attainment, what there other the market returns of having been to preschool?

In the absence of reasonable answers to those questions, a cost-effectiveness analysis of the program is our the best alternative in order to generate some comparison with alternative interventions. Despite demanding much less heroic assumptions, a cost effectiveness analysis, nonetheless still poses several challenges to the researcher. While in a cost benefit analysis there is one single monetary dimension to focus on, we can make a cost effectiveness analysis for many of the multiple benefits of being enrolled in preschool.

We choose the Ages and Stages Questionnaire as our main outcomes for the cost effectiveness analysis. [61] Although comparing a child ability to kick a ball, draw a line, make circles and align objects in Mozambique with learning high school chemistry in US might sound a bit odd, it is common practice in the education literature to use the standard deviation of scores in the control group as a reference scale in order to compare gains from different interventions. In this case, calculating the gains is straightforward and practical.

We start by calculating the present value of the impact, that was measured one year and half to two years and half after the start of the intervention. While some students went to preschool for two years, others only started going to preschool much later. We thus discount the impact for one period. Then we divide

[61] Choosing the schooling gains arising from the increase on the probability of being currently enrolled in primary school or gains on the Early Development Index in primary school would involve additional challenges we wish to avoid here. For example, part of the increase on the probability of being in primary school comes from enrolling the child in primary school at the correct age. For the EDI, spillover in the classroom would force us to calculate benefits for a group of kids that is larger than the group of beneficiaries.

the impact by the cost per child. This yields a cost of USD$ 85 for each additional standard deviation, or a gain of 1.14 standard deviation for each USD$ 100 invested.

We also analyzed the sensitivity of our estimate of cost-effectiveness, both to the imprecision of the estimate of the impact, but also to the different assumptions that had to be made to calculate the cost per student. We therefore estimate the upper bound for the cost effectiveness by dividing the upper 90% confidence interval of the estimate by our lowest estimate of the cost per student, resulting in 2.81 standard deviations of the ASQ score for each USD$ 100 invested. By the other hand, we calculate the lower bound of the cost effectiveness by dividing the lower 90% confidence interval by our highest estimate of cost per student, when teachers receive USD$ 100 per month. In this case, we have 0.33 standard deviation gain for each USD$ 100 spent.

Finally, we compare our estimates of cost-effectiveness to other estimates in the literature analyzed by Dhaliwal, Duflo, Glennerster and Tulloch (2013). The Mozambique preschools rural preschool impact of 1.14 std for each USD$ 100 ranks in between the cost effectiveness of Read-a-Thon program in the Phillipines (Abeberese et al, 2012) and the cost effectiveness of Minimum Conditional Cash Transfer in Malawi (Baird et al, 2011). The ranking, however, is sensitive to adjustments for the imprecision on the estimation of the program impact and on the calculation of costs. For example, if we had instead used a 10% discount when computing our cost per student, Mozambique preschool would be rank behind the extra contract teacher initiative in Kenya (Duflo, Dupas and Kremer, 2011 and Duflo, Dupas and Kremer, 2012) and the individually paced computer assisted program in India (Banerjee, Cole, Duflo and Linden, 2007). Ranking by the lower and upper estimates of the cost estimates would produce different ordering of the programs. Consequently, any of those ranks should be read with this caveat in mind.

Table 36

COST EFFECTIVENESS ANALYIS
TOTAL ASQ SCORE

Impact per child	0.370***
	(0.096)
Present value of impact:	0.3524
Cost per child:	USD 30.96
Cost per additional SD	USD 87.85
Additional SD per USD 100	1.14 SD

Table 37

SENSITIVITY ANALYSIS -TOTAL ASQ SCORE

90% CI of Impact Estimate (1)	Impact Estimate (SD) (2)	Lowest and highest Cost (3)	Cost per Additional SD (4)	Additional SD per $100 (5)
Point Estimate	0.35	USD 30.96	USD 87.85	1.14
Upper Bound	0.50	USD 17.90	USD 35.56	2.81
Lower Bound	0.20	USD 60.76	USD 301.50	0.33

Figure 14 – Cost effectiveness of other programs

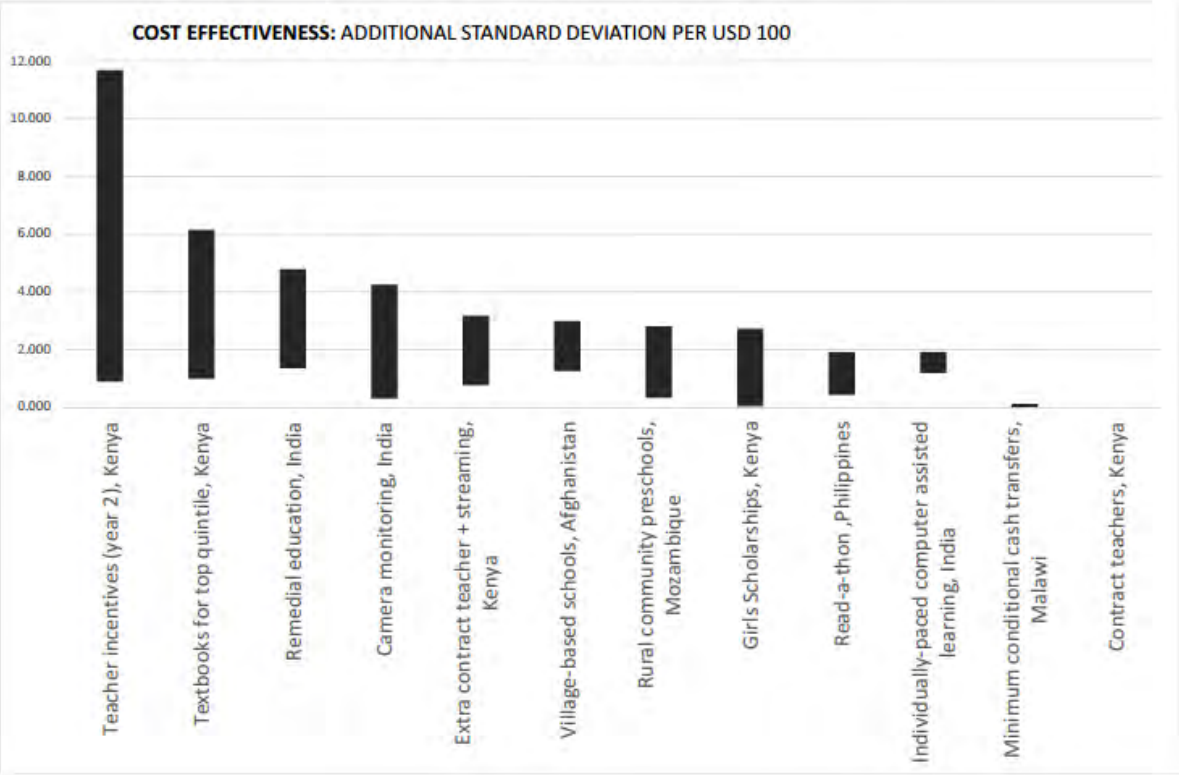


Table 38

COST EFFECTIVENESS OF PROGRAM IMPACT

Study Name	Country	Primary Source Papers	Cost per Additional SD (USD)	Additional Standard Deviation per USD$ 100 invested		
				Lower bound	Point estimate	Upper bound
			(1)	(2)	(3)	(4)
Rural community preschools	Mozambique	Martinez, Naudeau and Pereira. "The Promise of Preschool: Results of a Randomized Controlled Study in Rural Mozambique". Working paper, May 2016	$87.85	0.33	1.140	2.81
Unconditional cash transfers	Malawi	Baird, Sarah, Craig McIntosh, and Berk Ozler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly Journal of Economics* 126 (4): 1709-1753.		No significant impact		
Minimum conditional cash transfers	Malawi	Baird, Sarah, Craig McIntosh, and Berk Ozler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly Journal of Economics* 126 (4): 1709-1753.	$1,667.43	0.002	0.060	0.118
Girls Scholarships	Kenya	Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *The Review of Economics and Statistics* 91 (3): 437-456.	$72.26	0.035	1.384	2.733
Village-based schools	Afghanistan	Burde, Dana and Leigh Linden. "The Effect of Village-Based Schools: Evidence from a Randomized Controlled Trial in Afghanistan." Working Paper, May 2012.	$47.05	1.257	2.126	2.994

Providing earnings information	Madagascar	Nguyen, Trang. "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar." Working Paper, January 2008.	$0.85	16.187	118.338	220.490
Reducing class size	Kenya	Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101 (August 2011): 1739-1774. Duflo, Esther, Pascaline Dupas, and Michael Kremer. "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." NBER Working Paper #17939, June 2012.		No significant impact		
Textbooks	Kenya	Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics* (1) 1: 112-135.		No significant impact		
Textbooks for top quintile	Kenya	Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics* (1) 1: 112-135.	$28.06	0.982	3.563	6.145
Flipcharts	Kenya	Glewwe, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz. 2004. "Retrospective vs. Prospective Analyses of School Inputs: the Case of Flip Charts in Kenya." *Journal of Development Economics* 74: 251-268.		No significant impact		
Reducing class size	India	Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122(3):1235-1264.		No significant impact		
Building/improving libraries	India	Borkum, Evan, Fang He, and Leigh Linden. "School Libraries and Language Skills in Indian Primary Schools: A Randomized Evaluation of the Akshara Library Program." Working Paper, December 2009.		No significant impact		

School committee grants	Indonesia	Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Arya Gaduh, and Rima Prama Artha. "Improving Educational Quality Through Enhancing Community Participation: Results from a Randomised Field Experiment in Indonesia." Working Paper, April 2012.		No significant impact		
School committee grants	Gambia	Blimpo, Moussa, and David Evans. "School-Based Management and Educational Outcomes: Lessons from a Randomized Field Experiment." Working Paper, November 2011.		No significant impact		
Adding computers to classrooms	Colombia	Barrera-Osorio, Felipe and Leigh Linden. "The Use and Misuse of Computers in Education: Evidence from a Randomized Controlled Trial of a Language Arts Program." Working Paper, March 2009.		No significant impact		
One Laptop Per Child	Peru	Cristia, Julián, Pablo Ibarrán, Santiago Cueto, Ana Santiago, and Eugenio Severín. "Technology and Child Development: Evidence from the One Laptop per Child Program." *IZA Discussion Paper* No. 6401, March 2012.		No significant impact		
Diagnostic feedback	India	Muralidharan, Karthik and Venkatesh Sundararaman. 2010. "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India." *The Economic Journal* 120: F187-F203.		No significant impact		
Read-a-thon	Philippines	Abeberese, Ama Baafra, Todd Kumler, and Leigh Linden. "Improving Reading Skills by Encouraging Children to Read: A Randomized Evaluation of the Sa Aklat Sisikat Reading Program in the Philippines." Working Paper, June 2012.	$85.07	0.432	1.176	1.919
Individually-paced computer assisted learning	India	Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India."*The Quarterly Journal of Economics* 122(3):1235-1264.	$64.46	1.186	1.551	1.917

Extra contract teacher + streaming	Kenya	Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101 (August 2011): 1739-1774. Duflo, Esther, Pascaline Dupas, and Michael Kremer. "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." NBER Working Paper #17939, June 2012.	$50.74	0.768	1.971	3.174
Remedial education	India	Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122(3):1235-1264.	$32.59	1.350	3.069	4.788
Streaming by achievement	Kenya	Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." American Economic Review 101 (August 2011): 1739-1774. Duflo, Esther, Pascaline Dupas, and Michael Kremer. "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." NBER Working Paper #17939, June 2012.	$2.87	9.750	34.784	59.818
Contract teachers	Kenya	Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." American Economic Review 101 (August 2011): 1739-1774. Duflo, Esther, Pascaline Dupas, and Michael Kremer. "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." NBER Working Paper #17939, June 2012.	-$334.54	-0.424	-0.299	-0.174

Teacher incentives (year 1)	Kenya	Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (July): 1-25.		No significant impact		
Teacher incentives (year 2)	Kenya	Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (July): 1-25.	$15.90	0.888	6.291	11.694
Teacher incentives (long-run)	Kenya	Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (July): 1-25.		No significant impact		
Camera monitoring	India	Duflo, Esther, Rema Hanna, and Stephen Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." American Economic Review 102(4): 1241–1278.	$43.90	0.294	2.278	4.262
Training school committees	Indonesia	Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Arya Gaduh, and Rima Prama Artha. "Improving Educational Quality Through Enhancing Community Participation: Results from a Randomised Field Experiment in Indonesia." Working Paper, April 2012.		No significant impact		
Grants & training for school committee	Gambia	Blimpo, Moussa, and David Evans. "School-Based Management and Educational Outcomes: Lessons from a Randomized Field Experiment." Working Paper, November 2011.		No significant impact		
Electing school committee & linking to local government	Indonesia	Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Arya Gaduh, and Rima Prama Artha. "Improving Educational Quality Through Enhancing Community Participation: Results from a Randomised Field Experiment in Indonesia." Working Paper, April 2012.	$7.50	3.891	13.337	22.784
Linking school committee to local government	Indonesia	Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Arya Gaduh, and Rima Prama Artha. "Improving Educational Quality Through Enhancing Community Participation: Results from a Randomised Field Experiment in Indonesia." Working Paper, April 2012.	$2.89	11.496	34.624	57.752

Notes: The table shows the cost effectiveness of different programs, based on J-PAL compilation and sensitivity analysis reported in Dhaliwal, Duflo, Glennerster and Tulloch (2013). Columns 2 and 4 presents the lower and upper bound for the cost-effective analysis using the 90% CI for the impact estimate. Our estimate of lower and upper bounds for cost effectiveness, in addition to considering the imprecision of the impact estimate, also uses different estimates for the cost per student.

6
Appendix Chapter 2

6.1. Institutional environment

The school system is decentralized in Brazil. According to the Constitution, the governments of the 26 states and the Federal District are responsible for primary and secondary education, while the municipalities are responsible for early childhood and primary education.

The Federal Government is directly responsible for tertiary education. In the state of Rio de Janeiro, 96% of public high schools are managed by the state government, and 4% by the federal government. The share of primary schools run by state governments varies widely from state to state. In the state of Rio de Janeiro, 20% of public primary schools are run by the state, but this distribution also varies by municipality. In Rio de Janeiro capital city, 95% of primary schools from 1st to 9th grade are run by the municipality.

The academic year runs from February to December, and the school year's length is determined by federal law to be at least 200 days and 800 hours long. In order to pass a grade, children need to have at least 75% attendance and attain a passing grade. There are no official achievement standards required for attaining passing grades. The process of promoting or retaining a student is discretionary, and based on student scores from tests prepared and graded by each teacher, as well as teacher subjective evaluations of student behavior.

The school system is divided into three categories: Early primary education corresponds to 1st to 5th, and late primary education corresponds to 6th to 9th. High school in regular schools lasts three years, from 10th to 12th grade. Children are expected to start school 1st grade when 6 years old, and cannot legally work in formal jobs before age 14. Since the nineties, several reforms have been implemented in order to increase education attainment and school accountability. At the same time,

the federal government has been trying to increase tertiary enrollment by expanding college scholarships and credit to students at private institutions, which concentrate 75% of total college enrollment (Censo da Educação Superior, 2011).

Despite all these efforts, the increase in high school completion rates observed in the period has lagged behind targets set by the National Plan of Education and by civil society movements (Todos pela Educação, 2014). According to PNAD, only 54% of students younger than 19 graduate from high school, while the intermediate target set for 2012 was 68%. In addition, historical disparities on school attainment according to family income still remain. Eighty percent of youths from the first quintile of the income distribution, ie, whose per capita income is higher than R$ 1050, graduate from high school. In contrast, only a third of youths whose per capita income is lower than R$ 100 graduate from high school before turning 24.

In the state of Rio de Janeiro, the scenario is even more dramatic. Notwithstanding recent improvements in high school graduation rates at public schools at other states, the percentage of high school graduates under 19 years old has been stagnant since 2008. According to PNAD, the percentage of youths under 19 who graduated from high school has actually fallen from 55% to 53% between 2008 and 2011.

6.2. Renda Melhor, Cartão Família Carioca and Renda Melhor Jovem programs

Program Renda Melhor is a means tested cash transfer program that works over and above program Bolsa Família in the State of Rio de Janeiro, for all municipalities except for the State's capital. The main goal of Renda Melhor transfers is to complement transfers from Bolsa Família in order to raise family per capita income to the State poverty line of R$ 100 per capita.

Beneficiary families are targeted according to an index of living conditions calculated using the Cadastro Único para Programas Sociais data (CadUnico), the household administrative data from program Bolsa Família. The index predicts family income and is measured in Brazilian Reais. Families whose predicted per capita income after accounting for governmental transfers is lower than R$ 100

receive a cash transfer whose total amount is equal to the difference between R$ 100 and the per capita predicted income after transfers, multiplied by the number of household members. The minimum transfer is R$ 30 and the maximum is R$ 300.

In the capital of the State, the city of Rio de Janeiro, Bolsa Familia transfers are matched through another program, called Cartão Familia Carioca, managed by the city government. The program is very similar to Renda Melhor. The main difference is that it matches Bolsa Família transfers to a poverty line set at R$ 108, and that it incentives children to attain good grades in primary school. The program also targets beneficiaries though a living conditions index and it was implemented in December 2010, five months before the start of Renda Melhor. The minimum transfer is R$ 20, the maximum is R$ 400 and only up to 3 children and youths aged less than 17 are counted when calculating the total transfer.

6.3. Data

In this section I better detail the datasets used in the paper:

1- Renda Melhor Jovem Program roster. This is the main administrative data from Renda Melhor Jovem Program. It is provided by the Secretariat of Social Protection and Human Rights from the State of Rio de Janeiro and contains information from 58,883 students who were eligible to receive Renda Melhor Jovem award between 2011 and 2012. The data has identifying variables such as student's full name, parents' names, date of birth, NIS, matriculation ID, school code, municipality, and student grade. In addition, it also includes crucial information for program administration, such as students' status regarding completion of each step of the registration process, bank agency designated to open the account, name of the person who opened the account and amount transferred to each student.

2- Renda Melhor program Roster. The administrative data from program Renda Melhor is provided by the Secretariat of Social Protection and Human Rights from the State of Rio de Janeiro. It contains the amount transferred to each family enrolled in the program, as well as their predicted household income and predicted per capita

income. The data is detailed at the household level. Beneficiaries are usually the mother and are identified by their NIS.

3. Secondary school enrollment records. This data is provided by the Secretariat of Education of the State of Rio de Janeiro and contains the records of all students enrolled in regular public high schools in the State of Rio from 2010 to 2012. The data includes all students enrolled at any time of the year, including students who drop out during the year. It contains all students' names, date of birth, mother's name, school, grade and Matriculation ID, school shift (morning, afternoon or night). It also includes the final situation by the end of the year, which can be: passing grade, failing grade as a result of low grades, or failing because of absenteeism. Typically, students who drop out during the school year are registered as having failed due to absenteeism, as they will reach the limit of absent days or they will fail to show up to final exams.

4. School Census from INEP. The school census is collected every year by INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), a research institute connected to the Ministry of Education (MEC). School Census provides information on school infrastructure and on the prole of students and teachers by each grade and school, for all schools in Brazil, including private and public schools. Participation in the census is mandatory for all schools.

The census is collected in two steps. First, around May, schools are required to send the roster of all students and teachers at the school to the Ministry of Education. At this step, the principal also reports the information about school infrastructure. Students and teachers are assigned a unique census ID that is used to follow students and teacher over time and across schools. Later, around February of the subsequent year, after the end of previous school year, principals fill in the information about the final situation of each student, including the students who were enrolled at the school after the first phase of the census in May.

There are 5 possibilities for final situation: approved, failed, dropped, transferred to another school or deceased. With all this information in hands, the Ministry of Education calculates some key statistics, which are then made publicly available.

5- Grade passing, grade failure and dropout rates per school, from INEP. After receiving all the information from the second phase of the census, INEP calculates the approval, retention and dropout rates for each school, which by construction sum to one. Rates are calculated by dropping deceased students and by counting students who transferred during the school year at their destination school.

Dropout means the student was absent or was not found by the end of the previous school year, but has nor died and has not been transferred to another school. It does not necessarily mean that the student will not be enrolled in next year. Grade passing means that the student was approved to enroll in the next highest grade in the following year, but does not mean that the student was promoted, as he can still leave school. Similarly, grade failure rates are not repetition rates. Those measures, however, are very correlated with dropout, promotion and repetition rates.

Grade passing rates for each school are then used to calculate an official index of school quality, IDEB (Índice de Desenvolvimento da Educação Básica), that combines passing grades and test scores and is used to allocate some resources across schools by the Ministry of Education.

6- Grade-age distortion average class sizes per school from INEP: By using the information from the first phase of the Census, INEP also calculates the average number of students per class by dividing the number of enrolled students by the number of classes and the grade-age distortion rate for each grade and school. The grade-age distortion rate is the proportion of students that are 2 or more years older than their expected age for their grade. For instance, students are expected to be aged 15 when starting high school at 10^{th} grade and are considered to be over age if they are enrolled at 10^{th} grade and are 17 or older. Grade-age distortion is not calculated for students enrolled in special night classes for adults.

7- Students per grade and school characteristics from Census micro data set: The microdataset from the census has some basic information from each student

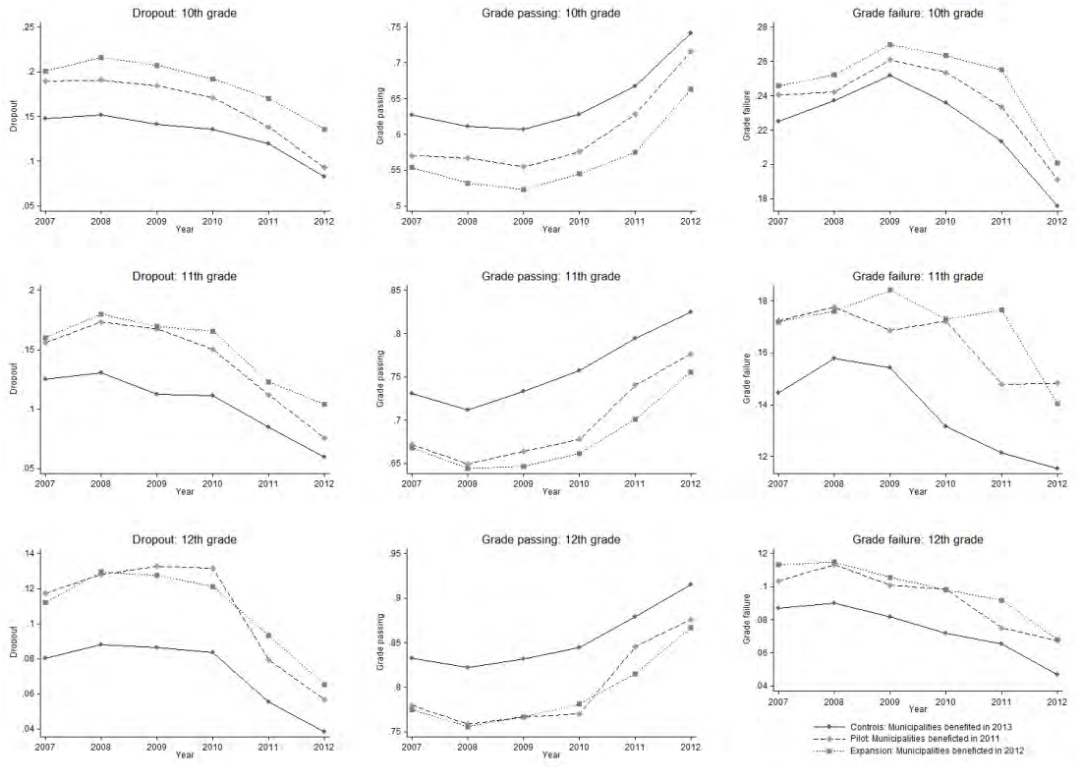
such as race, gender, date of birth, disability, municipality of birth, municipality of residence and grade. It also contains some characteristics of teachers such as race, age, gender, subject taught, background. Finally, it contains school characteristics such as the number of students per grade, number of teachers and information on school infrastructure such as presence of potable water, sanitation, number of classrooms, number of bathrooms, number of computers, Tvs, presence of broadband internet internet connection, among other items.

8-SAERJ test scores from the Secretariat of Education of the State of Rio de Janeiro. SAERJ (Sistema de Avaliação da Educação do Rio de Janeiro - Education Evaluation System for the State of Rio de Janeiro) is an item response theory (IRT) calibrated high stakes test on language and math applied every year to all senior high school students from public schools in the State of Rio de Janeiro.

By combining SAERJ scores and passing grades, the Secretariat of Education of the State of Rio de Janeiro calculates the school quality index of Rio de Janeiro (IDERJ- Índice de Desenvolvimento da Educação do Estado do Rio de Janeiro). SAERJ exam also contains a socio-economic questionnaire with basic information on students' socio-economic conditions, as number of rooms and toilets at home, ownership of car, computers, tvs and dvd players and parental educational. In addition, it includes questions on the frequency with which students read books and newspapers and perceptions about teachers, colleagues and school principals. Data is available from 2010 to 2012.

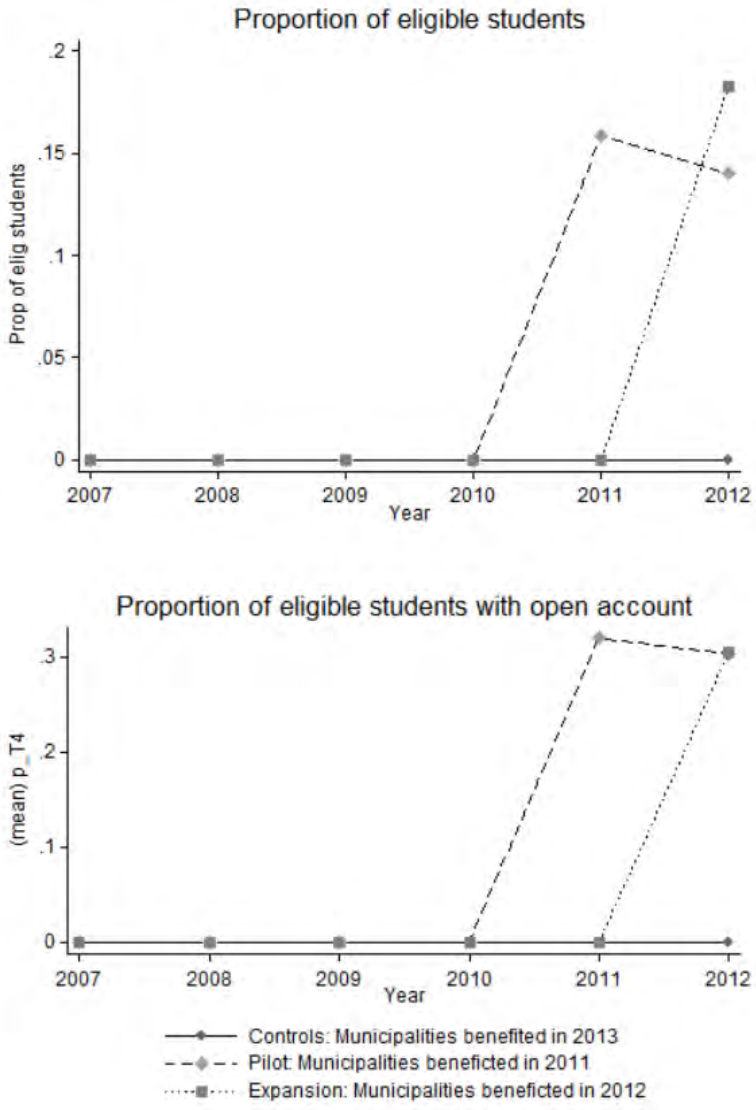
6.4.Additional figures

Figure 1- Trends by grade



Notes: Grade-school level data from INEP. Control municipalities only received the program in 2013, and are represented by the continuous line. Pilot municipalities that received the program in 2011 are represented by the dashed line, while municipalities that received the program in the first wave of expansion in 2012 are represented by the dotted line.

Figure 2- Eligibility and take up over time

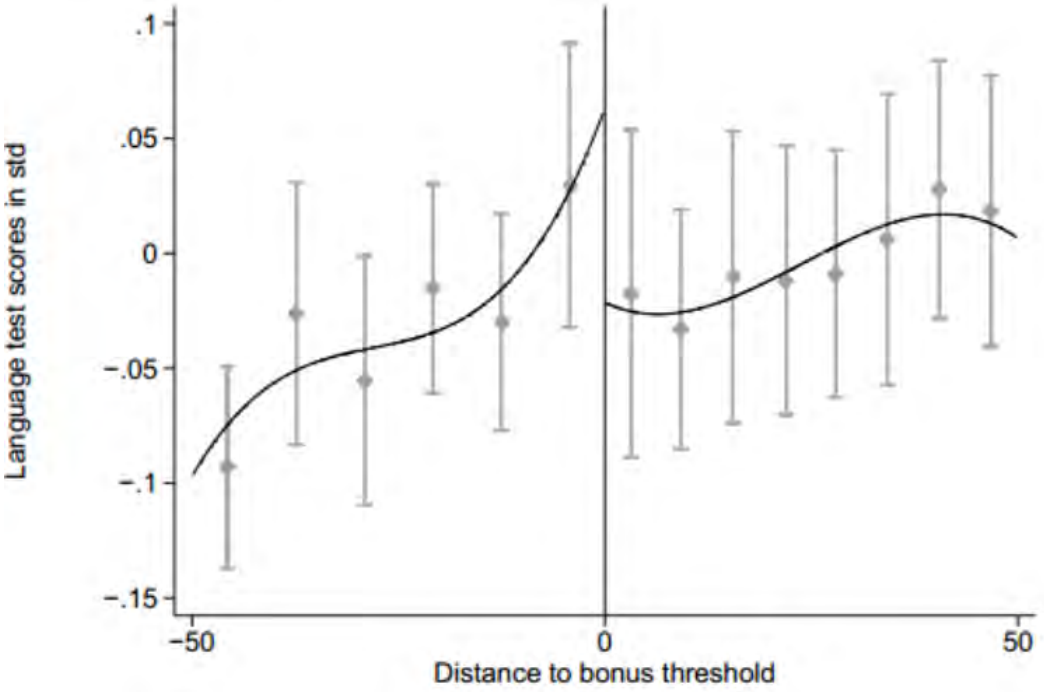


Notes: Control municipalities only received the program in 2013, and are represented by the continuous line. Pilot municipalities that received the program in 2011 are represented by the dashed line, while municipalities that received the program in the first wave of expansion in 2012 are represented by the dotted line.

7
Appendix Chapter 3

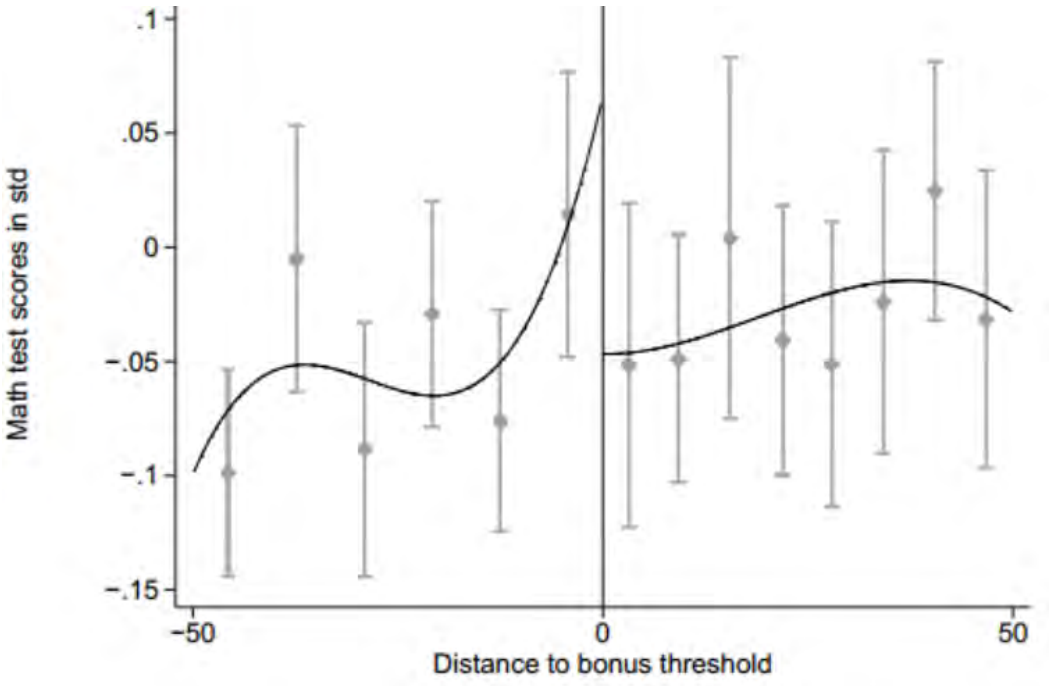
7.1. Additional Figures

Figure 1- Language test scores



Note: Figures 1 presents local linear estimates of student test scores against the distance to the bonus threshold. Schools at the right side of the threshold receive the bonus.

Figure 2- Math test scores



Note: Figures 2 presents local linear estimates of student test scores against the distance to the bonus threshold. Schools at the right side of the threshold receive the bonus.

Figure 3-Language test scores (2009)

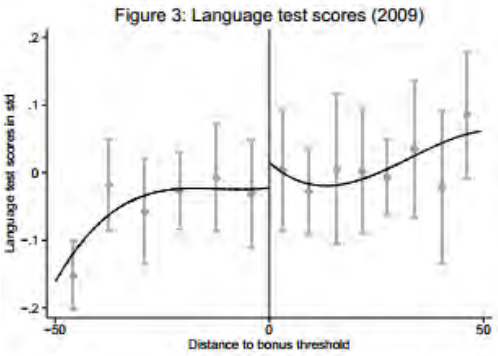
Figure 4-Language test scores (2010)

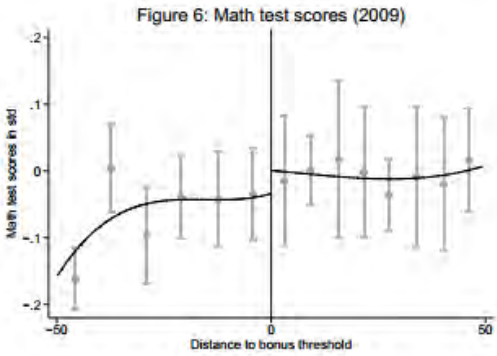
Figure 5-Language test scores (2011)

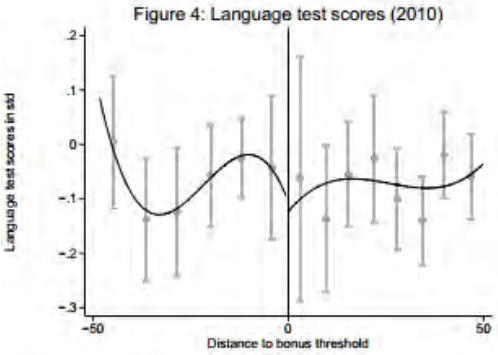
Figure 6-Math test scores (2009)

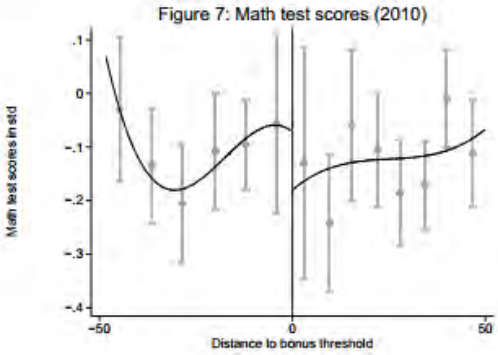
Figure 7-Math test scores (2010)

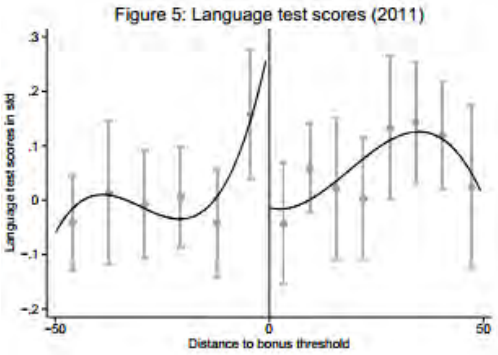
Figure 8-Math test scores (2011)

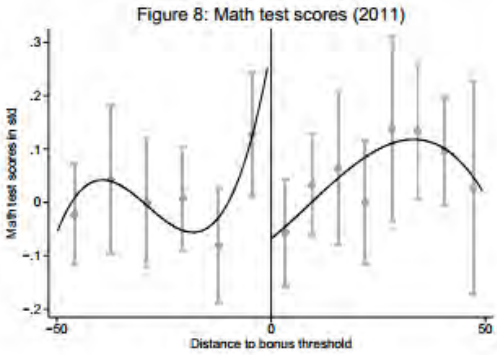






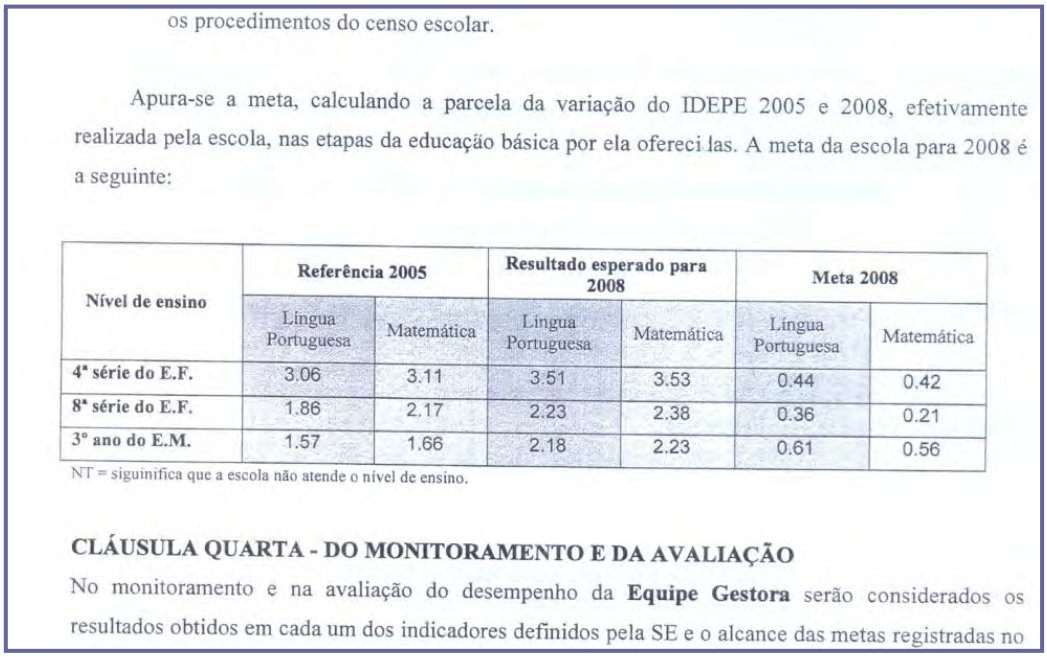






Notes: Figures 3 to 8 present local linear estimates of student test scores against the distance to the bonus threshold. Schools at the right side of the threshold receive the bonus.

Figure 9- Agreement between school principals and the Secretariat of Education



os procedimentos do censo escolar.

Apura-se a meta, calculando a parcela da variação do IDEPE 2005 e 2008, efetivamente realizada pela escola, nas etapas da educação básica por ela oferecidas. A meta da escola para 2008 é a seguinte:

Nível de ensino	Referência 2005		Resultado esperado para 2008		Meta 2008	
	Língua Portuguesa	Matemática	Língua Portuguesa	Matemática	Língua Portuguesa	Matemática
4ª série do E.F.	3.06	3.11	3.51	3.53	0.44	0.42
8ª série do E.F.	1.86	2.17	2.23	2.38	0.36	0.21
3º ano do E.M.	1.57	1.66	2.18	2.23	0.61	0.56

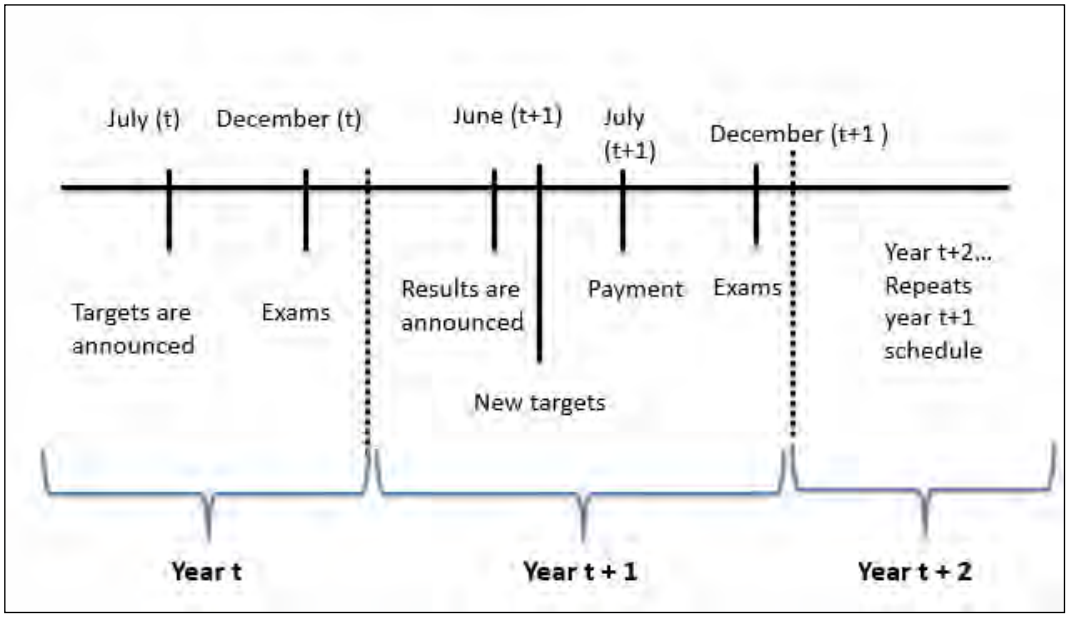
NT = siguinifica que a escola não atende o nível de ensino.

CLÁUSULA QUARTA - DO MONITORAMENTO E DA AVALIAÇÃO

No monitoramento e na avaliação do desempenho da **Equipe Gestora** serão considerados os resultados obtidos em cada um dos indicadores definidos pela SE e o alcance das metas registradas no

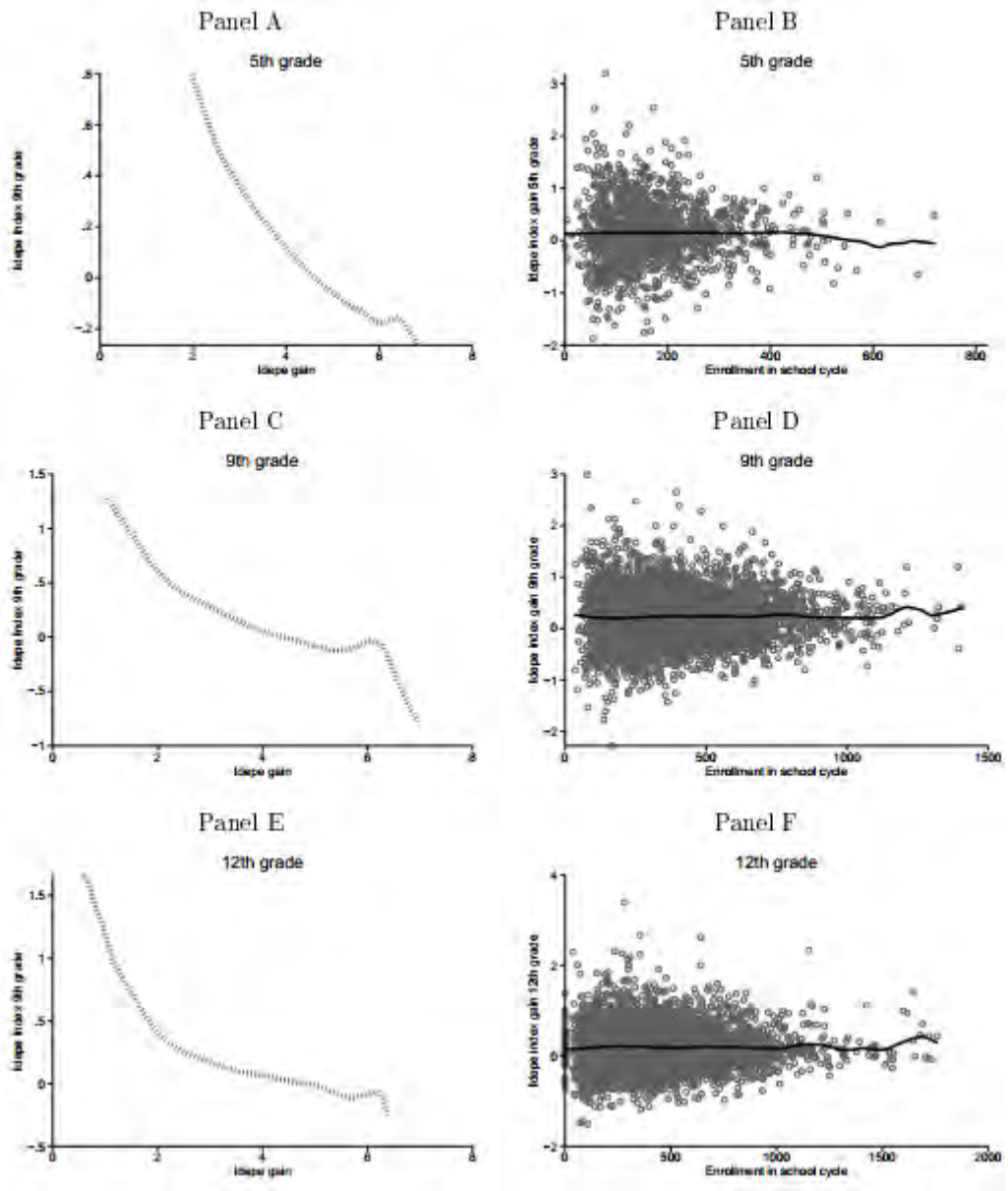
Notes: Figure 9 shows a picture of part of the agreement signed between the school principal and the Secretariat of Education of Pernambuco. The table shown contains the school targets, for language and math, for each school segment. First 2 columns show previous scores, upon which targets are calculated. Columns 3 and 4 show the expected levels of quality index for language and math. Columns 5 and 6 show the performance gains needed to fully achieve the target. The main index that determines if the school receives the bonus is the average of the targets from columns 5 and 6, weighted by the relative number of students in each segment.

Figure 10- Timeline of the bonus scheme



Note: Figure 11 shows the timeline of events of the pay for performance scheme in Pernambuco. The school year runs from February to December. Every year, targets for the bonus are announced around May or June, immediately after results from the exams from December of the previous year come out. The payment of the bonus is also made right after results come out. Exams are taken by December. The next year, the results are compared with targets and the bonus is paid.

Figure 11- Reversion to the mean



Notes: Panels A, C and E show a local linear prediction of yearly gains of Idepe, based on current levels of Idepe index, for years 2008 to 2014. Panels B, D and F plot yearly Idepe gains against student enrollment for each school cycle, from 2008 to 2014. The black line is a local linear estimate Idepe gains.

7.2. Socio Economic Index

Table 1-Socioeconomic index-factor loadings

SOCIO ECONOMIC INDEX FACTOR LOADINGS AFTER OBLIQUE ROTATION		
	Loadings	Uniqueness
Sex (female=1)		
Race (white=1)	0.0509	0.9974
VHS player	0.7796	0.3923
Fridge	0.6486	0.5793
Freexer	0.505	0.7449
Washing machine	0.7098	0.4962
Computer	0.6353	0.5963
Resides with mother	0.1253	0.9843
Resides with father	-0.0924	0.9915
Works	-0.1091	0.9881
Started studying at preschool	-0.3713	0.8621
Moved from school	-0.2127	0.9548

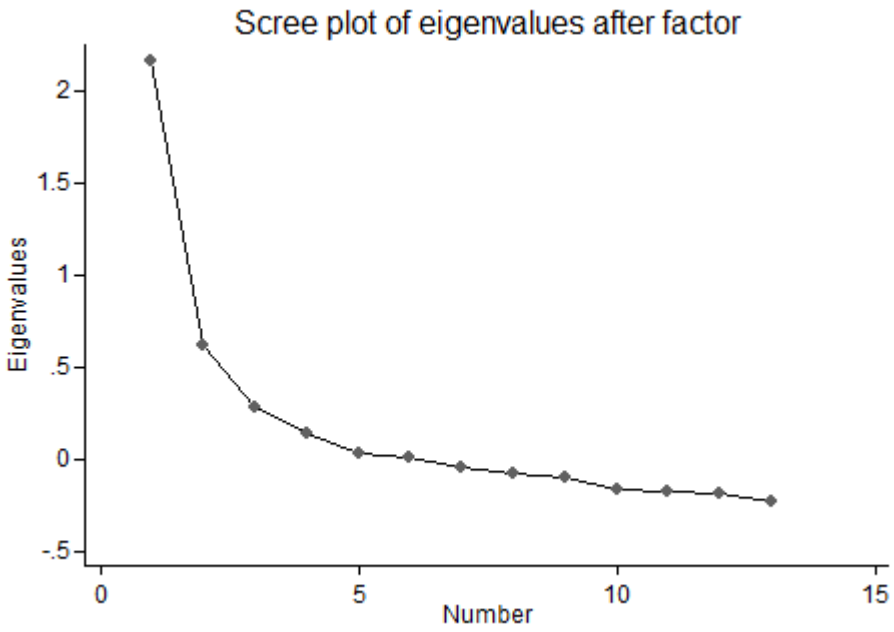
Source: SAEPE student survey 2008

Table 2-Socioeconomic index-KMO

FACTOR - ADEQUACY	
Kaiser-Meyer-Olkin measure of sampling adequacy	
Sex (female=1)	0.6837
Race (white=1)	0.6196
VHS player	0.7477
Fridge	0.768
Freexer	0.8269
Washing machine	0.8066
Computer	0.8278
Resides with mother	0.6248
Resides with father	0.6229
Works	0.5668
Started studying at preschool	0.7982
Moved from school	0.7268

Source: SAEPE student survey 2008

Figure 12- Socioeconomic index- screeplot



7.3.Indices of working conditions

7.3.1.Principal leadership

Table 3-Principal leadership-factor loadings

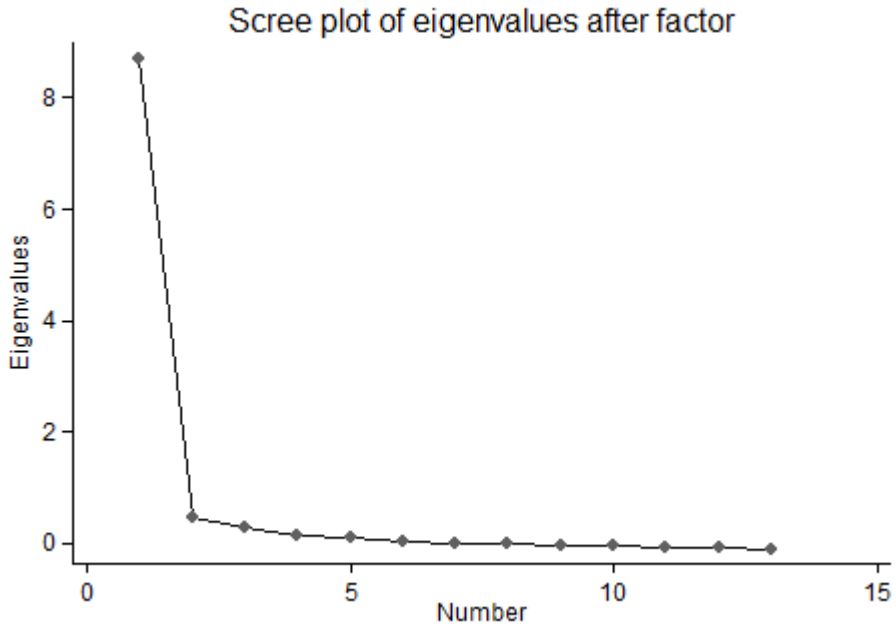
PRINCIPAL LEADERSHIP INDEX
FACTOR LOADINGS AFTER OBLIQUE ROTATION

	Loadings	Uniqueness
The school principal encourages me and motivates me to work.	0.8675	0.2475
I have full professional confidence in the school principal.	0.8945	0.1999
The school principal manages to make teachers commit themselves to the school	0.8383	0.2972
The school principal stimulates innovative activities.	0.8838	0.2189
The school principal pays particular attention to aspects related to student learning	0.8852	0.2164
The meetings held by school principal are dynamic.	0.7979	0.3634
The school principal is frequently absent from school.	-0.3839	0.8526
The school principal is particularly engaged in improving the school	0.8689	0.245
I feel respected by the school principal.	0.8572	0.2653
I respect the school principal.	0.7659	0.4134
The school principal, teachers and other members of the school staff collaborate to make this school work	0.7418	0.4497
The school principal implements clear rules.	0.8241	0.3209
The school principal supports me when I need.	0.8653	0.2512

Table 4- Principal leadership-KMO

PRINCIPAL LEADERSHIP INDEX FACTOR ADEQUACY	
Kaiser-Meyer-Olkin measure of sampling adequacy	
The school principal encourages me and motivates me to work.	0.9597
I have full professional confidence in the school principal.	0.9594
The school principal manages to make teachers commit themselves to the school	0.9686
The school principal stimulates innovative activities.	0.9635
The school principal pays particular attention to aspects related to student learning	0.9669
The meetings held by school principal are dynamic.	0.9833
The school principal is frequently absent from school.	0.9592
The school principal is particularly engaged in improving the school	0.9777
I feel respected by the school principal.	0.9329
I respect the school principal.	0.9407
The school principal, teachers and other members of the school staff collaborate to make this school work	0.9544
The school principal implements clear rules.	0.9698
The school principal supports me when I need.	0.9634
Overall	0.9617

Figure 13- Principal leadership-Screeplot



7.3.2.Teamwork

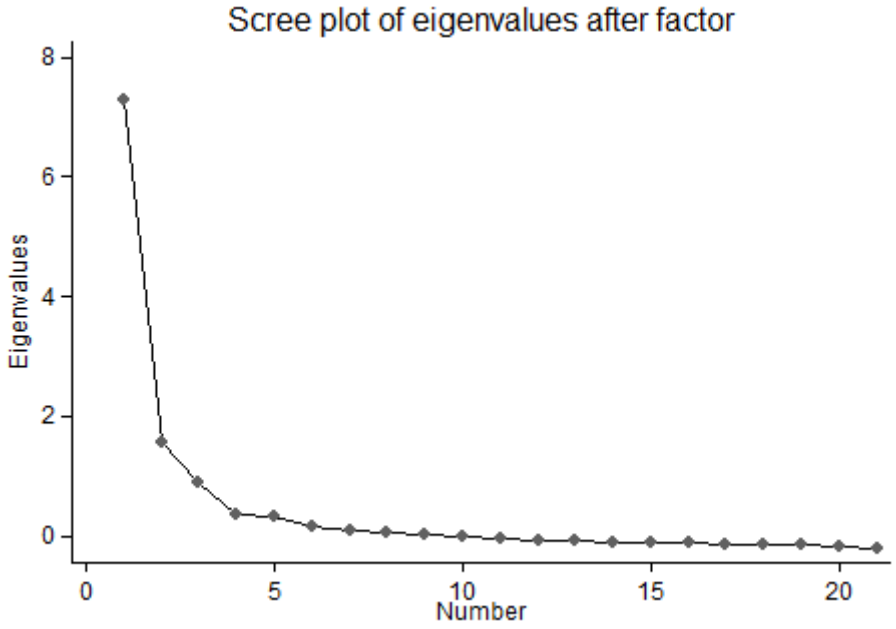
Table 5- Teamwork-factor loadings

TEAMWORK INDEX FACTOR LOADINGS AFTER OBLIQUE ROTATION	Loadings	Uniqueness
The political-pedagogical project of this school is a result of the exchange of ideas between teachers	0.6194	0.6164
Most teachers are committed to improving their classes.	0.7029	0.5059
Few teachers take responsibility for improving the school.	-0.4599	0.7885
Most teachers maintain high expectations about the learning of their students	0.6537	0.5726
Few teachers are willing to take on new charges to improve the school	-0.4407	0.8057
Most teachers are receptive to the implementation of new ideas.	0.6464	0.5822
Most teachers are overloaded with work, which harms lessons planning	-0.3958	0.8433
Most teachers feel responsible for student performance.	0.6091	0.629
I take into account suggestions from other colleagues.	0.5297	0.7194
The pedagogical proposal is discussed in a team with the participation of teachers from the same grade/ subject.	0.6413	0.5887
In this school, there is a high turnover of teachers.	-0.1436	0.9794
The curriculum is discussed in team with the participation of teachers.	0.5794	0.6643
In this school, I have few opportunities to discuss the pedagogical proposal of my class with the school team.	-0.6159	0.6207
In this school, I have difficulties in sharing my concerns and disappointment.	-0.5948	0.6463
There are too many projects in this school, but I cannot get an overview of them.	-0.4922	0.7578
In this school, a few teachers exchange ideas and experiences in order make everybody learn.	-0.6501	0.5773
In this school, I have few opportunities to discuss ideas about teaching and learning process.	-0.725	0.4744
In this school teachers use the results of external evaluations to evaluate their teaching practices.	0.4912	0.7588
I participate in decisions related to the school.	0.6697	0.5515
The teaching staff takes into consideration my ideas.	0.6798	0.5378
The teaching that the school offers to its students is greatly influenced by the exchange of ideas between the school members.	0.7195	0.4823

Table 6- Teamwork-KMO

TEAMWORK INDEX- FACTOR ADEQUACY	
Kaiser-Meyer-Olkin measure of sampling adequacy	
The political-pedagogical project of this school is a result of the exchange of ideas between teachers	0.9586
Most teachers are committed to improving their classes.	0.9349
Few teachers take responsibility for improving the school.	0.8536
Most teachers maintain high expectations about the learning of their students	0.939
Few teachers are willing to take on new charges to improve the school	0.8561
Most teachers are receptive to the implementation of new ideas.	0.9617
Most teachers are overloaded with work, which harms lessons planning	0.9518
Most teachers feel responsible for student performance.	0.9454
I take into account suggestions from other colleagues.	0.9484
The pedagogical proposal is discussed in a team with the participation of teachers from the same grade/ subject.	0.9421
In this school, there is a high turnover of teachers.	0.8589
The curriculum is discussed in team with the participation of teachers.	0.9388
In this school, I have few opportunities to discuss the pedagogical proposal of my class with the school team.	0.9347
In this school, I have difficulties in sharing my concerns and disappointment.	0.9309
There are too many projects in this school, but I cannot get an overview of them.	0.9565
In this school, a few teachers exchange ideas and experiences in order make everybody learn.	0.9361
In this school, I have few opportunities to discuss ideas about teaching and learning process.	0.9264
In this school teachers use the results of external evaluations to evaluate their teaching practices.	0.9648
I participate in decisions related to the school.	0.9322
The teaching staff takes into consideration my ideas.	0.9267
The teaching that the school offers to its students is greatly influenced by the exchange of ideas between the school members.	0.955
Overall	0.9356

Figure 14- Teamwork-Screeplot



7.3.3.Trust

Table 7- Trust-factor loadings

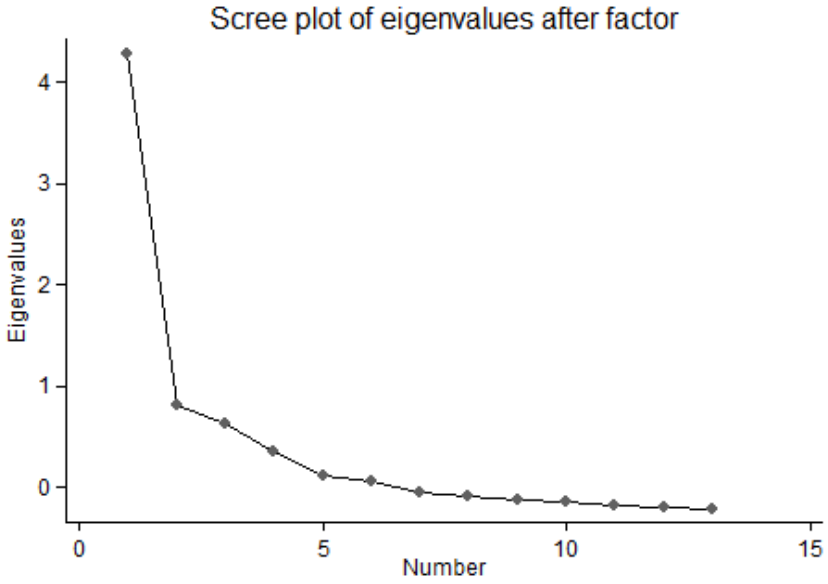
TRUST INDEX
FACTOR LOADINGS AFTER OBLIQUE ROTATION

	Loadings	Uniqueness
The staff at the school are willing to help their colleagues.	0.7275	0.4707
The staff at the school have the same opinion about what is right and wrong	0.5732	0.6714
The staff at the school are people I can trust.	0.7872	0.3803
If I needed to borrow R$ 30.00 in an emergency, I could ask someone from this school	0.7042	0.5041
I am a person in whom others can trust.	0.557	0.6898
If someone from the school staff needed to borrow $ 30.00 in an emergency, he could borrow.	0.6606	0.5636
You can trust most of the people in your community.	0.4838	0.7659
You can trust most of your school staff.	0.7751	0.3992
Most of the school staff would try to take advantage of you if they had the chance	-0.5483	0.6994
Most of the time, people are only mainly concerned with themselves	-0.5064	0.7436
It is possible to be successful on my own, I do not need a large group of people helping each other	-0.2897	0.9161
Having money is important to be happy.	-0.1317	0.9827
People who strive working end up, usually, in a better situation	0.2628	0.9309

Table 8- Trust-KMO

TRUST INDEX FACTOR - ADEQUACY	
Kaiser-Meyer-Olkin measure of sampling adequacy	
The staff at the school are willing to help their colleagues.	0.9000
The staff at the school have the same opinion about what is right and wrong	0.8650
The staff at the school are people I can trust.	0.8799
If I needed to borrow R$ 30.00 in an emergency, I could ask someone from this school	0.8657
I am a person in whom others can trust.	0.8691
If someone from the school staff needed to borrow $ 30.00 in an emergency, he could borrow.	0.8034
You can trust most of the people in your community.	0.8109
You can trust most of your school staff.	0.8591
Most of the school staff would try to take advantage of you if they had the chance	0.8336
Most of the time, people are only mainly concerned with themselves	0.8176
It is possible to be successful on my own, I do not need a large group of people helping each other	0.8123
Having money is important to be happy.	0.7050
People who strive working end up, usually, in a better situation	0.8954
Overall	0.8528

Figure 15- Trust-screeplot



7.4.Student behavior

Table 9- Student behavior-factor loadings

	FACTOR LOADINGS AFTER OBLIQUE ROTATION	
	Loadings	Uniqueness
Learning is hampered by irregular attendance of students.	0.3726	0.8611
With all the attractions that students can access today is very difficult for the school to make its work	0.4461	0.801
Learning is hindered by a lack of support from the parents to the student.	0.4385	0.8077
In this school, I have a suitable pedagogical support.	-0.465	0.7838
Learning is hindered by a lack of discipline of students.	0.4221	0.8219
Given the conditions of this school, any student learning is already satisfactory	0.3358	0.8872
Student learning is hampered by the poor condition of school facilities	0.5372	0.7114
This school lacks teachers for some subjects.	0.4551	0.7929
This school lacks administrative support staff.	0.4868	0.763
The educational resources necessary to my work are available in this school	-0.4481	0.7992
In this school, there is enough pedagogical support staff to help me working	-0.4787	0.7709
For students of this school to learn, it is necessary that the school has more pedagogical resources	0.3987	0.841
Learning is hindered by a lack of skills and abilities of students.	0.4222	0.8217
Learning is hindered by a lack of textbooks.	0.4186	0.8248
Student learning is compromised by the lack of family support.	0.4333	0.8122
Students of this school have no desire to learn.	0.5399	0.7085
Learning is hindered by a lack of interest and effort by the students.	0.4531	0.7947
I can do very little for my students because they come from disadvantaged families	0.4263	0.8182
My students do not do the duties.	0.4772	0.7722
The pedagogical coordination of this school helps in my activities.	-0.4183	0.825
In this school, there is theft and vandalism.	0.4024	0.838
In this school, there are enough computers for the activities of the students.	-0.2207	0.9513
In this school, there are enough computers for the use of teachers.	-0.2698	0.9272
Student learning is hindered by the failure to comply with the curriculum	0.4665	0.7823
In this school, there is too much paperwork.	0.4149	0.8279
The learning of my students is hampered by inadequate curriculum	0.5317	0.7173

Notes: In all our regressions, we multiply the index of student behavior by -1, so that ratings of positive sentences (As “The educational resources necessary to my work are available in this school”) indicate a higher student behavior index.

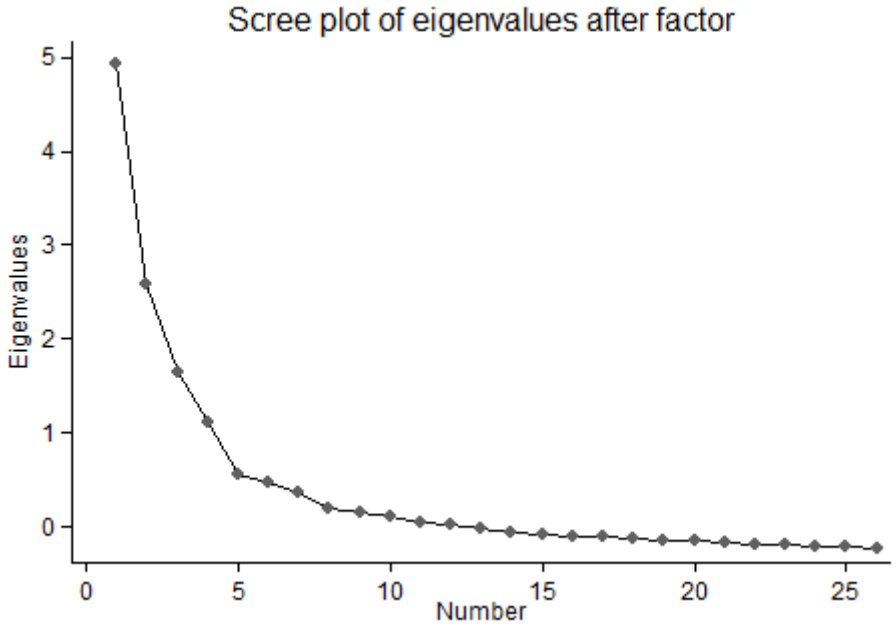
Table 10- Student behavior- KMO

STUDENT BEHAVIOR
FACTOR - ADEQUACY

Kaiser-Meyer-Olkin measure of sampling adequacy

Learning is hampered by irregular attendance of students.	0.8945
With all the attractions that students can access today is very difficult for the school to make its work	0.9283
Learning is hindered by a lack of support from the parents to the student.	0.8026
In this school, I have a suitable pedagogical support.	0.8402
Learning is hindered by a lack of discipline of students.	0.9064
Given the conditions of this school, any student learning is already satisfactory	0.8322
Student learning is hampered by the poor condition of school facilities	0.8968
This school lacks teachers for some subjects.	0.8424
This school lacks administrative support staff.	0.8312
The educational resources necessary to my work are available in this school	0.8843
In this school, there is enough pedagogical support staff to help me working	0.8409
For students of this school to learn, it is necessary that the school has more pedagogical resources	0.8561
Learning is hindered by a lack of skills and abilities of students.	0.8993
Learning is hindered by a lack of textbooks.	0.9094
Student learning is compromised by the lack of family support.	0.8273
Students of this school have no desire to learn.	0.8675
Learning is hindered by a lack of interest and effort by the students.	0.8566
I can do very little for my students because they come from disadvantaged families	0.892
My students do not do the duties.	0.9141
The pedagogical coordination of this school helps in my activities.	0.8322
In this school there is theft and vandalism.	0.9266
In this school there are enough computers for the activities of the students.	0.6206
In this school there are enough computers for the use of teachers.	0.6396
Student learning is hindered by the failure to comply with the curriculum	0.8545
In this school there is too much paperwork.	0.8967
The learning of my students is hampered by inadequate curriculum	0.8479
Overall	0.8453

Figure 16- Student behavior-screeplot



7.5.Perception about standardized tests

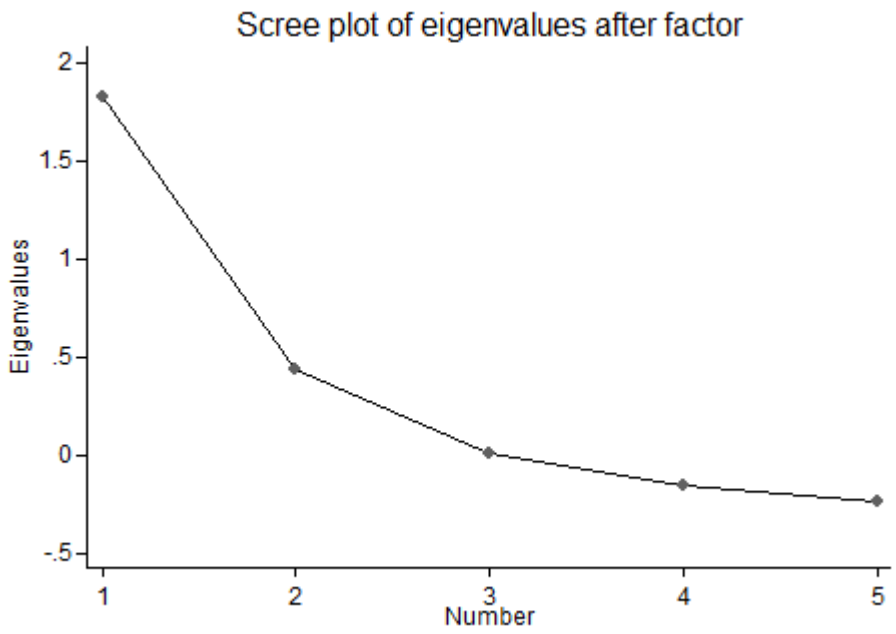
Table 11-Perception about standardized tests- factor loadings

PERCEPTION ABOUT STANDARDIZED TEST FACTOR LOADINGS AFTER OBLIQUE ROTATION	Loadings	Uniqueness
I use the results of evaluations to review my teaching practices.	0.6922	0.5208
Teachers from other subjects, beyond Language and Math, think that the results of large scale evaluations are relevant	0.4634	0.7852
I consider standardized tests unnecessary because I know well my students.	-0.2204	0.9514
In this school teachers use the results of evaluations to review their teaching practices	0.7895	0.3767
Discussing the results of large-scale assessments helps to reflect on my work inside the classroom	0.7192	0.4828
Teachers of other subjects beyond Language and Math, use the results of evaluations to review their teaching practices.	0.7722	0.4038
The results of external evaluations have contributed to improve student performance in this school	0.7322	0.464
The results of external evaluations are used to review the political pedagogical project of this school	0.6723	0.5479

Table 12--Perception about standardized tests- KMO

PERCEPTIONS ABOUT STANDARDIZED TEST FACTOR - ADEQUACY	
Kaiser-Meyer-Olkin measure of sampling adequacy	
I use the results of evaluations to review my teaching practices.	0.8714
Teachers from other subjects, beyond Language and Math, think that the results of large scale evaluations are relevant	0.8945
I consider standardized tests unnecessary because I know well my students.	0.8451
In this school teachers use the results of evaluations to review their teaching practices	0.8544
Discussing the results of large-scale assessments helps to reflect on my work inside the classroom	0.9019
Teachers of other subjects beyond Language and Math, use the results of evaluations to review their teaching practices.	0.8652
The results of external evaluations have contributed to improve student performance in this school	0.8416
The results of external evaluations are used to review the political pedagogical project of this school	0.8374
Overall	0.8635

Figure 17--Perception about standardized tests- screeplot



7.6.Teaching effort/Pedagogical practices

Table 13-Teacher effort- factor loadings

TEACHING EFFORT/PEDAGOGICAL PRACTICES
FACTOR LOADINGS AFTER OBLIQUE ROTATION

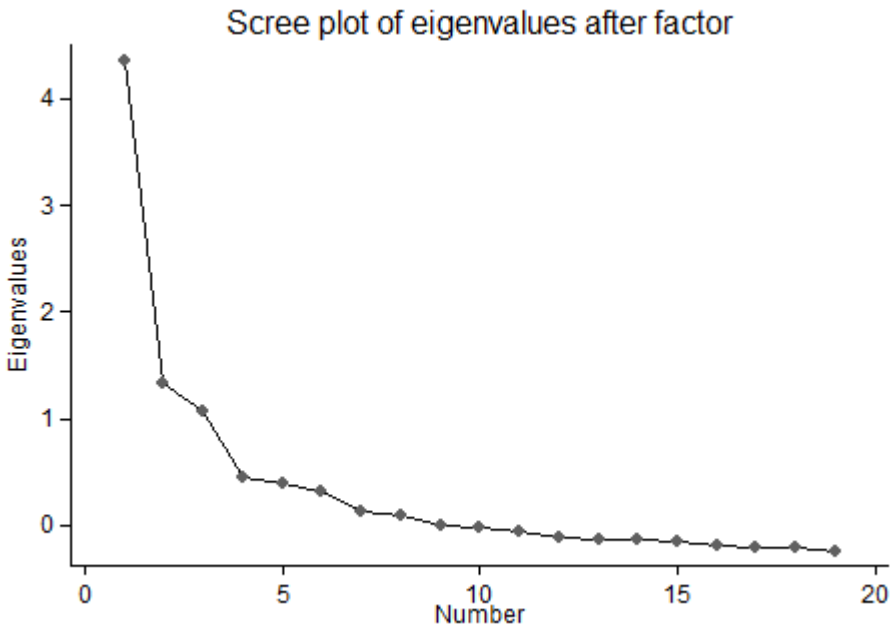
	Loadings	Uniqueness
The textbook is essential for planning lessons.	0.348	0.8789
I have freedom to prepare the lesson plan.	0.4469	0.8003
The education activities I participated in the last two years helped me improving my teaching practice	0.5205	0.729
I prioritize the cognitive aspects when assessing my students' learning	0.5744	0.6701
I prioritize attitudinal aspects in assessing the learning of my student	0.4997	0.7503
I ask home activities from my students.	0.7018	0.5075
I use the textbook only to assign exercises.	-0.2316	0.9464
I do not correct, but I clarify the doubts of students at their home activities.	-0.2666	0.9289
I managed to successfully meet the curriculum this year.	0.4787	0.7709
I use the political pedagogical project of this school as a reference for planning my lessons	0.4405	0.806
I assign homework.	0.6152	0.6215
I correct homework in the classroom with students.	0.5563	0.6905
I correct homework alone and then I give students the results	-0.0555	0.9969
I correct homework alone and then discuss the doubts of students.	0.0099	0.9999
I use computing resources in preparing the lessons.	0.1276	0.9837
I use the textbook in my classes.	0.4156	0.8273
I lose a lot of time with the organization of the class, with roll call and disciplinary warnings	-0.1534	0.9765
I use newspapers and magazines in preparing school lessons.	0.5095	0.7404
I use newspapers and magazines in class.	0.5156	0.7341
I ask home activities from my students.	0.6592	0.5655
I pay extra attention to students with poor performance or more learning difficulties.	0.5496	0.6979

Table 14- Teacher effort-KMO

TEACHING EFFORT/PEDAGOGICAL PRACTICES
FACTOR ADEQUACY

Kaiser-Meyer-Olkin measure of sampling adequacy	
The textbook is essential for planning lessons.	0.8029
I have freedom to prepare the lesson plan.	0.89
The education activities I participated in the last two years helped me improving my teaching practice	0.9181
I prioritize the cognitive aspects when assessing my students' learning	0.811
I prioritize attitudinal aspects in assessing the learning of my student	0.799
I ask home activities from my students.	0.873
I use the textbook only to assign exercises.	0.7392
I do not correct, but I clarify the doubts of students at their home activities.	0.7561
I managed to successfully meet the curriculum this year.	0.9098
I use the political pedagogical project of this school as a reference for planning my lessons	0.8844
I assign homework.	0.8833
I correct homework in the classroom with students.	0.8694
I correct homework alone and then I give students the results	0.5554
I correct homework alone and then discuss the doubts of students.	0.5607
I use computing resources in preparing the lessons.	0.7807
I use the textbook in my classes.	0.8412
I lose a lot of time with the organization of the class, with roll call and disciplinary warnings	0.7914
I use newspapers and magazines in preparing school lessons.	0.6704
I use newspapers and magazines in class.	0.6759
I ask home activities from my students.	0.851
I pay extra attention to students with poor performance or more learning difficulties.	0.9288
Overall	0.7999

Figure 18- Teacher effort-screeplot



7.7.Principal leadership

Table 15-Leadership- factor loadings

PRINCIPAL LEADERSHIP INDEX – SAEPE 2011
SAMPLING ADEQUACY

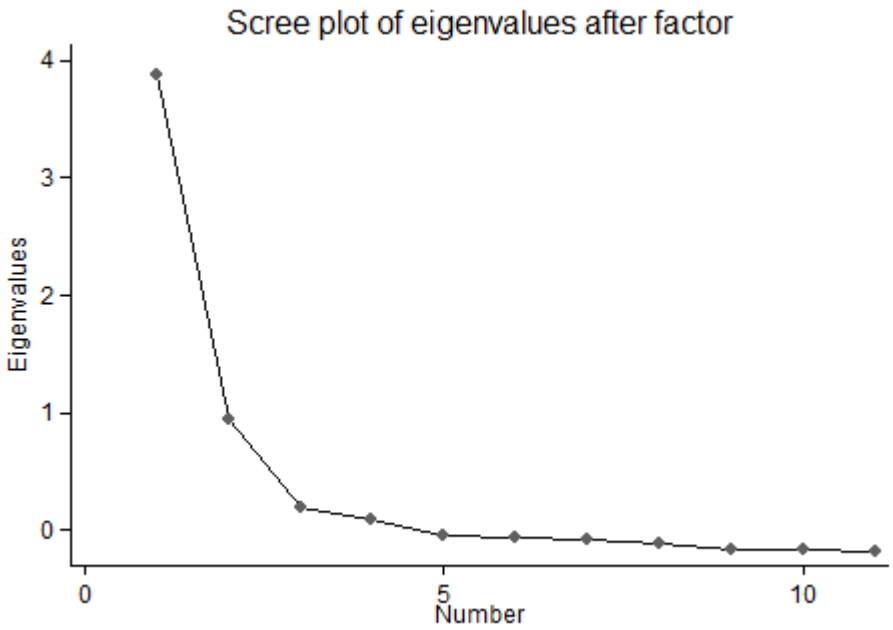
Kaiser-Meyer-Olkin measure of sampling adequacy	
The principal encourages and supports innovative activities.	0.872
I miss school administration support to implement projects	0.855
The director is zealous and demanding in compliance.	0.858
Quarrels and confusion, when they occur, are resolved at the school.	0.908
Everyone knows what can and what cannot be done at school.	0.908
The school is more rigorous when evaluating some teachers than others	0.858
The school principal promotes several important activities	0.906
If the principal had more initiative, the school would be better off	0.861
I participate from decisions about my work	0.887
The principal never did anything about Saepe results	0.935
Overall	0.880

Table 16- Leadership-KMO

PRINCIPAL LEADERSHIP INDEX-SAEPE 2011
FACTOR LOADINGS AFTER OBLIQUE ROTATION

	Loadings	Uniqueness
The principal encourages and supports innovative activities.	0.824	0.322
I miss school administration support to implement projects	-0.559	0.687
The director is zealous and demanding in compliance.	0.753	0.433
Quarrels and confusion, when they occur, are resolved at the school.	0.602	0.638
Everyone knows what can and what cannot be done at school.	0.529	0.721
The school is more rigorous when evaluating some teachers than others	-0.507	0.743
The school principal promotes several important activities	0.497	0.753
If the principal had more initiative, the school would be better off	-0.673	0.547
I participate from decisions about my work	0.503	0.747
The principal never did anything about Saepe results	-0.566	0.680

Figure 19- Leadership-screeplot



7.8.Teacher pedagogical practices-Saepe 2011

Table 17-Teacher effort 2011- factor loadings

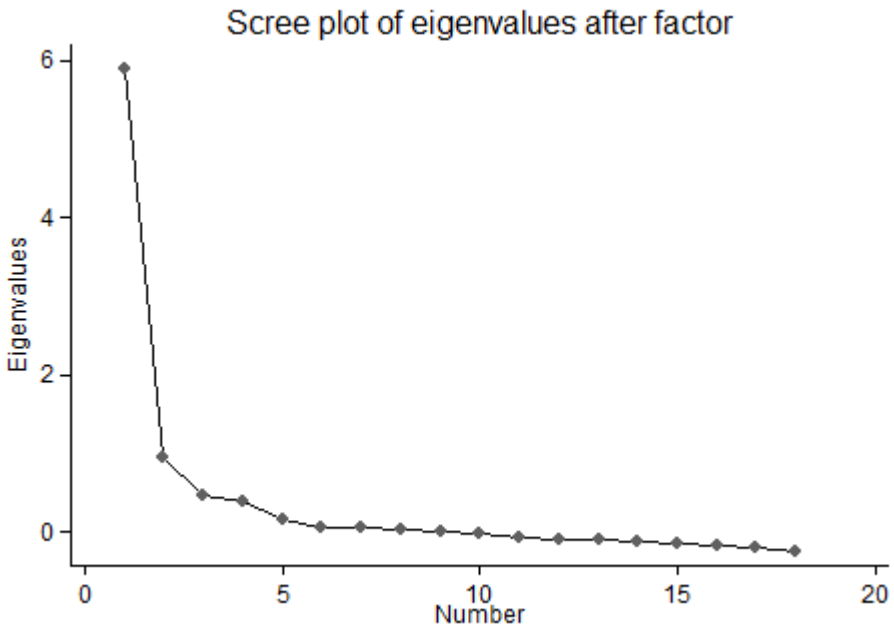
TEACHER PEDAGOGICAL PRACTICES-SAEPE 2011 FACTOR - ADEQUACY	
Kaiser-Meyer-Olkin measure of sampling adequacy	
39. I have to wait a long time for students to make silence.	-0.022
40. One thing will not give up: students pay attention in class	0.050
41. I can build the classroom an atmosphere of order and respect.	0.082
42. I rarely miss classes.	0.028
43. In my classes all have the opportunity to express their views.	0.122
44. I am always available to answer questions from students.	0.162
45. I like when students come to me to help them.	0.195
46. I treat all students equally, without distinction.	0.077
47. With students with special needs I use different teaching resources	0.044
48. I always correct and discuss the homework with students.	0.085
37. I always inform parents about the performance of their children.	0.058
51. I do not accept work badly done by students.	0.029
54. I teach the same subject in various ways, if necessary, to make students learn	0.076
55. I am responsible for student learning.	0.043
57. My classes are always well planned.	0.126
58. I always use the textbook to teach.	0.024
59. I try to use various features to make the most attractive classes.	0.131
60. I try to assign interesting and challenging tasks for my students.	0.092

Table 18 Teacher effort 2011- KMO

TEACHER PEDAGOGICAL PRACTICES-SAEPE 2011
FACTOR LOADINGS AFTER OBLIQUE ROTATION

	Loadings	Uniqueness
39. I have to wait a long time for students to make silence.	-0.199	0.960
40. One thing will not give up: students pay attention in class	0.476	0.774
41. I can build the classroom an atmosphere of order and respect.	0.539	0.709
42. I rarely miss classes.	0.371	0.863
43. In my classes all have the opportunity to express their views.	0.729	0.469
44. I am always available to answer questions from students.	0.796	0.366
45. I like when students come to me to help them.	0.805	0.352
46. I treat all students equally, without distinction.	0.673	0.547
47. With students with special needs I use different teaching resources	0.431	0.815
48. I always correct and discuss the homework with students.	0.655	0.572
37. I always inform parents about the performance of their children.	0.536	0.713
51. I do not accept work badly done by students.	0.200	0.960
54. I teach the same subject in various ways, if necessary, to make students learn	0.633	0.599
55. I am responsible for student learning.	0.461	0.787
57. My classes are always well planned.	0.671	0.550
58. I always use the textbook to teach.	0.237	0.944
59. I try to use various features to make the most attractive classes.	0.679	0.538
60. I try to assign interesting and challenging tasks for my students.	0.632	0.601

Figure 20- Teacher effort 2011- screeplot



7.9.Student satisfaction

Table 19-Student satisfaction- factor loadings

STUDENT SATISFACTION-SAEPE 2011
FACTOR LOADINGS AFTER OBLIQUE ROTATION

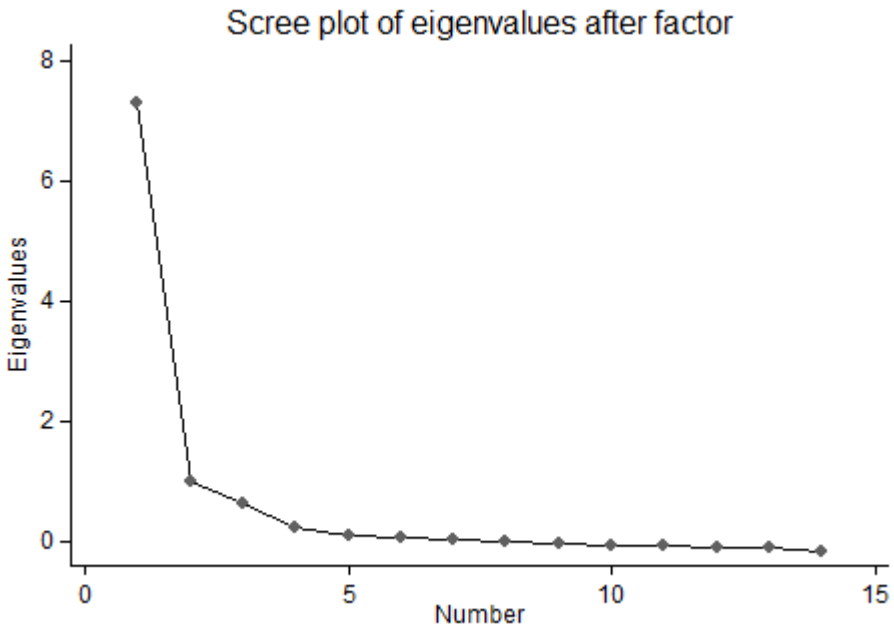
	Loadings	Uniqueness
I think it is worth studying in this school.	0.755	0.430
I'm always learning new things in school.	0.741	0.451
I feel safe in this school.	0.659	0.565
I get along with everyone in this school.	0.565	0.681
I like being with my colleagues.	0.560	0.687
At school all are treated with respect.	0.617	0.619
The school holds parties and events in which everyone participates.	0.646	0.583
I feel well taken care of this school.	0.831	0.309
I feel I am valued in this school.	0.790	0.376
I am proud to be a student of this school.	0.859	0.262
I like to study in this school.	0.848	0.281
I feel full (a) energy and excited (a) at school.	0.759	0.424
I like to go to school.	0.651	0.577

Table 20 – Student satisfaction -KMO

STUDENT SATISFACTION-SAEPE 2011
FACTOR - ADEQUACY

Kaiser-Meyer-Olkin measure of sampling adequacy	
I think it is worth studying in this school.	0.947
I'm always learning new things in school.	0.958
I feel safe in this school.	0.962
I get along with everyone in this school.	0.932
I like being with my colleagues.	0.959
At school all are treated with respect.	0.947
The school holds parties and events in which everyone participates.	0.968
I feel well taken care of this school.	0.946
I feel I am valued in this school.	0.962
I am proud to be a student of this school.	0.937
I like to study in this school.	0.930
I feel full (a) energy and excited (a) at school.	0.962
I like to go to school.	0.955
Overall	0.950

Figure 21- Student satisfaction-screeplot



7.10.Student assessment about teacher practices

Table 21- Student assessment about teachers -factor loadings

STUDENT ASSESSMENT OF TEACHER PRACTICES-SAEPE 2011
FACTOR LOADINGS AFTER OBLIQUE ROTATION

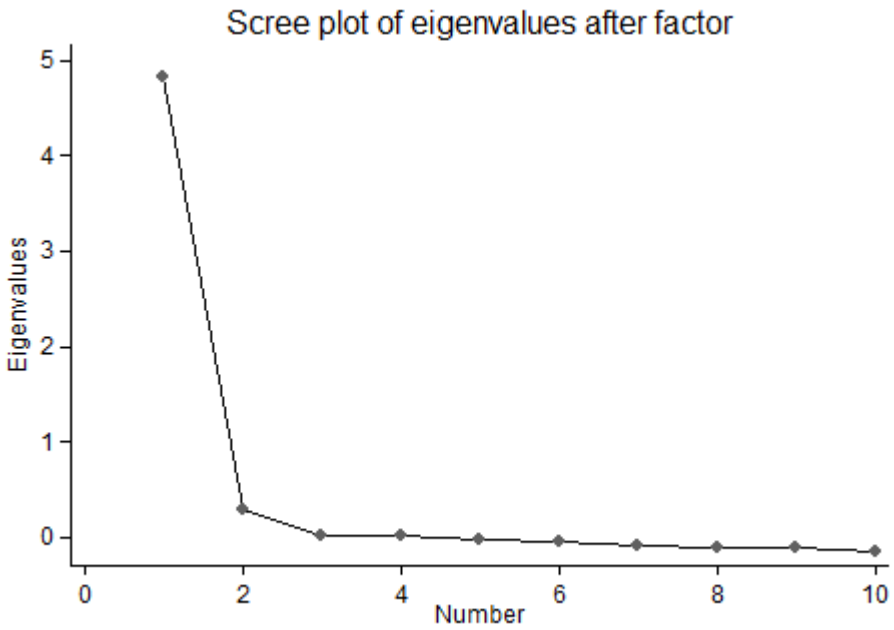
	Loadings	Uniqueness
In class the teacher listens to the claims of students.	0.6903	0.5234
The teacher always clarifies my doubts.	0.7150	0.4888
The teacher helps more some students than others.	-0.0930	0.9914
I learn the material that the teacher teaches.	0.7187	0.4834
The teacher always corrects the homework.	0.7250	0.4744
The teacher explains until everyone understands the lesson.	0.8235	0.3219
For the teacher the whole class can learn.	0.7072	0.4999
The teacher is clear when explaining	0.8154	0.3351
Lessons are interesting and animated	0.7366	0.4574
The teacher always uses textbook to teach.	0.6332	0.5991

Table 22- Student assessment about teachers-KMO

STUDENT ASSESSMENT OF TEACHER PRACTICES-SAEPE 2011
FACTOR - ADEQUACY

Kaiser-Meyer-Olkin measure of sampling adequacy	
In class the teacher listens to the claims of students.	0.9473
The teacher always clarifies my doubts.	0.9352
The teacher helps more some students than others.	0.5927
I learn the material that the teacher teaches.	0.9547
The teacher always corrects the homework.	0.9505
The teacher explains until everyone understands the lesson.	0.936
For the teacher the whole class can learn.	0.956
The teacher is clear when explaining	0.9406
Lessons are interesting and animated	0.9417
The teacher always uses textbook to teach.	0.9456
Overall	0.9427

Figure 22-Student assessment about teachers- screeplot



7.11.Mechanisms

Table 23 – Mechanisms (2009)

MECHANISMS: ESTIMATES OF THE IMPACT OF NOT WINNING THE BONUS (2009)

		RD: Imbens and Kalyanaraman	RD: Cross validation	RD: Half IK bandwidth	RD: Twice IK bandwidth	OLS: Polynomial 3rd order	OLS: Polynomial 2nd order	OLS: Polynomial 1st order
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pedagogical Practices	Estimate:	0.142	0.133	0.245	0.016	0.141	0.108	-0.096
	Std error:	(0.135)	(0.118)	(0.205)	(0.101)	(0.109)	(0.104)	(0.072)
	N:	390	515	177	738	912	912	912
Principal behavior	Estimate:	0.123	0.106	0.104	0.090	0.003	-0.012	0.055
	Std error:	(0.108)	(0.115)	(0.152)	(0.090)	(0.123)	(0.119)	(0.076)
	N:	696	540	291	804	912	912	912
Teamwork	Estimate:	0.026	0.030	0.101	0.059	-0.105	-0.159	-0.053
	Std error:	(0.143)	(0.137)	(0.210)	(0.107)	(0.152)	(0.149)	(0.086)
	N:	502	540	248	778	911	911	911
Trust	Estimate:	0.122	0.126	0.145	0.109	0.129	0.120	0.049
	Std error:	(0.102)	(0.107)	(0.139)	(0.083)	(0.103)	(0.098)	(0.064)
	N:	695	540	284	803	911	911	911
Standardized tests	Estimate:	0.088	0.056	0.050	-0.016	0.042	0.021	-0.123*
	Std error:	(0.136)	(0.115)	(0.204)	(0.096)	(0.120)	(0.117)	(0.070)
	N:	389	539	178	738	911	911	911

Note: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school at a given year. All indices constructed from SAEPE data. All regressions include year dummies and a dummy for full time schools.

Table 24 - Mechanisms (2010)

MECHANISMS: ESTIMATES OF THE IMPACT OF NOT WINNING THE BONUS (2010)

		RD: Imbens and Kalyanaraman (1)	RD: Cross validation (2)	RD: Half IK bandwidth (3)	RD: Twice IK bandwidth (4)	OLS: Polynomial 3rd order (5)	OLS: Polynomial 2nd order (6)	OLS: Polynomial 1st order (7)
Pedagogical Practices	Estimate:	0.381**	0.350**	0.488*	0.215*	0.391**	0.441***	0.195**
	Std error:	(0.181)	(0.153)	(0.252)	(0.120)	(0.161)	(0.158)	(0.095)
	N:	260	328	134	510	912	912	912
Principal behavior	Estimate:	-0.057	-0.030	-0.058	-0.041	0.221	0.253*	0.031
	Std error:	(0.137)	(0.126)	(0.187)	(0.104)	(0.142)	(0.138)	(0.095)
	N:	278	328	144	537	916	916	916
Teamwork	Estimate:	0.064	0.063	0.075	0.027	0.209	0.250	0.054
	Std error:	(0.143)	(0.142)	(0.198)	(0.113)	(0.157)	(0.154)	(0.101)
	N:	324	328	166	601	915	915	915
Trust	Estimate:	-0.112	-0.122	-0.139	-0.091	-0.071	-0.047	-0.080
	Std error:	(0.114)	(0.151)	(0.158)	(0.101)	(0.162)	(0.158)	(0.095)
	N:	578	328	302	808	915	915	915
Standardized tests	Estimate:	0.137	0.107	0.308	0.043	0.278**	0.308**	0.081
	Std error:	(0.168)	(0.141)	(0.248)	(0.115)	(0.141)	(0.138)	(0.088)
	N:	256	328	132	506	915	915	915

Note: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school at a given year. All indices constructed from SAEPE data. All regressions include year dummies and a dummy for full time schools.

Table 25- Mechanisms (2011)

MECHANISMS: ESTIMATES OF THE IMPACT OF NOT WINNING THE BONUS (2011)

		RD: Imbens and Kalyanaraman (1)	RD: Cross validation (2)	RD: Half IK bandwidth (3)	RD: Twice IK bandwidth (4)	OLS: Polynomial 3rd order (5)	OLS: Polynomial 2nd order (6)	OLS: Polynomial 1st order (7)
Pedagogical Practices	Estimate:	0.127	0.145	0.158	0.147	0.071	0.045	0.107
	Std error:	(0.127)	(0.150)	(0.183)	(0.115)	(0.157)	(0.154)	(0.097)
	N:	495	353	250	681	897	897	897
Pedagogical practices reported by students	Estimate:	0.138*	0.131*	0.217*	0.087	0.195**	0.189**	0.040
	Std error:	(0.080)	(0.076)	(0.119)	(0.062)	(0.079)	(0.078)	(0.051)
	N:	326	354	163	565	904	904	904
Student satisfaction	Estimate:	0.157	0.150	0.260*	0.063	0.168*	0.156	-0.017
	Std error:	(0.099)	(0.097)	(0.135)	(0.079)	(0.097)	(0.095)	(0.063)
	N:	339	354	176	590	904	904	904
Principal behavior	Estimate:	0.034	-0.020	-0.056	0.058	-0.016	0.034	0.166*
	Std error:	(0.127)	(0.149)	(0.175)	(0.112)	(0.137)	(0.134)	(0.095)
	N:	510	354	262	705	904	904	904
Standardized tests	Estimate:	-0.062	-0.059	-0.155	0.026	-0.078	-0.091	0.135
	Std error:	(0.142)	(0.135)	(0.207)	(0.105)	(0.129)	(0.127)	(0.085)
	N:	326	354	163	566	904	904	904

Note: Table presents regression discontinuity estimates of the effect of not winning the bonus. Each observation is one school at a given year. All indices constructed from SAEPE data. All regressions include year dummies and a dummy for full time schools.