FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO

LYCIA SILVA E LIMA

ESSAYS ON HUMAN CAPITAL INVESTMENTS IN BRAZIL

São Paulo
2019

LYCIA SILVA E LIMA

ESSAYS ON HUMAN CAPITAL INVESTMENTS IN BRAZIL

Tese de doutorado apresentada à Escola de Administração de Empresas de São Paulo da Fundação Getulio Vargas, como requisito parcial para a obtenção de título de Doutora em Administração Pública e Governo.

Linha de Pesquisa: Política e Economia do Setor Público

Orientador: Prof. Dr. Ciro Biderman

São Paulo

2019

Lima, Lycia Silva e.
Essays on human capital investments in Brazil / Lycia Silva e Lima. - 2019.
128 f.

Orientador: Ciro Biderman.
Tese (doutorado CDAPG) – Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo.

1. Educação - Brasil. 2. Capital humano. 3. Qualificações profissionais. 4. Mercado de trabalho - Brasil. I. Biderman, Ciro. II. Tese (doutorado CDAPG) – Escola de Administração de Empresas de São Paulo. III. Fundação Getulio Vargas. IV. Título.

CDU 37(81)

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

LYCIA SILVA E LIMA

ESSAYS ON HUMAN CAPITAL INVESTMENTS IN BRAZIL

Tese apresentada ao Programa de Doutorado Escola de Administração de Empresas de São Paulo da Fundação Getulio Vargas, na Linha de Política e Economia do Setor Público.

Campo de Conhecimento: Microeconomia Aplicada, Capital Humano
Linha de Pesquisa: Política e Economia do Setor Público

Data:
15/05/2019

Banca examinadora:

Prof. Dr. Ciro Biderman (Orientador)
FGV EAESP

Prof. Dr. André Portela Souza
FGV EESP

Prof. Dr. Bruno Ferman
FGV EESP

Prof. Dr. Ricardo Paes de Barros
INSPER

AGRADECIMENTOS

"Digo: o real não está na saída nem na chegada: ele se dispõe para a gente é no meio da travessia" - Guimarães Rosa

O que eu construí ao longo desse doutorado vai muito além da tese. Grande parte das minhas conquistas eu devo a algumas pessoas especiais com as quais eu tive a sorte de cruzar nesse caminho, tornando tudo mais leve, mais inspirador e mais feliz. O primeiro muito obrigada vai para o Ricardo Paes de Barros, a minha maior inspiração desde o começo. É uma honra ter você na minha vida, a minha gratidão é infinita. Sem você eu não estaria aqui.

Ao meu orientador, Ciro Biderman, pela genialidade, pela parceria, pela energia e por estar sempre on board com as minhas ideias. Sorte a minha, você foi muito mais que um orientador. Ao Andre Portela, por tudo que construimos juntos ao longo desses anos e por me proporcionar o melhor trabalho do mundo. Tudo que aprendi com você foi fundamental para o meu desenvolvimento como pesquisadora. Obrigada pela leveza, pela amizade e pelo apoio incondicional.

Ao Bruno Ferman, pela oportunidade de construir juntos, que tanto me agregou e enriqueceu nos últimos anos e, em especial, ao longo da tese. Ao Pedro Carneiro, prova viva que genialidade e humildade caminham juntos, por me receber na UCL e por tantas horas pensando juntos.

Aos colegas de PESP, amigos da FGV pra vida toda: foi um verdadeiro presente ter vocês ao longo desses anos. Em especial, ao Arthur Fisch, Adriano Borges, Claudinha Oshiro, Patricia Alencar, por me levantar toda vez que eu caí, e Marianna Sampaio, por me fazer ter certeza de que estou do lado certo. À Lara, Leozinho e aos demais queridos colegas do CEPESP, que sempre fazem com que eu me sinta querida e em casa. E ao George Avelino, pela doçura, e por me ensinar a olhar pra fora da caixa. Vejo o mundo com outros olhos depois de conhecer você.

Eu sou eternamente grata pela generosidade de algumas pessoas especiais que dedicaram muito tempo me ajudando de várias formas. Juliana Camargo, amiga, irmã, braço direito, companhia de todas as horas. Não cabe nesta página o tamanho da minha gratidão a você. Gabriel Weber, pela excelência e boa vontade com as quais você faz tudo, pela revisão impecável, e por deixar tudo assim tão lindo. À Dani Stucchi, por me ajudar a deixar tudo sob controle e por tantas conversas que me fazem colocar a vida em perspectiva. À Thais Dietrich, pela cereja do bolo. Ao Samuel e à Andrezza e, em especial, à querida Camilinha Mata Machado, por me ajudarem sempre com tanta boa vontade. Ao Lucas Finamor, meu menino prodígio, por tanto desde sempre, pela construção conjunta e por estar perto mesmo longe. E ao Flavio Riva, pela parceria, por me contaminar com sua empolgação e com sua forma de pensar o mundo e por todas as intervenções brilhantes ao longo desses anos.

Ao Pedro Olinto, por não me deixar desistir nunca e por me fazer acreditar que tudo é possível. Ao Afonso Henriques Borges Ferreira, minha gratidão eterna pelo apoio desde a graduação. E ao Alvaro Villalobos, por continuar sendo família.

Aos outros tantos queridos, atuais ou alumni do CLEAR: Dalila, minha minion amada, e aos queridos Aline, Sammara, Eduardo Cenci, Tatiana Sandim, Matheus, e tantos outros, pela amizade sincera e valiosa. Agradeço a todos vocês por tudo que fizemos e vivemos juntos. Trabalhar com vocês ao longo

desses anos fez trabalho parecer férias. Em especial, à Amandinha Arabage, pela convivência leve e pelo carinho; e ao Tomas Malaga, por alegrar nosso ambiente com a sua existência. E à Patricia dos Anjos, minha ídola, pela amizade que torna a minha vida mais feliz e pela competência que torna a minha vida muito mais fácil.

Também o meu muito obrigada àqueles amigos que estiveram sempre presentes e fizeram a vida ser mais doce e divertida ao longo desses anos. Aos queridos Ronan Cunha, Fabio Morosini, Federica Iorio, Thiago de Lucena, Thais Herdy, Clarissa Huguet, Joana Hardy e Guilherme Marinho. Em especial à Julia Mata Machado e à Melissa Menezes, pela fortaleza da nossa rede que é responsável por eu nunca me sentir desamparada; à Mariana Andrade, pelo exemplo de coragem e equilíbrio e por estar sempre ao meu lado, mesmo longe; e à minha grande amiga Amanda Schutze, por todo o apoio e por tanto mais. Não só o doutorado mas a vida é mais fácil com você por perto.

Ao professor Daniel Hidalgo e ao grupo Gov/Lab por ter me recebido durante meu período no MIT. Ao CGIS de Harvard, e ao Harvard-MIT Data Lab, home away from home. E aos amigos de Cambridge, em especial ao Danilo Limoeiro, Rafael Mouallem, João Dabbur, Natalie Catlett, Laura Oller, Julie Ricard, Candice Vianna, Maira Machado, Joao Abreu e Tarcila Reis, que fizeram Boston ser inesquecível.

À Fundação Getulio Vargas por ter me proporcionado tanto. Ao Departamento de Administração Pública e Governo e à Escola de Economia de São Paulo, em especial, ao Ronaldo Toniete, por todo o apoio e por todas as risadas.

À Marcia, por me ajudar na construção da base de tudo.

Um agradecimento especial à minha família, por me lembrar onde é meu lugar no mundo. À minha irmã Thais Boaventura, meu maior presente e meu maior orgulho, por alegrar a minha existência diariamente, sendo a melhor amiga e irmã que alguém pode ter. Ao meu irmão, Otavio Boaventura, pelo exemplo de persistência, por todas as risadas, e por me fazer ter certeza que eu sempre tenho alguém por mim incondicionalmente. À minha prima-irmã Erika Magalhães, minha grande companheira, por estar sempre ao meu lado independente da distância; e ao Vitor Quinet, o irmão que eu escolhi, obrigada por ter me escolhido também, e por tornar a minha vida mais leve e divertida. À minha madrinha maravilhosa Du, meu exemplo de luz, e à minha avó linda Zulma, nossa fortaleza, e aos meus tios e primos amados.

E, por fim, minha homenagem à minha mãe e ao meu pai, meu porto seguro, por me ensinarem o valor da educação e por me motivarem a correr sempre atrás dos meus sonhos. Tudo que eu consegui na vida eu devo a vocês. À minha mãe, Valeria, a mulher mais forte e destemida que eu já conheci, que me ensinou que ninguém é melhor que ninguém, e com quem eu aprendi a me indignar com os absurdos do mundo. É ela que me salva quando eu preciso, e foi seguindo o exemplo dela que eu aprendi que resistir e questionar sempre é o que dá sentido à vida. Foi vendo o mundo através dos seus olhos que eu decidi seguir esse caminho e trabalhar com algo que pudesse contribuir para melhorar a vida das pessoas. Nos momentos de extremo cansaço, ela sempre me motivava dizendo que *"tudo na vida é difícil, é assim mesmo, e se doutorado fosse fácil todo mundo tinha"*. E ao meu pai, Aldemar, com quem eu aprendi a não ter medo de sair pelo mundo pra crescer e progredir, que me ensinou o valor do trabalho, que lutou e se sacrificou pra me dar a melhor educação possível, por ser meu exemplo de bondade e resiliência, e por dizer sempre, nos dias mais difíceis, que *"se não der, tudo bem, filhinha"*. Este título é para vocês.

ABSTRACT

Essay 1 - Public Childcare, Child Development and Labor Market Outcomes: Experimental Evidence from Rio de Janeiro, Brazil. This paper describes the short and medium term impacts of free public childcare provision in Rio de Janeiro using a lottery that allocated access to the program in 2008. Lottery winners had large gains in anthropometrics measures which persisted throughout time. They also had higher scores in cognitive function tests, which had faded out 7 years after the lottery took place. The program also had significant impacts on household income and employment outcomes of family members.

Key-words: Early childhood development, childcare, public childcare, daycare, RCT.

Essay 2 - Are Public Schools Ready to Integrate Math Classes with Khan Academy? We study the impacts of the program *Khan Academy in Schools* using a randomized control trial in Brazilian primary public schools. Once a week, teachers would take their students to the school's computer lab and teach using the Khan Academy platform, instead of their standard math classes. We find positive effects of the program on measures of attitudes towards math, which were not translated to a positive average treatment effect on students' math proficiency. We also explore treatment heterogeneity by quality of implementation, suggesting that the program can have positive effects when there are no infrastructure problems and when the implementation modality is based on one computer per student.

Key-words: Ed-tech, computer-aided-learning, education, RCT.

Essay 3 - Human Capital Investments Among Vulnerable Youth in Rio de Janeiro: Experimental Evidence from the* Protejo *Program. This paper presents experimental evidence to describe the impacts of the Project of Youth Protection in Vulnerable Territories (Protejo), implemented in Rio de Janeiro under the coordination of the Ministry of Justice with the purpose of fostering cognitive and noncognitive skills of youth participants to promote education and employment and, ultimately, reduce crime and victimization. Our results indicate that, although we find no average effects on education or noncognitive skills two years after the program, the intervention was successful in increasing beneficiaries' probability of formal employment from 3 years after the end of the program onwards.

Key-words: Human capital, vocational training, skill formation, RCT.

RESUMO

Ensaio 1 - Creches Públicas, Desenvolvimento Infantil e Mercado de Trabalho: Evidência Experimental do Rio de Janeiro, Brasil. Este artigo descreve os resultados de curto e médio prazos das creches municipais no Rio de Janeiro a partir de uma loteria que selecionou os beneficiários do programa em 2008 de forma experimental. Crianças que ganharam a loteria tiveram ganhos nas variáveis antropométricas, que persistiram ao longo do tempo. Foram também observados ganhos em funções cognitivas, que desapareceram 7 anos após a loteria. O programa também teve impactos significativos sobre renda das famílias e variáveis de mercado de trabalho de membros da família.

Palavras-chave: Creches, primeira infância, avalição experimental.

Ensaio 2 - As Escolas Estão Prontas para Integrar Aulas de Matemática com Khan Academy? Este artigo estuda o impacto do programa *Khan Academy nas Escolas* em escolas públicas de ensino fundamental a partir de evidência experimental. Uma vez por semana, professores levavam os alunos ao laboratório de computador para uma aula usando o Khan Academy. Nossos resultados mostram que o programa teve impacto positivo sobre atitudes com relação à matemática, que não se traduziu em ganhos em proficiência em matemática. Nós também exploramos heterogeneidade no tratamento por qualidade da implementação, e os resultados sugerem que o programa pode ter efeitos positivos quando não há problemas com a infraestrutura e quando a modalidade de implementação é baseada na modalidade um aluno por computador.

Palavras-chave: Educação, tecnologia, avaliação experimental.

Ensaio 3 - Investimentos em Capital Humano de Jovens Vulneráveis no Rio de Janeiro: Evidência Experimental do Programa Protejo. Este artigo estuda, a partir de evidência experimental, os impactos do programa Protejo (Projeto de Protecao de Jovens em Territorios Vulneraveis), coordenado pelo Ministério de Justiça e implementado no Rio de Janeiro em áreas de vulnerabilidade social. O programa tinha como objetivo aperfeiçoar as habilidades cognitivas e não-cognitivas dos jovens, para aumentar escolaridade, empregabilidade, e reduzir a vitimização e criminalidade. Nossos resultados indicam que, apesar de não encontrarmos impacto sobre educação ou habilidades não cognitivas, a intervenção teve impacto significativo sobre a probabilidade de estar empregado no setor formal da economia de 3 anos após o fim do programa em diante.

Palavras-chave: Capital humano, educação vocacional, habilidades, avaliação experimental.

List of Figures

List of Tables

TABLE OF CONTENTS

1 Public Childcare, Child Development and Labor Market Outcomes: Experimental Evidence from Rio de Janeiro, Brazil

1.1 Introduction

A large body of literature shows that early childhood shocks are decisive for shaping human capital of children, impacting outcomes which are determinant for their welfare at later stages in life. Empirical research findings show that interventions before the age of five can have substantial impacts on a variety of outcomes, ranging from cognitive and noncognitive skills to schooling, labor productivity, earnings and health conditions, among many others, such as criminal behavior and teenage pregnancy (CURRIE; ALMOND, 2011; HECKMAN, 2008; KNUDSEN et al., 2006). Although policies to promote child development can be extremely costly, early life cycle interventions seem to have in general much higher economic returns than human capital investment at later stages, often generating high cost benefit ratios (CARNEIRO, 2003; CUNHA et al., 2006; HECKMAN, 2008).

Early childhood development initiatives (*henceforth* ECD) may combine interventions targeting different dimensions of child development, including activities that enhance nutrition and health, foster favorable social and home environments and promote cognitive and noncognitive stimulation. ECD services may be offered in a variety of settings: home-based, center-based, formal preschool and parent/community-based arrangements (LEROY et al., 2011). Public provision of subsidised ECD services is not uncommon, being particularly relevant in the context of developing countries, where poverty is more pronounced and hence more children are lagging behind by facing several dimensions of deprivation. A very common setting for the provision of children care in many different countries is daycare centers, which may offer a variety of ECD services, from basic childcare to integrated nutritional, health and educational services targeting child development through many different channels. Many governments, in both developed and developing countries, subsidize daycare centers primarily as an initiative to foster labor market participation of the child carer, aiming at enabling families to be economically self-sufficient. Therefore, it is not surprising that public childcare programs potentially impact both children as well as family outcomes.

In this paper, we study the short and medium-term impacts of a large-scale public daycare program in Rio de Janeiro (*creches*) over a wide range of family and child development dimensions. In conjunction with the Secretariat of Education of Rio de Janeiro, in 2007 we conducted a lottery to assign approximately 10,000 slots among 24,000 applicants (children 0-3 years old) to public daycare for the academic year of 2008. Our results show that winning the lottery led to an increase in *creche* attendance of approximately one semester, and had strong impacts on household income in 2008 and 2012, which faded out by 2015. Large short run effects on several labor market outcomes were also found, particularly meaningful for grandparents and siblings, but most differences between lottery winners and losers faded out by 2015, seven years after the lottery was carried out.

In terms of child development, our results indicate substantial effects on anthropometrics measures that persisted throughout the years. Gains in cognitive development of children were also registered for lottery winners. For 2012, we find significant impacts on our aggregate measure of cognitive development,

driven by the strong effects on vocabulary development. For 2015, we find positive effects on the perceptual reasoning dimension of the IQ index (which reflects children's ability to examine a problem, organize their thoughts and create solutions), although no effect was captured by the aggregate IQ score. Our mediation analysis findings suggest that improvements in cognitive development of children are largely explained by the positive impacts daycare attendance had on household income and home environments. On the other hand, the strong effects on anthropometrics measures seem to be unrelated to the improvements in family outcomes, and are likely to be mostly explained by an enhancement of children's nutritional intake while being serviced at daycare centers.

Our findings contribute to the existing empirical evidence on the effects of subsidized childcare programs, which is far from conclusive. For the labor market dimension, most studies for developed countries seem to converge to positive impacts (HERBST, 2017; LEFEBVRE; MERRIGAN, 2005; BAKER et al., 2008; ROTHBART et al., 2001; NOLLENBERGER; RODRÍGUEZ-PLANAS, 2015). However, the evidence for developing countries is mixed, and different studies have found both positive (BERLINSKI et al., 2011; BARROS et al., 2011) and no impacts on labor market outcomes (MANLEY et al., 2013; VERA, 2009).

The literature investigating the effects of daycare on child development also registers heterogeneous findings. For developed countries, subsidized childcare programs have been found mostly beneficial but sometimes harmful for children's development (BLACK et al., 2014; HAVNES; MOGSTAD, 2011; HERBST, 2017; BAKER et al., 2008). For developing countries, the existing evidence is very limited and largely relies on non-experimental research that mostly investigates the impacts of community-based childcare centers. There is no consensus on whether these are beneficial for children's development, as different studies have reached diverging conclusions (ATTANASIO et al., 2013; ATTANASIO; VERA-HERNANDEZ, 2004; ANGELES et al., 2014). To the best of our knowledge, only two studies investigated the effects of institutional daycare in developing countries, both focused on pre-school programs that target children 3-5 years old (BERLINSKI et al., 2011; BERLINSKI et al., 2008). These find positive effects on school performance and attendance, but do not present evidence on measures of child development. In this context, this paper is of particular relevance, as it is the first study to generate experimental evidence on the effects of a large-scale public daycare program (targeting children 0-3) in a middle-income country over child development and labor market outcomes of all household members.

The remainder of this paper is organized as follows. Section 1.2 discusses the relationship between daycare, child development and labor market outcomes and reviews the existing empirical literature. Section 1.3 describes the context of the study and the experimental design. Section 1.4 describes the sample selection and data sources. Section 1.5 presents the empirical strategy, while in section 1.6 we describe the results on main outcomes. Section 1.7 attempts to discuss the channels through which daycare may be impacting child development, based on a mediation analysis. Section 1.8 discusses the results and section 1.9 concludes with a summary of our findings and potential policy implications.

1.2 Related Literature: Daycare, Child Development and Labor Market Outcomes

Very popular in middle-income countries, most government subsidised daycare programs are launched primarily as initiatives to foster women's labor market participation, enabling families to be

economically self-sufficient. Therefore, not surprisingly, there is a vast literature investigating the effects of daycare on labor market outcomes of the carer (MORRISSEY, ; BLAU; CURRIE, 2006). However, a large body of literature has also dedicated efforts to understanding the effects of center-based childcare on the development of children (BLAU; CURRIE, 2006; CURRIE; ALMOND, 2011; LEROY et al., 2011). In this section, we discuss the channels through which we expect daycare to affect labor market outcomes and child development, and review the existing empirical literature.

1.2.1 Daycare and Labor Market Outcomes

In an attempt to promote female labor force participation, governments in both developed and developing countries in the past decades have invested in childcare subsidies to decrease the cost of maternal employment and foster their labor market participation (MORRISSEY, ; BLAU; CURRIE, 2006). For developed countries, which usually have higher maternal employment rates, several studies have used non-experimental methods to investigate the relationship between subsidized childcare and maternal employment. Positive impacts of childcare subsidies have been found in the United States (HERBST, 2017), Canada (LEFEBVRE; MERRIGAN, 2005; BAKER et al., 2008), Germany (BAUERNSCHUSTER; SCHLOTTER, 2015), Italy (BRILLI et al., 2016), Netherlands (BETTENDORF et al., 2015) and Spain (NOLLENBERGER; RODRÍGUEZ-PLANAS, 2015). Although most of the studies indicate a strong positive relationship, there is some diverging evidence in the literature. For Sweden (BLACK et al., 2014), for instance, childcare subsidies have no effect on labor market participation.

Evidence for developing countries, where female labor force participation is lower, is scanter and more heterogeneous. Berlinski et al. (2011) find positive maternal employment effects of pre-primary school availability in Argentina based on a differences-in-differences identification strategy. Positive effects are also registered for Brazil, in a study by Barros et al. (2011) that relies on the same database used in this study for the year 2008. Positive effects on maternal employment as a result of subsidized childcare was also found for Mexico, in a study by Angeles et al. (2014). In Chile, some studies investigated an expansion of pre-school slots for poor children, and find no positive impact on mother's labor force participation (MANLEY et al., 2013; VERA, 2009). However, a randomized experiment carried out by Martínez and Perticará (2017) show that an afterschool care program for children aged 6-13 substantially increases labor force participation and employment of mothers. While further studies on the impacts of subsidized daycare on labor market outcomes in developing countries are certainly of interest, shifting childcare from mother's responsibility to daycare centers' also has effects on child development that deserves attention, since this is an issue largely underexplored in the literature, especially for the developing country context.

1.2.2 Daycare Programs and Child Development

Daycare programs may impact child development outcomes through a variety of different channels. One the one hand, subsidised childcare provision may act indirectly on child development through the income channel, since it enables carers to work while their children are at daycare, hence enhancing labor market outcomes of the carer and raising household income. Higher household income is likely to be translated into higher living standards, potentially increasing household food security, which may improve children's dietary intake. Higher income may also be translated into parents spending

more on child stimulation and health care. Several experimental and quasi-experimental studies in both developed and developing countries analyze the impacts of different types of income shocks on child development, including both conditional and unconditional cash transfer programs, tax benefits programs, worker displacement shocks, among others. Although the findings in this literature are mixed, with some contributions exhibiting no or very small effects (see, for instance, Shea (2000) and Løken (2010)), the majority of studies suggest that, especially in lower income settings, parents' income has a positive and significant relationship with children's short and long run outcomes - ranging from birthweight, health measures and test-scores to adult labor market outcomes (see Currie and Almond (2011) for a summary and also Macours et al. (2012), Fernald et al. (2008), Oreopoulos et al. (2008)).

Attending daycare may also impact child development directly, either in a positive or negative way. The impacts of daycare attendance on child development outcomes are inextricably linked to the quality of services provided in comparison to the counterfactual environment of the child. In lower income settings, where home environments of disadvantaged families are less favorable, good quality childcare is more likely to have more favorable results. Daycare may boost child development if it increases social interaction, enriches the child's environment and enhances nutritional food intake and health services. Daycare may also have a positive impact on children if it fosters good parenting practices and home environments. Parents may learn from the interactions with care providers how positive attitudes towards the child are important (such as avoiding aggressive attitudes and physical punishment) and how to best care for the children (nutrition and health recommendations and child stimulation best practices such as limited TV time and increased reading and playing time). However, daycare may also have adverse effects, exposing children to communicable diseases from the interaction with other children and also infectious diseases depending on the hygiene and sanitation levels of the centers. Daycare may also have the adverse effect of reducing parents' interaction with the child, and even potentially reducing the amount or quality of food families offer to the child, if they consider children are being fully serviced in those fronts while at the daycare centers (CURRIE; ALMOND, 2011; LEROY et al., 2011).

A number of studies - the majority using data from developed countries - have generated robust evidence on this issue, with heterogeneous findings. Many studies find positive short run effects and, although these are certainly of interest, a crucial policy question is whether these programs are able to deliver long lasting results. One interesting issue uncovered by several empirical studies of ECD interventions is the fact that impacts on outcomes that reflect cognitive abilities - such as test scores - often fade out after a few years. However, an increasing body of evidence shows that early childhood interventions impact also noncognitive skills dimensions, which are just as important for adult life as cognitive skills. One of the most famous cases to illustrate this is the Perry Preschool project, which randomly assigned 58 children to a high quality preschool program in Michigan in the 1960s, and followed them and also the 65 children assigned to control group through adulthood. Although the Perry program initially boosted the IQs of the participants, this effect faded after a few years after the program finished. However, the program showed persistent effects on personality and social skills, and later on several different outcomes such as labor market variables and reduced criminal behavior (HECKMAN et al., 2006; BLAU; CURRIE, 2006; ATTANASIO, 2015). There is extensive evidence on the effects of small scale interventions such as the Perry Preschool program. From the perspective of public policy, however, studies that look at the impacts of large-scale subsidized government programs are of particular interest.

In Norway, childcare subsidies have been growing steadily since the 1970s, as a result of the Kindergarden Act that was passed in 1975, and the subsidies are administered at the municipality level. Some municipalities have a single price for all income groups, while prices vary by family income in other municipalities. Black et al. (2014) use a sharp discontinuity in the price of child care in some municipalities in Norway to study the effects of child care subsidies on children's and parental outcomes, for birth cohorts from 1986 to 1992. Surprisingly, they find no effect of subsidies on childcare utilization or parents' labor market participation, but significantly large effects on children's educational performance in high school, indicating that increments in disposable income induced by the subsidies acted as the mechanism though which child outcomes were improved. Havnes and Mogstad (2011) analyze the longer run impacts of the large-scale expansion of childcare subsidies in Norway, which constituted a massive positive shock to the supply of subsidised formal childcare. Their difference-in-differences identification strategy explore the fact that the size of supply shocks to formal childcare varied among the different areas of the country. Their findings show large positive effects on children's adult life outcomes measured in their early 30's, including labor market attachment, welfare dependency and education outcomes.

Herbst (2017) studies the effects on maternal employment and children's outcomes of the Lanham Act, a near universal and heavily subsidized child care program, implemented in 635 communities in the US during World War II as a temporary war emergency measure to enable women to augment the economy's labor market. Through a triple difference approach and based on census data, the authors explore for identification the substantial cross-state variations in the size of the program. Their results show that the program increases substantially maternal employment, which persisted even after the program was terminated. In terms of child outcomes, their findings show gains in educational attainment and health, employment rates, earnings, and dependency of cash assistance programs, particularly concentrated on more economically disadvantaged individuals. For more advantaged individuals, the program had neutral and even small negative effects in some dimensions.

The available evidence, however, does not always deliver promising results. Baker et al. (2008) study the introduction of universal childcare subsidies in Quebec, Canada, finding a large impact on utilization of daycare services (14 percent, or one third of baseline rate), substantial increase in women employment and negative effects on a wide range of child outcomes, including child's anxiety, aggressiveness, social and motor skills, health status and illness (reported by parents). Detrimental effects of childcare subsidies were also registered on parental relationship quality and parent health. A later study by Baker et al. (2015) show these negative results in noncognitive dimensions persist as children get older, revealing not only worse self-reported health and life satisfaction among teens, but also increased criminal behavior, all particularly pronounced for boys.

In face of such heterogeneity of results, it is a challenge to draw a conclusion on whether childcare should be scaled up into large-scale or universal programs subsidised by governments. A meta-analysis by Huizen and Plantenga (2015) synthesizes empirical evidence from 30 quasi-experimental studies that use data from either Canada, the US or Eastern Europe aiming at explaining the heterogeneity in the effects of universal childcare. Their findings indicate that quality is a crucial determinant for the universal child care impacts on children, the largest gains are among the disadvantaged children's group and overall effects do not fade out over time. The existing literature indicates for large gains for lower income families, although

this conclusion is in general coming from richer countries which are more likely to have higher quality center-based childcare services. Although there is a growing number of contributions on the effects of large scale daycare programs in developed countries, the evidence available for developing countries is more limited.

For developing countries, a systematic review by Leroy et al. (2011) examines the impacts of six daycare programs on health, nutrition and development of children under five years old, all based on non-experimental approaches. The review considers mostly community-based centers, but also includes two studies of institutional daycare, and points out that drawing conclusions from the available evidence is challenging. The effects on child health and nutrition are mixed, but overall the literature indicates positive effects on child development. Attanasio et al. (2013) use an instrumental variable approach to estimate the impact of a community daycare program in Colombia on nutritional status, finding significant positive effects on children's height (0.44σ), while Bernal and Fernández (2013) find positive effects of the same program on cognitive child development (0.15σ to 0.3σ after at least 15 months of program exposure). Attanasio and Vera-Hernandez (2004) also study the long run impacts of the same program, and identify that children that attended community daycare were more likely to be in school at age 13-17. Berlinski et al. (2011) examine the effects of institutional pre-primary education for children aged 3-5 in Argentina and find positive effects on test scores and also noncognitive dimensions related to self-control. Evidence on the impacts of pre-school education to children 4-5 is also available for Uruguay, in a study by Berlinski et al. (2008). Based on IV estimations, the authors find positive effects on years of education for treated children. Angeles et al. (2014) study, based on a non-experimental design, the effects of a childcare subsidies for community-based centers in Mexico, finding no effects on child development and negative effects on child health.

1.3 Background: Context and Study Design

In 2007, the year the lottery explored in this study was carried out, Rio de Janeiro had a population of around 6 million people (7 percent of which were children aged 0-4), which corresponded to roughly 40 percent of the state's population and 3.5 percent of entire Brazilian population.[1] Although the Brazilian constitution grants every child the right to education and states free access to daycare as a fundamental social right, in practice there are often in Brazil not enough public *creches* to fully meet the demand. Provision of daycare services in Brazil may be private or public, the latter being free of charge. In 2007, there were 244 public daycare centers in Rio de Janeiro servicing 6.8 percent of the city's 0-4 population, and 352 private centers servicing 7.3 percent of the children.[2] Therefore, not surprisingly, there was an excess demand for the available public daycare slots. In this context, the municipal government agreed to implement in 2007, for entry in the academic year of 2008, a lottery to distribute the available vacancies.

[1] Rio de Janeiro is a relatively high income city within Brazil, accounting for 5 percent of the national GDP. When compared to the average of the country, in 2007 Rio de Janeiro had a higher GDP per capita (11.477 USD as opposed to the 7.374 USD country average), and the poorest 10 percent individuals in the city had a per capita monthly income of 58 USD, substantially higher than the 34 USD in the rest of the country. In terms of income inequality, however, Rio de Janeiro was worse off than the Brazilian average, displaying a Gini coefficient of 0.56, slightly higher than the 0.55 national index in 2007. Source of data: Brazilian Institute of Geography and Statistics - IBGE.

[2] Source of data: School Census - INEP/MEC.

1.3.1 The Creches Program in Rio de Janeiro

Rio de Janeiro's public daycare program is an integrated early childhood development program offered by the municipal government free of charge for children aged 0-3 living in low-income neighborhoods.[3] The program includes full-time daycare during weekdays (7am to 4:30pm), which consists of a variety of center-based activities specifically tailored to the four different age-groups.[4]

Daycare services offered in *Creches* include a variety of daycare activities throughout the day mixed with times for relaxation and rest. Among daycare activities, there are physical activities as well as other learning activities that involve playing with instructional toys, art, music, storytelling, among others. Children also have access to five meals throughout the day, which are offered to all *creches* according to a standardized menu developed by a nutritionist to ensure a balanced diet.[5] In addition, the program foresees involvement by parents as a way of improving knowledge about good parenting practices, and also offers integrated health services (dental and medical) as government health professionals pay frequent visits to each *creche* to monitor the health status of the children and intervene when needed.

1.3.2 Selection Criteria and Experimental Design

The allocation of available slots in *creches* in Rio de Janeiro, up to 2006, was decentralized and assigned under the responsibility of each *creche*'s management. Government's guidelines for allocation indicated a general criteria of prioritization according to social vulnerability, which involved taking into account whether the children: i) had special needs, ii) had any chronic diseases, iii) were living in poor households, iv) were in households with members in conflict with the law and v) had parents that needed access to daycare to be able to work. However, as public *creches* are mostly located in the low-income neighborhoods of the city, the vast majority of the children applying to the available slots met at least one of these criteria, so that final allocation decision was mostly within *creche*'s management discretion. Therefore, in 2007, the municipal government decided to implement a lottery to allocate the available slots in a more structured and transparent way for the upcoming 2008 academic year.

For 2008, there were 244 public day care centers providing these services spread around most low-income neighborhoods of the city. As not all *creches* offered services for all four age groups, the total of creche-age group combinations for the 2008 academic year were 847. A total of 25,511 children applied for the available 11,640 slots offered in *creches* by the municipal government in that year.

Children considered high priority and children with special needs, a total of 957 and 660 respectively, were granted a slot in a *creche* without the need to participate in the lottery.[6] Therefore, a lottery was carried out to distribute the remaining 10,033 vacancies among all the other 23,904 applicants, which all met at least one of the vulnerability criteria above mentioned.

[3] The Brazilian constitution states that early childhood care provision is under the mandate of municipal governments.

[4] *Bercario I*: from 0-11 months old; *Bercario II*: from 12 to 23 months old; *Maternal I*: from 24 to 35 months and *Maternal II*: from 36 to 47 months.

[5] Standardized menus available at the government's website at: http://prefeitura.rio/web/sme/exibeconteudo?id=6482166.

[6] High priority children were selected by each creche's management, considered well equipped to identify children under exceptionally vulnerable circumstances.

The selection of beneficiaries was made through a lottery specific to each creche-age group. A public lottery was carried out only in those groups in which there were vacancies being offered and excess demand. Those not selected through the lottery were ranked in a waiting list.

1.4 Sample and Data

1.4.1 Sample

A balanced sample of 4350 children in 232 creche-age groups was selected for the impact evaluation among the creche-age groups that participated in the lottery. The number of children selected for the sample in each creche-age group varied between 5, 10, 15 or 20 children from both treatment and control groups, depending on the number of vacancies offered and the size of the waiting list. Creche-age groups with less than 7 vacancies or less than 7 children on the waiting list were not included. In each group, the first children in the lottery and the last children in the waiting list were chosen to compose the sample, in an attempt to avoid contamination.

1.4.2 Administrative Data

Administrative data for every applicant to a *creche* slot stems from a short questionnaire answered by the person responsible for the child at the time of application, which included identification information as well as information to identify vulnerability status of the child at the time. This information was used to assess whether the child met at least one of the vulnerability criteria that was a requirement for being eligible to enter the lottery.[7]

1.4.3 Surveys

Primary data was collected in three waves of surveys. The first round of survey data was collected between July and December 2008 (6-11 months after the lottery winners were exposed to childcare). The main purpose of this survey was to collect information to assess the short term impacts of the program on family welfare, including labor market outcomes, time allocation of the child's main carer, household income and assets and stress of the mother. This survey also collected data on whether children in the sample were enrolled in a public *creche* or any other daycare alternative.

Two additional surveys were carried out, in 2012 and 2015, several years after lottery winners had the opportunity to enrol in *creches*, with the purpose of measuring the medium-term impacts of the program on family labor market outcomes and on child development. By 2012, less than 1 percent of our sample was still attending *creches*. Collecting data to measure longer term impacts was crucial since many early childhood interventions previously evaluated have been shown to have effects that fade away with time.

[7] Identification information included name and gender of the child, date of birth and *creche* to which the child was applying to. In addition, information to identify vulnerability status of the child included household size, work status of the person responsible for the child, whether the person depended on daycare to be able to work, whether the child had any chronicle disease, whether the child had special needs, whether any member of the family was involved in substance abuse or was ever imprisoned, and whether the family lived in the community.

The survey implemented in 2012, due to financial constraints, only interviewed a subsample of 64 *creches*, corresponding to approximately 40 percent of the sample. The socioeconomic questionnaire, answered by the person responsible for the child, included information on income and assets, labor market outcomes for all family members, stress of the mother and home environment characteristics. It also had a module to record a detailed history of daycare attendance by the child and a module in which the enumerator answered a few observational questions about the attitudes of the person responsible to the child towards the child during the interview.

In 2012, we also collected data to assess child development in terms of cognitive function, child behavior and anthropometrics (height and weight of the child). Cognitive function assessment stems from data based on the following tests:[8]

- The TVIP Peabody picture vocabulary test (DUNN et al., 1986; LIMA, 2007), which measures vocabulary development.
- The following measures of executive function skills, which relate to to working memory, mental flexibility, and self-control: a) Head Toes Knees Shoulder (PONITZ et al., 2008), b) Pencil Tapping Test (DIAMOND; TAYLOR, 1996), and c) Stroop Test (STROOP, 1935).
- Two batteries of the Woodcock-Johnson-Munoz tests (WOODCOCK et al., 2005) related to visual-spatial thinking and associative memory: a) WJ Visual Integration and b) WJ Memory for Names.

Child behavior is measured based on the CBQ - Child Behavior Questionnaire (ROTHBART et al., 2001), administered on the mother, aimed at providing a detailed assessment of temperament in children 3 to 7 years old. The CBQ also has five subscales which we analyze separately:

- Frustration, which is associated with the negative reactions to interruption of ongoing tasks.
- Attention focusing, related to the ability of maintaining attention focus upon tasks.
- Soothability, which is related to rate of recovery from excitement or distress.
- Impulsivity, which relates to speed of response initiation.
- Inhibitory control, which is related to the capacity to suppress inappropriate responses under instructions.

In the 2015 round, the same household questionnaire was applied on the entire sample. Child development measures of anthropometrics and child behavior were collected again using the same instruments applied in 2012, but cognitive development data in 2015 stems from the Wechsler Intelligence Scale for Children-IV (WECHSLER, 2003), which measures IQ as an aggregate measure of its components: [9]

[8] All cognitive tests carried out measure factors of the Cattell-Horn-Carroll theory on the structure of human cognitive abilities (MCGREW, 2005; ALFONSO et al., 2005).

[9] The IQ test measures factors of the Cattell-Horn-Carroll theory on the structure of human cognitive abilities (MCGREW, 2005; ALFONSO et al., 2005)

- Verbal comprehension index measure, which assesses children's verbal concept formation.
- Perceptual reasoning index measure, which evaluates non-fluid and fluid reasoning.
- Working memory index measure, which assesses children's working memory.
- Processing speed index measure, which assesses speed of information processing.

Details on each of the assessments are presented in Appendix F.

1.4.4 Attrition and Balancedness

Between June and October 2008, a sample of 3,777 households were surveyed out of the universe of 4,348 households. Of the 3,777 households successfully interviewed in 2008, due to financial constraints, only 2,124 were approached for interview in 2012. These families correspond to all families residing in 6 out of the 10 education districts in the original sample. Of these 2,124 families, 1,462 were actually found. In 2015, these families were surveyed again, together with the families from the districts not covered in the 2012 survey. In this round, 1,125 were found among those interviewed in 2012 and 925 were found in the remaining group (interviewed in 2008 and not interviewed in 2012), making a total sample of 2,050 households in 2015. Table 1 shows the difference in proportion of attrition between lottery winners and losers between the different waves of data collection. Although survey rates are higher in the treatment group in all waves (at most 3 percentage points), we show in Appendix Table A.1 that differential selective attrition does threaten the validity of our results.

Table 2 shows that characteristics are balanced across treatment arms. Although 4 out of 21 variables display significant differences between treatment and control groups, coefficients are very small and we cannot reject the null hypothesis that all coefficients are equal to zero (p-value joint=0.130).

1.5 Empirical Strategy

Winning the lottery guaranteed a slot in a *creche*, but individuals may not have taken up the offer. Losing the lottery did not prevent children from reapplying to the lottery in the following year. Therefore, lottery results affect the probability of daycare attendance but are not a perfect predictor of enrollment. Our main analysis focuses on intention to treat (*ITT*) estimates, which reflect the impacts of being offered a slot in *creches* on our outcomes of interest.

Therefore, our ITT estimates are based on the following regression:

$$y_{igc} = \alpha + \beta_{\text{ITT}} L_{igc} + \Gamma \mathbf{X}_{igc} + \delta_{gc} + \epsilon_{igc} \tag{1.1}$$

where y_{igc} is an outcome of interest for individual i, who participated in the lottery for age group g in day care center c, L_{igc} is an indicator variable that takes value 1 if individual i is a lottery winner and 0 otherwise, $\mathbf{X}_{igc}$ is a set of baseline individual level controls (race and gender of the child), δ_{gc} is a set of strata fixed effects (for each age group-day care center pair at which each lottery took place), and ϵ_{igc} is an

error term. $\beta_{\texttt{ITT}}$ is the ITT coefficient, which measures the impact of winning the lottery on the outcome of interest.

Our main results (tables 3 - 12) show estimates of equation 1.1 for years the child spent in *creche*, household income and labor market outcomes of family members, child development and home environment measures, for each of the survey years separately. Tables in Appendix D show estimates of equation 1.1 for the main outcomes, restricted to the observations for which we have a balanced panel. In addition, tables in Appendix E show estimates for the impact of the lottery on the probability of being above the 2nd, 4th, 6th and 8th deciles of the distribution of each outcome of interest, estimated jointly in a seemingly unrelated regressions model.

Since children not offered a slot in *creches* in 2008 were eligible for entering the lottery in the subsequent years, many of the no offer group children eventually enrolled in public *creches*. Some children also enrolled in alternative daycare arrangements, such as private daycare centers or community-based daycare centers. In order to go beyond the intention to treat estimates and measure the actual effect of attending *creches* on our main outcomes of interest, we also use an instrumental variables strategy (IV), where the lottery status is the instrument for *creche* attendance, since it is correlates directly with treatment but not with our outcomes of interest. In this setting, the IV estimation identifies a local average treatment effect, which corresponds to the impacts on the individuals induced to go to daycare by a change in the instrument. Our measure of *creche* attendance is a variable that reflects years attending daycare, ranging from zero to four, based on self-reported data collected on the different survey waves.[10]

Our IV estimates are based on the following two-stage approach:

$$\begin{aligned} y_{igc} &= \alpha + \beta_{\texttt{LATE}}\widehat{T}_{igc} + \Gamma\mathbf{X}_{igc} + \delta_{gc} + \epsilon_{igc} \\ T_{igc} &= \gamma + \eta L_{igc} + \phi\mathbf{X}_{igc} + \delta_{gc} + \xi_{igc} \end{aligned} \tag{1.2}$$

where y_{igc} is an outcome of interest for individual i, L_{igc} is the instrumental variable for predicting years in *creche* T_{igc}, $\widehat{T}_{igc}$ is the predicted value of years in *creche*, $\mathbf{X}_{igc}$ is a set of baseline individual level controls, δ_{gc} is a set of fixed effects for each age group-daycare center pair, and ϵ_{igc} and ξ_{igc} are error terms. $\beta_{\texttt{LATE}}$ is the local average treatment effect coefficient.

Tables in Appendix C display the results for the instrumental variables estimation for the main outcomes.

1.6 Results on Main Outcomes

1.6.1 Daycare Enrollment

The lottery guaranteed a slot in public *creche* for the lottery winners in 2008, but it did not prevent children from entering the lottery in the following year, or from enrolling in alternative daycare provision arrangements. In this section, we show that winning the lottery had a strong effect on *creche* attendance,

[10] Survey collected detailed data on history of daycare attendance, including detailed information on which center the child attended in each semester. The variable years in *creche* takes value 0 if child never attended daycare, 1 if child attended between 0 and 2 semesters, 3 if child attended between 4 and 6 semesters and 4 if child attended more than 6 semesters

based on self-reported survey data, collected in 2012 and 2015. Several years after the randomization took place, 97 percent of lottery winners and 78 percent of lottery losers reported to have attended daycare.

Table 3 presents estimates of equation 1.1 on *creche* attendance. Column 1 shows the impact of the lottery on years attending *creches*, while columns 2-5 show results for the probability of being enrolled in *creche* for more then 1, 2, 3 and 4 years respectively, estimated jointly in a seemingly unrelated regressions model. The estimates show that winning the lottery led to approximately one extra semester in daycare (0.637 years) and increased the probability of attending 1 or more years of *creche* by 19 p.p., 2 years or more years by 21 p.p., 3 years or more by 17 p.p. and 4 years or more by 7 p.p. Figure A.1 in Appendix B displays years in *creches* by lottery status.[11]

In the next subsections, we look at the impacts of winning the lottery and creche attendance on a wide range of child and family outcomes.

1.6.2 Household Income and Labor Market Outcomes

This section examines the impact winning the lottery had on household income and labor market outcomes of the family members living in the household. ITT results in Table 4 show estimates of equation 1.1 on household income for 2008, 2012 and 2015. Winning the lottery had a very strong impact on monthly household income in 2008 (of approximately 50 Brazilian Reais, which corresponded to approximately 8 percent of control group mean) and also in 2012 (approximately 111 reais, or 10 percent of control group mean). The gains in 2008 household income were registered throughout the entire distribution, as displayed in Appendix Table A.10, while the 2012 gains were more concentrated at the top of the distribution. In 2015, however, the results on income fade out as there are no statistically significant differences between lottery winners and losers, although the coefficient is still large, corresponding to approximately 5 percent of control group mean. The results for the balanced panel in Table A.7 corroborate the ITT results with the full sample, ruling out the possibility that our results are being driven by different samples in each year. IV estimates presented in Appendix Table A.2 show that an additional year of child attending *creches* increases substantially household income, by 101 and 185 Brazilian Reais respectively in 2008 and 2012 (16 and 17 percent of control group mean).

Table 5 displays, in panels A-D, respectively, ITT estimates for monthly income, employment status, weekly working hours and contribution to social security, which may be an indication of whether the individual is formally employed.[12] Separate regressions are estimated for different household members (aged 16 or above): i) child's carer (columns 1-3), ii) parent or step-parent (columns 4-5) iii) grandparent (columns 6-7), iv) sibling (columns 8-9) and v) other family members (columns 10-11). For all household members, we have results for 2012 and 2015, with the exception of the outcomes for the carer, for which we also have 2008 results on some of the dimensions. The carer, whose identity is displayed in Figure A.2 in Appendix B, is accounted in the other household member categories if they live in the household. In 2008, 89 percent of children had their mother or father as the main carer, and 7 percent had their

[11] The variable years in *creche* takes value 0 if child never attended *creche*, 1 if child attended between 0 and 2 semesters, 3 if child attended between 4 and 6 semesters and 4 if child attended more than 6 semesters

[12] Monthly income variable is equal to zero if the individual did not work; employment status variable is equal to one if individual is currently employed and 0 otherwise; weekly working hours variable is equal to zero if individual did not work.

grandmothers caring for them most of the time. For 2015, 84 percent of the children were cared for by one of their parents, while 10 percent had as the main carer their grandmothers, and 2,5 percent were cared for by their siblings.

Results show that winning the lottery generated in 2012 systematic effects on labor market outcomes, particularly for grandparents, with very large and significant results for all four dimensions considered. Estimates show a roughly 55 percent increase in monthly income (245 Reais), a 21 percentage points increase in employment, a roughly 50 percent increase in weekly working hours (10.68 hours) and a 22 percentage points increase in social security contribution. These effects, however, fade out with time and are no longer significant in 2015. Strong effects are also registered for siblings above 15 years old, which had significant increases in monthly income, employment and contribution to social security. For the parents, there was a significant increase in monthly income in 2012 (of 53 Reais or approximately 7 percent of the control group mean), although no significant results are found for the other variables, and no significant differences in 2015. For the carer, very strong impacts are registered in 2008, soon after the lottery took place (approximately 5 percentage points higher employment rate and an increase of 1.9 in weekly working hours, which corresponds to roughly ten percent of control group mean). These results persist through 2012, when a significant effect is also found on monthly income (14 percent of control group mean), employment (5 p.p.) and working hours (2.471, roughly ten percent of control group mean), but again no differences between treatment and control persist throughout 2015.

IV estimates (Appendix Table A.6) show an additional year of *creche* attendance largely increases monthly income of household members, particularly significant for the carer, parents and grandparents in 2012 (101, 86 and 405 Reais respectively). Significant impacts are also observed on employment (8 p.p. for the carer in 2008 and 2012, 34 p.p. for the grandparent in 2012, and 12 p.p. for the siblings above 15 in 2015), weekly working hours (2.9 and 3.9 for the carer in 2008 and 2012 and 19.8 for the grandparent in 2012) and contribution to social security (36 p.p. for grandparents as well as siblings in 2012).

1.6.3 Child Development Outcomes

Child development outcomes are measured in this study in terms of anthropometrics, cognitive function and child behavior (the first two stemming from the measures and tests applied on the child and the latter from the child behavior questionnaire answered by the mother). The age range at which each instrument was applicable varied across instruments.[13] Throughout the paper we standardized all scores to have mean zero and standard deviation one within age, and within the sample. Height and weight were standardized using the World Health Organization (WHO) growth standards to calculate height for age and weight for age z-scores (HFA and WFA).[14]

Table 6 reports estimates of the impacts of winning the lottery on anthropometrics measures, HFA and WFA, for both 2012 and 2015. The estimated coefficients for all anthropometrics regressions are large and statistically significant. For 2012, ITT estimates are 0.163σ and 0.199σ for HFA and WFA

[13] Details on each instrument and age applicability follow in Appendix F.

[14] The WHO growth standards only has standards for WFA up to age 114 months. Therefore, for 2015, since some children were older than 114 months, we imputed age=114 to compute WFA z-score for the entire sample. This imputation did not implement major distortions, as we did the same exercise for height as a robustness check, and the results were similar to those calculated with the correct age.

respectively. On average, this corresponds to 0,51 cm increase in height and 600 grams increase in weight.[15] Effects persist throughout time and, in 2015, although the effects are smaller, they are still significant and large (0.11σ and 0.14σ). Results from the balanced panel in Table A.9 corroborate these results. IV estimates presented in Appendix Table A.4 show one additional year attending *creches* increased, for 2012 and 2015 respectively, HFA by 0.269σ and 0.170σ and WFA by 0.327σ and 0.217σ. In Appendix Table A.11, we show that the gains in HFA were more concentrated at the bottom of the distribution for 2012, but spread throughout the distribution for 2015. For WFA, gains were dispersed throughout the distribution for both 2012 and 2015. These results are encouraging as they are consistent with the effects of nutrition-specific interventions in the literature. A meta-analysis by Bhutta et al. (2008) indicates that interventions specifically targeted at enhancing children nutrition increase height-for-age score by approximately 0.25σ in populations with sufficient food, which is the case of our sample, since the z-score of the control group is above the mean according to the WHO standards.

The ITT results on cognitive function, based on estimates of equation 1.1 for the 2012 and 2015 cognitive development tests are presented in panels A and B of Table 7. Column 1 shows the impact of winning the lottery on an aggregate measure of the cognitive function tests, which is constructed by using factor analysis to combine the child's test results presented in columns 2-5, and then standardizing it within the sample and within age group to obtain a variable with zero mean and σ equal to 1. Results for each of the tests considered are individually presented in columns 2-5: respectively, TVIP, an aggregate of executive function tests, Woodcock-Johnson-Munoz's Memory for Names and Visual Integration. Winning the lottery had a positive significant impact of 0.067σ on the z-score of our aggregate measure of cognitive development, driven by the significant impacts (0.112σ) on vocabulary development measured by the TVIP test. No significant impacts were found on average for the other measures of cognitive development. However, we do find distributional effects for the executive function score, for which the lottery increased the probability of being above the 6th decile of the distribution, as displayed in Appendix Table A.12. Significant positive effects were also registered at the bottom of the distribution for the Memory for Names score. Positive results on cognitive development were still present in 2015, as displayed in Panel B of Table 7, that shows in columns 1-5 the effects on the aggregate IQ z-score as well as the effects of winning the lottery on the four different IQ's components. We find a significant impact of 0.091σ on the Perceptual Reasoning index, although no significant effects were captured by the aggregate IQ measure or the other components. Perceptual reasoning is related to non-verbal and fluid reasoning, which is reflected on child's ability to examine a problem, organize their thoughts and create solutions.

Appendix Table A.3 shows IV results for the main cognitive function measures for which we registered a noticeable ITT effect. One additional year attending *creches* increased the aggregate cognitive score by 0.112σ and the TVIP score by 0.191σ for 2012, and IQ-Perceptual reasoning index by 0.144σ for 2015. These results are consistent with other findings in the literature, such as in Bernal and Fernández (2013), which find community-based daycare in Colombia enhanced cognitive development by 0.15σ to 0.3σ after at least 15 months of exposure to the program.

Table 8 displays in columns 1-6 the results for the last set of children's development outcomes for 2012 (Panel A) and 2015 (Panel B): the aggregate child behavior index z-score and its components,

[15] We get these estimates by using the raw variables instead of standardized ones, and including age dummies as controls.

indices of frustration, attention, soothability, impulsivity, and inhibition, taken from the Child Behavior Questionnaire answered by the mother. All indexes are standardized within age and within the sample. We do not find substantial or statistically significant impacts of daycare attendance on average scores of child behavior. Further examining the impacts of winning the lottery on the distribution of the CBQ scores (appendix table A.14) we confirm that winning the lottery had no impact on this dimension, on the average or the distribution.

1.6.4 Home Environment and Assets

The previous sections have shown that winning the lottery had strong effects on income and other labor market variables, as well as on several child development dimensions. This section looks at whether lottery winners experienced a significant change in other income-related variables as well as on home environments. Table 9 shows the impacts of the lottery on mean expenditures, asset index, access to bank account and access to credit in panels A-D respectively. Columns 1-3 show the results for 2008, 2012 and 2015. Results show that lottery winners' higher income was accompanied by an increase in assets, captured by a 0.066σ and 0.131σ increase in the asset index in 2008 and 2012 respectively. Our findings also show that there was an expansion in monthly expenditures of 28 Reais in 2012, approximately 5 percent of the control group mean. Lottery winners were also more likely to have at least one household member with a bank account (by 7 p.p.), which indicates greater access to the financial system, although no impacts were found on access to credit. Nevertheless, the impacts found had faded away by 2015.

IV estimates presented in Table A.2 show that an additional year of child attending *creches* increased, for 2012 outcomes, mean expenditures by 45 Brazilian Reais, asset index by 0.218σ and probability of having at least one household member with a bank account by 12 percentage points.

Winning the lottery also had a positive impact on home environments. Table 10 shows the effects of the lottery in panels A-E on time carer spends with the child, ever singing or reading for the child, number of children's books in the household, attitudes towards the child and a measure of stress of the person responsible for the child. Columns 1-3 show estimates for 2008, 2012 and 2015 respectively. As expected, children that won the lottery had less time spent with the carer in 2008, but this difference did not persist in 2012 and 2015, when most children were out of daycare already. Children that won the lottery were 7 percentage points more likely to have someone from the household ever reading or singing for them in 2012. We also have a measure of survey respondent's interaction with the child, based on questions answered by the enumerator on the nature (positive or negative) of the survey respondent's attitudes towards the child during the interview.[16] In 2012, we find that a significantly smaller share of adults had negative attitudes towards the child among lottery winners, when compared to those that had not won the lottery (impact of -0.013 on negative attitudes, approximately one third of control group mean). No significant differences are registered for the positive attitudes measure, and no results are statistically different between lottery winners and losers in 2015. One potential explanation is that this may reflect

[16] For this measure, the enumerator answered 9 questions on the nature of the interaction of respondent with the child. Each question admitted a yes/no answer and asked questions such as whether the respondent talked to the child at least twice during the interview, whether they kissed or hugged the child, hit, depreciated or yelled at the child, among others. The aggregate measure of positive/negative attitudes towards the child is the percentage of positive/negative answers recorded.

the parents learning from the daycare staff about good parenting practices. We also measured carer's stress based on the Perceived Stress Scale, validated for Portuguese by Luft et al. (2007). Based on the instrument's answers (in a Likert 5-point format), we constructed a stress z-score, and find that winning the lottery significantly reduced the stress of the mothers in 2008 by approximately 0.08σ, however this effect had faded out when we measured the same outcome years later, in 2012 and 2015.[17]

Appendix table A.5 shows IV results for home environment outcomes, indicating that an extra year of *creche* attendance reduces the stress of the mother in 2008 by 0.120σ, increases the probability of anyone in the household ever reading or singing for the child in 2012 by 11 p.p., and reduces the percentage of adults that had negative attitudes towards the child during the interview by 2 percentage points.

1.7 Channels through which Creches Impact Child Development

Our results from the previous section show that winning the lottery had very strong impacts on some dimensions of child development. The literature indicates that daycare programs may impact child development directly through stimulation and nutrition that happen while child is being serviced at the daycare centers, but also indirectly through enhancing intermediate outcomes in the family environment. At home, potential channels may be higher household income (due to carer's increased labor market attachment) and improved home environments (due to parents learning good parenting practices while interacting with daycare staff). The purpose of this section is to examine the extent to which observed effects are related with improvements in intermediate outcomes, potential mediators that may act upon fostering child development.

Following Heckman and Pinto (2015), we rely on the assumption that child development outcomes are a function of both measured and unmeasured inputs, affected by treatment, as well as baseline characteristics, not affected by treatment. For this exercise, we have data on intermediate family outcomes, related to household income and home environments (which are affected by *creches*, as presented in the previous section), but we do not have measures of other intermediate outcomes at the *creche* level that may also act as potential mediators.

Therefore, our mediation analysis exercise, following Oreopoulos et al. (2017), is based on the following specification:

$$y_{igc} = \alpha + \beta_{\texttt{residual}} L_{igc} + \Gamma \mathbf{X}_{igc} + \sum_j \Theta^j \mathbf{D}^j_{igc} + \delta_{gc} + \epsilon_{igc} \tag{1.3}$$

in which we assume child development outcomes y_{igc} are a linear function of child individual characteristics $\mathbf{X}_{igc}$ and j mediators $\mathbf{D}_{igc}$, represented by variables related to household income and home environment of the child. L_{igc} is an indicator variable that takes value 1 if individual i is a lottery winner and 0 otherwise, δ_{gc} is a set of strata fixed effects, and ϵ_{igc} is an error term. $\beta_{\texttt{residual}}$ is the coefficient that reflects how much of the child development outcome y_{igc} is not explained by changes in the intermediate family outcomes, hence capturing the effect of all other relevant unmeasured inputs.

[17] The stress z-score was computed through item response theory estimation.

The coefficients on the potential mediators must be interpreted with caution. If there are unmeasured inputs that determine the outcome of interest y_{igc} and are positively correlated with the mediators in specification 1.3, estimates for Θ^j are likely to be overestimated, as they will capture the effects of these inputs as well.

Results for estimates of equation 1.3 for two dimensions of child development (cognitive function and anthropometrics) are displayed in Tables 11 and 12 respectively, for selected outcomes in years 2012 and 2015. For cognitive function, we present mediation analysis for TVIP in 2012 and WISC-Perceptual reasoning in 2015, the two test scores that registered significant ITT effects. For each outcome, keeping the estimation sample constant, we report four columns: i) ITT results, ii) estimates of 1.3 including the income-related variables as mediators, iii) estimates of 1.3 including home environment variables as mediators and iv) estimates of 1.3 including all income-related and home environment as mediators. For the mediation analysis on anthropometrics, for each outcome reported (HFA and WFA), three columns are presented for each year (keeping the sample constant): i) the ITT results, ii) estimates of 1.3 including the income variables as mediators and iii) estimates of 1.3 including all income-related variables as mediators.

The results suggest that the mediators we consider seem to be related to children's cognitive gains. When controlling for income-related variables, the ITT coefficient on the TVIP of 0.096 is reduced to 0.055, which is the $\beta_{\texttt{residual}}$. For the WISC-Perceptual reasoning index, coefficients drop from 0.084 to 0.064 after controlling for income. Home environments are also correlated to gains in TVIP (as coefficients drop from 0.096 to 0.068 after including the home environment mediators), although for WISC-Perceptual reasoning the coefficients do not change much as a result of including these controls. This suggests that the differences observed in the 2012 cognitive development indicators are potentially linked to the positive impacts *creches* had on both household income and home environments, while the 2015 gains in cognitive development are to a great extent related to the increased household income.

For the anthropometrics measures, we observe a different picture. As we have no strong reason to believe our measures of home environments have an impact on height and weight of children, we only include income-related controls as mediators. When controlling for these, the $\beta_{\texttt{residual}}$ coefficients on HFA and WFA remain significant (marginally significant for HFA in 2015), and are not substantially smaller than the ITT coefficients. Increases in family income seem to have played a minor role in the improvements registered on anthropometrics outcomes. This suggests that the observed effects are mostly explained by factors happening at the *creche* environment, not at home. These findings may be explained by the fact that children attending *creches* are mostly from lower income families that may not provide a balanced diet for children at home, not only due to insufficient income, but possibly because they lack information on the importance of a nutritional diet for child development. Therefore, having access to a nutritional diet offered in daycare through five meals a day had, not surprisingly, substantial positive impacts on children's height and weight.

1.8 Discussion of Results

Our findings presented in the previous sections showed that winning the lottery increased *creche* attendance by approximately one semester. Instrumental variables estimates for effects of one additional year attending *creches* are roughly consistent with intention to treat results by being, overall, nearly twice

the size of the ITT coefficients. *Creche* attendance benefited lottery winners and their families through a wide range of dimensions.

From the household welfare perspective, we find very strong effects on household income and labor market outcomes, including monthly income and employment status, not only for the main carer, but also for other household members, particularly for grandparents and also for the siblings. With the exception of siblings, which experienced sustainable gains on employment, the income and labor market effects for household members are mostly concentrated in the shorter run, as their point estimates become smaller over time and indistinguishable from zero by 2015. Other income-related variables (such as asset index and mean expenditures) also display significant effects for lottery winners up to four years after the lottery took place, but these again fade out by 2015, 7 years after the lottery took place.

This result suggests that increased labor market attachment of older household members due to enrolling the child in *creches* was at most a temporary effect, which did not seem to have significant impacts on skills accumulation, at least for parents and grandparents. The fact that the differences in household labor market outcomes between lottery winners and losers are indistinguishable from zero 7 years after the lottery took place suggests that the additional labor market experience induced by children's increased *creche* attendance was not enough for shifting parents' and grandparents' labor market trajectories to higher productivity and higher paid jobs in the longer run. For siblings, however, enrolling the child in *creches* seems to have had more lasting benefits. The impact of the lottery on employment of siblings remained positive and significant by 2015, by ten percentage points. Although we don't find significant 2015 results in terms of income, working hours and contribution to social security, our findings of higher employment rates for siblings of lottery winners 7 years after the lottery took place are, at the very least, an indication that their labor market trajectories have been shifted, as they are being exposed to more working experience, which may be translated in income gains in the future.

Nevertheless, the ability to enrol the child in *creches* enabled a temporary increase in family's disposable income which, according to our mediation analysis, had an important role to play in the observed gains in children's cognitive development in both 2012 and 2015. For both years, income gains explain a significant share of the cognitive development observed treatment effects.

Winning the lottery also had an effect on home environments, which also appeared to be an important factor through which enrolling in *creches* affected child cognitive development, particularly for the effects observed in 2012. One potential hypothesis to explain these results is that increased income combined with potentially learning best parenting practices from care providers may have enabled parents to provide a more stimulating environment for children at home, boosting their accumulation of cognitive skills.

While *creches* had significant cognitive function impacts registered for both 2012 and 2015, significant treatment effects were not observed in all cognitive dimensions measured. The 2012 effect was driven by the vocabulary development dimension, while in 2015 the main driver was the perceptual reasoning index, which relates to non-verbal and fluid reasoning, determinants of child's ability to examine and solve a problem. Although there are no effects on the mean executive function z-score, lottery winners had a higher probability of being above 60 percent of the distribution of the executive function index. Executive function is a measure of cognitive skills that is also associated to self-control, which is linked to

noncognitive dimensions (SCORZA et al., 2016). Although our study does not have a direct measure of noncognitive dimensions, we do not rule out the possibility of identifying in the future significant effects of the lottery on noncognitive skills, as the ECD literature has records of programs that had cognitive function effects fading out in the medium run, but persistent effects on personality and social skills (HECKMAN et al., 2006; BLAU; CURRIE, 2006).

While gains in cognitive development of the child seem to have been mediated by the impacts *creches* had on elements happening at home, our mediation analysis results also show that the very large gains observed in the height and weight of the child are likely happening through a different channel. The fact that the anthropometrics results remain large and significant after controlling for household income and income-related mediators suggests these are likely not the main intermediate outcomes through which daycare affects height and weight of the child.

Children in public *creches* in Rio de Janeiro have access to five meals a day, which is standardized for all daycare units in the city. All meals are planned following guidelines for a healthy and balanced diet developed by the *Instituto de Nutricao Annes Dias*, a government institute in charge of developing and monitoring the implementation of healthy eating habits in all government-related institutions, including hospitals and schools.[18] Therefore, it seems that children in *creches* had access to a healthier diet than the one they would have had access to if they had stayed at home. This hypothesis is not unlikely, given most of the children accessing public daycare centers in Rio de Janeiro come from lower income households that may not only face income restrictions, but also lack the awareness of the importance of securing a nutritional and balanced diet for child anthropometric development. The lack of information is a reasonable hypothesis in this case since the observed higher household income does not play a major mediation role for the impacts on anthropometrics measures.

1.9 Conclusion and Policy Implications

The evidence to inform the policy debate on whether governments should subsidize universal public childcare programs has been characterized by heterogeneous findings. Although there are several empirical studies that investigate this issue in a developed country setting, evidence for developing country programs is more limited. The purpose of this paper is to generate novel experimental evidence on the impacts of a large-scale subsidized institutional daycare program in a middle-income country over a wide range of child and family outcomes.

The findings of our study show that attending *creches* in Rio de Janeiro had positive impacts on household income, labor market outcomes of the family members and home environments, which mostly faded out with time. For siblings, however, the effects on employment persisted by 2015, seven years after the lottery took place. The program also had noticeable treatment effects on child cognitive development, which seems to have emerged partially through the temporary increase in family income. *Creches* also had very large effects on children's anthropometric development. Both cognitive and anthropometric development are extremely relevant as they are inextricably linked with the level of welfare the child will experience later in life.

[18] http://www.rio.rj.gov.br/web/sms/alimentacao-e-nutricao

Children that won the lottery had enhanced height for age and weight for age z-scores, likely derived from better nutrition at the *creche* level. Improving child nutrition is important not only to secure individuals' welfare, but also from a society's perspective, because it fosters human capital formation and hence economic growth. Poor nutrition is associated with losses in productivity from individuals' poor physical status and poor cognitive function, in addition to the adverse health effects, which increases a society's health care costs (SHEKAR et al., 2006). Interventions that enhance childhood nutrition have been shown to boost physical health, schooling, and adult economic productivity (HODDINOTT et al., 2008; BEHRMAN; DEOLALIKAR, 1988).

Favorable early childhood environments, including cognitive and noncognitive stimulation and good parenting practices, are a major determinant for the development of cognitive, social and emotional competencies. The early development of these skills directly increases the efficiency of learning at later stages of children's lives, and correlate with positive adult schooling and labor market outcomes. There is no shortage of evidence showing that disadvantaged early environments lead to adult disadvantage in a range of social and economic measures. Therefore, programs that have positive impacts on cognitive development of young children, such as *creches*, promote not only social justice but also enhance the productivity in the economy, benefiting the society overall (HECKMAN, 2006; HECKMAN, 2008; ALMOND; CURRIE, 2010)

Although we find positive impacts on children's cognitive skills as a result of *creche* attendance, our study does not shed much light on the wide range of noncognitive skills dimensions, which are all also crucial for adult success. It is not unreasonable to expect that *creches* may have a role in developing noncognitive skills through the promotion of social interaction since early ages. The literature suggests that, while cognitive effects of ECD interventions often fade out with time, treated children often experience long lasting effects due to increased noncognitive skills (HECKMAN, 2006; CARNEIRO, 2003). For the widely known Perry program, for example, treatment effects on IQ scores faded out by the age of ten, but treated children had long lasting advantages in terms of schooling, labor market outcomes, welfare dependency and criminal record, explained by higher levels of noncognitive skills (MANNING; PATTERSON, 2006).

Our study finds no impact, on average, on executive function skills, which refer to higher order cognitive skills but are intimately connected with the capacity to regulate emotions, linked to a range of interpersonal behavioral outcomes (SCORZA et al., 2016). We also find no effects on children's behavior based on maternal self report data. However, further research is needed as the limited evidence we have on these potentially associated dimensions are not enough to reach a conclusion on the impacts of *creches* on noncognitive outcomes.

The significant treatment effects of *creches* on both child and family dimensions are indisputable. However, an important issue to be discussed is whether there could potentially be other lower cost policy options to deliver the same development results of *creches*, given daycare programs are known to be very costly. Since effects on labor market outcomes are only temporary for parents and grandparents, we find that their main role is to act as an income channel for families to provide better conditions and a more stimulating environment to advance children's cognitive development. In this sense, other mechanisms transferring income for lower income families with children could potentially act in the same direction.

However, income transfers to families alone would likely not be effective in enhancing children's anthropometrics, as the treatment effects on height and weight do not seem to have emerged through increased household income. In this sense, an intervention to deliver similar results should be able to promote better nutrition. There are interventions in the ECD literature that have been proven effective in improving nutritional status of children, such as the home visiting programs and other nutritional supplement programs (WALKER et al., 1991).

All in all, the policy question that remains is how to deliver the same development results more effectively and at the lowest cost. For a precise cost benefit analysis, all potential outcomes of interest should be taken into account. Although our study sheds light on many dimensions, further investigation is needed to shed light on the longer run impacts over cognitive and noncognitive skills of children and labor market outcomes of the siblings, as well as on other subsequent socioeconomic measures.

TABLES

Table 1 – Difference in the Proportion of Non-Attriters Between Lottery Winners and Losers

	2008 to registry	2012 to registry	2012 to 2008	2015 to registry	2015 to 2008	2015 relative to 2012 (matching sample)
	(1)	(2)	(3)	(4)	(5)	(6)
Lottery winner	0.025***	0.023*	0.024*	0.032**	0.028*	0.023
	(0.009)	(0.012)	(0.013)	(0.015)	(0.016)	(0.023)
Control group mean	.856	.331	.366	.456	.504	.750
N	4349	4349	3776	4349	3776	1486

Note – This table shows attrition results for the different waves of our surveys. Each column reports results of a regression of an indicator of whether a given lottery participant had data for a given year (relative to registry or previous wave of survey) on an indicator of winning the lottery and strata fixed effects. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**\,p \leq .05$, $***p \leq .01$.

Table 2 – Covariates Balance - Lottery Winners and Losers

	Lottery loser	Lottery winner	Regression adjusted difference	N
	(1)	(2)	(3)	(4)
Male child	0.507	0.533	0.026	3897
	(0.500)	(0.499)	(0.017)	
White child	0.324	0.346	0.023	3887
	(0.468)	(0.476)	(0.015)	
Black child	0.122	0.105	-0.017	3887
	(0.327)	(0.307)	(0.010)	
Mixed race child	0.524	0.523	-0.002	3887
	(0.500)	(0.500)	(0.016)	
Other race child	0.030	0.026	-0.005	3887
	(0.170)	(0.158)	(0.005)	
Birthweight in quilos	3.189	3.206	0.024	3742
	(0.615)	(0.612)	(0.020)	
Birth height in centimetres	49.26	49.29	0.038	3722
	(4.056)	(4.233)	(0.136)	
Planned Birth	0.329	0.346	0.017	3770
	(0.470)	(0.476)	(0.015)	
First Born	0.442	0.426	-0.014	3764
	(0.497)	(0.495)	(0.016)	
Age of the Mother at Birth	20.28	20.37	0.089	3767
	(4.890)	(4.968)	(0.157)	
Prenatal Care	0.948	0.944	-0.003	3765
	(0.223)	(0.230)	(0.007)	
Natural Birth Delivery	0.691	0.662	-0.028*	3768
	(0.462)	(0.473)	(0.015)	
Premature Birth	0.121	0.131	0.008	3762
	(0.327)	(0.337)	(0.011)	
Breastfed up to 6 Months	0.772	0.751	-0.022	3770
	(0.420)	(0.433)	(0.014)	
HH per capita income	586.2	634.5	56.01	4103
	(1818.9)	(2841.3)	(70.49)	
HH size	4.547	4.638	0.107	4137
	(3.463)	(4.553)	(0.124)	
Age of carer	29.25	29.15	-0.142	3776
	(9.768)	(9.157)	(0.304)	
Carer can read and write	0.965	0.982	0.017***	3768
	(0.184)	(0.134)	(0.005)	
Carer has at least basic education	0.676	0.707	0.034**	3404
	(0.468)	(0.455)	(0.015)	
Carer has at least secondary education	0.325	0.356	0.031**	3404
	(0.468)	(0.479)	(0.016)	
Carer has at least higher education	0.013	0.015	0.001	3404
	(0.114)	(0.122)	(0.004)	
p-value joint		.130		

Note – This table considers covariates balance for the evaluation sample. Columns 1 and 2 show lottery loser and lottery winners means; column 3 displays the results of a regression of each covariate on a dummy variable indicating whether the individual was a lottery winner and strata fixed effects; column 4 reports the number of observations. Robust standard errors are in parenthesis. Data come from registry and 2008 survey. P-value for the F-test of overall significance is reported at the bottom of the table. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table 3 – First Stage: Impact of Lottery on Creche Attendance

	Years in Creche	Probability (Years in Creche $\geq i$)			
		i=1	i=2	i=3	i=4
	(1)	(2)	(3)	(4)	(5)
Lottery Winner	0.637***	0.190***	0.207***	0.168***	0.072***
	(0.048)	(0.012)	(0.017)	(0.019)	(0.014)
N	2410	2410	2410	2410	2410
P-value difference ($\forall i$)		0.000			

Note – This table displays the impact of winning the lottery on average years attending creches (Column 1), and on the probability of (years attending creches greater than i) (Columns 2 - 5). Column 1 shows ITT estimates from a regression that includes strata fixed effects and controls for race and gender of the child. Robust standard errors are in parentheses. Columns 2-5 consider estimates of simultaneous regressions of dummies for attending creches for 1+, 2+, 3+ and 4+ years on lottery status, controls for race and gender of the child and strata dummies. Standard errors are in parentheses. Creche attendance is based on self-reported survey data collected in 2012 and complemented with data from 2015 for the remainder of the sample not surveyed in 2012. The p-value at the bottom of the table is for for the F test of null hypothesis that the differences between all simultaneous regressions coefficients are equal to zero. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table 4 – ITT Estimated Effects on Household Income

	2008	2012	2015
	(1)	(2)	(3)
Lottery winner	49.968***	110.982**	66.011
	(14.880)	(50.031)	(58.307)
Control group mean	613	1102	1361
N	3762	1486	2049

Note – This table considers the impact of winning the lottery on the household income (in current reais) for years 2008, 2012, and 2015, based on self-reported survey data from these years. All ITT estimated effects are from regressions that include strata fixed effects and controls for race and gender of the child. For all years the table displays the control group mean. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table 5 – ITT Estimated Effects on Labor Market Outcomes for Carer and All Household Members

	Main carer			All Household Members							
	Carer			Parent or Step-parent		Grandparent		Sibling		Other	
	2008	2012	2015	2012	2015	2012	2015	2012	2015	2012	2015
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Monthly income											
Lottery winner		62.850**	5.962	52.930*	45.570	245.100***	65.970	117.400**	42.660	33.260	111.700
		(32.010)	(35.290)	(29.970)	(32.580)	(80.510)	(85.050)	(58.370)	(39.460)	(65.720)	(81.080)
Control group mean		461	602	712	769	441	376	230	170	549	555
N		1386	1946	2212	2985	438	478	244	557	615	667
Panel B: Currently employed											
Lottery winner	0.048***	0.048*	0.017	0.007	0.019	0.208***	0.077	0.162*	0.102*	-0.060	0.046
	(0.016)	(0.027)	(0.023)	(0.018)	(0.016)	(0.061)	(0.064)	(0.090)	(0.056)	(0.045)	(0.045)
Control group mean	0.41	0.58	0.61	0.77	0.76	0.51	0.51	0.37	0.34	0.71	0.68
N	3754	1386	1939	2212	2978	438	475	244	555	615	664
Panel C: Weekly working hours											
Lottery winner	1.855***	2.471**	-0.028	0.812	0.041	10.68***	3.407	4.138	1.797	-0.163	2.879
	(0.702)	(1.244)	(1.032)	(1.024)	(0.883)	(3.078)	(3.048)	(4.283)	(2.260)	(2.439)	(2.583)
Control group mean	17	20	21	32	31	20	17	16	11	30	27
N	3753	1352	1874	2126	2811	415	430	241	527	572	593
Panel D: Contribution to social security											
Lottery winner		0.022	-0.008	-0.009	0.020	0.218***	0.066	0.154**	0.020	-0.060	0.055
		(0.026)	(0.023)	(0.022)	(0.019)	(0.059)	(0.064)	(0.072)	(0.044)	(0.051)	(0.050)
Control group mean		0.33	0.42	0.52	0.51	0.28	0.38	0.19	0.17	0.44	0.44
N		1385	1923	2209	2953	436	471	243	551	610	656

Note – This table considers the impact of winning the lottery on i) Average monthly income (Panel A); ii) Employment (Panel B); iii) Weekly working hours (Panel C) and iv) Contribution to Social Security (Panel D). Columns 1-3 show estimates for these outcomes for the main carer of the child for years 2008 (data available only for panels B and C), 2012 and 2015. Columns 4-11 show, for 2012 and 2015, estimates for these outcomes for all family members older than 15 (whom may or may not be the carer), namely parent/step-parent, grandparent, sibling and other, respectively. There may be more than one observation per child in each category in columns 4-11. All ITT estimated effects are from regressions that include strata fixed effects and controls for race and gender of the child. For all measures we include the control group mean. Robust standard errors are in parentheses. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table 6 – ITT Estimated Effects on Anthropometrics: Height for Age (HFA) and Weight for Age (WFA)

	Height for Age		Weight for Age	
	2012	2015	2012	2015
	(1)	(2)	(3)	(4)
Lottery winner	0.163**	0.110**	0.199***	0.140**
	(0.066)	(0.055)	(0.073)	(0.069)
Control group mean	0.099	0.258	0.012	0.182
N	1433	1939	1436	1946

Note – This table considers the impact of winning the lottery on the mean z-scores of anthropometrics measures, HFA and WFA, using data collected in years 2012 and 2015. All ITT estimated effects are from regressions that include strata fixed effects and controls for race and gender of the child. Robust standard errors are in parenthesis. Height and weight were standardized using World Health Organization growth standards to calculate HFA and WFA z-scores. As the WHO only has standardized weight for children up to 114 months, age =114 was imputed to all children older than 114 months in 2015 to avoid losing observations. For HFA z-scores, no imputation was carried out as the WHO standards are available for older ages. This imputation exercise does not introduce much distortion, as the same exercise for HFA generates very similar results (slightly higher point estimates).$*p \leq 0.1$, $**\,p \leq .05$, $***p \leq .01$.

Table 7 – ITT Estimated Effects on Children's Cognitive Function

	2012 Cognitive measures				
	Aggregate cognitive z-score	TVIP	Executive Function	Memory for Names	Visual Integration
	(1)	(2)	(3)	(4)	(5)
Panel A: 2012 mean z-scores					
Lottery winner	0.067**	0.112**	0.059	0.085	0.041
	(0.032)	(0.052)	(0.051)	(0.053)	(0.052)
N	1486	1466	1481	1476	1486
	2015 Cognitive measures				
	Aggregate IQ z-score	Verbal comprehension	Perceptual reasoning	Working memory	Processing speed
	(1)	(2)	(3)	(4)	(5)
Panel B: : 2015 mean z-scores					
Lottery winner	0.044	-0.011	0.091**	0.045	-0.006
	(0.043)	(0.043)	(0.044)	(0.045)	(0.045)
N	1999	1999	1999	1999	1996

Note – This table considers the impact of winning the lottery on the mean z-score for different measures of children's cognitive function in years 2012 (Panel A) and 2015 (Panel B). Panel A displays the aggregate cognitive z-score in column (1), the TVIP vocabulary test in column (2), the aggregate z-score of executive function tests in column (3), the z-score for the Memory for Names Test in column (4) and the z-score for the Visual Integration test in column (5). Panel B displays the aggregate IQ z-score in column (1), and its components in columns (2)-(5), respectively verbal comprehension, perceptual reasoning, working memory and processing speed. All scores have been standardized to have mean zero and σ one within age and within the sample. All ITT estimated effects are from regressions that include strata fixed effects and controls for race and gender of the child. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**\,p \leq .05$, $***p \leq .01$.

Table 8 – ITT Estimated Effects on Child Behavior

	Child Behavior Questionnaire					
	Aggregate CBQ z-score	Frustration	Attention	Soothability	Impulsivity	Inhibition
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: 2012 mean z-scores						
Lottery winner	0.001	-0.004	0.006	0.025	-0.081	0.061
	(0.052)	(0.053)	(0.053)	(0.053)	(0.054)	(0.053)
N	1483	1483	1483	1483	1483	1483
Panel B: 2015 mean z-scores						
Lottery winner	0.004	0.003	-0.036	-0.025	0.012	0.053
	(0.068)	(0.068)	(0.067)	(0.067)	(0.067)	(0.066)
N	923	923	923	923	923	923

Note – This table considers the impact of winning the lottery on the mean z-score for our measures of children's behavior based on the child behavior questionnaire in 2012 (panel A) and 2015 (Panel B): the aggregate CBQ z-score in column (1), and its components in columns (2)-(6), respectively frustration, attention, soothability, impulsivity and inhibition. Child behavior data for 2015 is only available for approximately half of the sample. All scores have been standardized to have mean zero and σ one within age and within the sample. All ITT estimated effects are from regressions that include strata fixed effects and controls for race and gender of the child. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table 9 – ITT Estimated Effects on Household Expenditures, Asset Index, Bank Account and Credit

	2008	2012	2015
	(1)	(2)	(3)
Panel A: Mean expenditures			
Lottery winner		27.551*	-5.132
		(16.193)	(16.340)
Control group mean		557	620
N		1439	1971
Panel B: Asset index z-score			
Lottery winner	0.066**	0.131**	0.041
	(0.031)	(0.052)	(0.035)
N	3762	1486	2049
Panel C: Access to bank account			
Lottery winner		0.071***	0.019
		(0.026)	(0.022)
Control group mean		0.57	0.59
N		1482	2045
Panel D: Access to credit			
Lottery winner		0.019	0.022
		(0.026)	(0.022)
Control group mean		0.43	0.42
N		1481	2042

Note – This table considers, for 2008, 2012 and 2015, the impact of winning the lottery on the mean household expenditures (Panel A), asset index z-score (Panel B), mean access to bank account (at least one household member with a bank account) (Panel C) and mean access to credit (at least one household member holding a credit card) (Panel D). All ITT estimated effects are from regressions that include strata fixed effects and controls for race and gender of the child. For the non-standardized measures we include the control group mean. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**\,p \leq .05$, $***p \leq .01$.

Table 10 – ITT Estimated Effects on Home Environment

	2008	2012	2015
	(1)	(2)	(3)
Panel A: Total time carer spends with child			
Lottery winner	-12.334***	-1.024	0.782
	(1.121)	(0.952)	(0.937)
Control group mean	55	60	55
N	3762	1482	2049
Panel B: Ever reads or sings for the child			
Lottery winner		0.065***	0.009
		(0.025)	(0.022)
Control group mean		0.63	0.47
N		1484	2048
Panel C: Number of children' books at home ≥ 8			
Lottery winner		0.036	0.013
		(0.024)	(0.021)
Control group mean		0.265	0.289
N		1482	2045
Panel D: Attitudes towards the child			
Positive attitudes			
Lottery winner		-0.023	-.001
		(0.015)	(0.012)
Control group mean		0.558	0.530
N		1484	2034
Negative attitudes			
Lottery winner		-0.013**	-0.002
		(0.006)	(0.006)
Control group mean		0.048	0.021
N		1483	1124
Panel E: Stress of the carer z-score			
Lottery winner	-0.079**	0.036	0.070
	(0.031)	(0.057)	(0.044)
N	3762	1486	2048

Note – This table considers, for 2008, 2012 and 2015, the impact of winning the lottery on i) Total time in weekly hours carer spends with the child (Panel A); ii) Probability of anyone in the household ever reading or singing for the child (Panel B); iii) Probability of the household having at least 8 children's books (Panel C); iv) Positive and Negative attitudes towards the child (Panel D), based on observational data reported by the enumerator and v) Stress of the mother z-score (Panel E), based on self reported data collected through the *Perceived Stress Scale* by (LUFT et al., 2007). All ITT estimated effects are from regressions that include strata fixed effects and controls for race and gender of the child. For all measures we include at the bottom of each panel the control group mean. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table 11 – Mediation Analysis for Cognitive Function Outcomes

	TVIP - 2012				WISC- PERCEPTUAL REASONING - 2015			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Mean z-score								
Lottery winner	0.096*	0.055	0.068	0.045	0.084*	0.064	0.087*	0.071
	(0.055)	(0.053)	(0.054)	(0.053)	(0.046)	(0.045)	(0.046)	(0.046)
N	1332	1332	1332	1332	1811	1811	1811	1811
Panel B: Controls								
INCOME-RELATED MEDIATORS								
Income 2008		0.0000954**		0.0000900**		0.00000119		0.00000834
		(0.0000447)		(0.0000428)		(0.0000511)		(0.0000496)
Income 2012		-0.0000594*		-0.0000588*				
		(0.0000343)		(0.0000340)				
Income 2015						0.0000417**		0.0000387**
						(0.0000184)		(0.0000182)
Assets 2008		0.137***		0.131***		0.0919***		0.0772***
		(0.0309)		(0.0305)		(0.0266)		(0.0262)
Assets 2012		0.164***		0.132***				
		(0.0337)		(0.0336)				
Assets 2015						0.0556*		0.0422
						(0.0297)		(0.0302)
Access to credit 2012		0.0277		-0.00165				
		(0.0599)		(0.0583)				
Access to credit 2015						0.192***		0.175***
						(0.0530)		(0.0526)
Access to bank account 2012		0.0906		0.0670				
		(0.0613)		(0.0603)				
Access to bank account 2015						0.0938*		0.0889*
						(0.0530)		(0.0526)
Expenditures 2012		0.000137		0.0000800				
		(0.000110)		(0.000108)				
Expenditures 2015						-0.000107		-0.000105
						(0.0000689)		(0.0000682)
HOME ENVIRONMENT MEDIATORS								
Total time with the child 2008			-0.000308	-0.00000891			0.000924	0.00108
			(0.000790)	(0.000790)			(0.000751)	(0.000743)
Total time with the child 2012			-0.000895	-0.00127				
			(0.00151)	(0.00149)				
Total time with the child 2015							0.00104	0.00150
							(0.00115)	(0.00114)
TV hours 2012			0.0938***	0.0826***				
			(0.0243)	(0.0236)				
TV hours 2015							0.00982	0.0170
							(0.0229)	(0.0229)
Ever sings or reads for the child 2012			0.0703	0.0379				
			(0.0629)	(0.0621)				
Ever sings or reads for the child 2015							-0.0626	-0.0686
							(0.0505)	(0.0498)
Number of children's books in the household 2012			0.253***	0.175***				
			(0.0437)	(0.0438)				
Number of children's books in the household 2015							0.213***	0.157***
							(0.0354)	(0.0357)
Positive attitudes towards the child 2012			0.256**	0.179*				
			(0.101)	(0.100)				
Positive attitudes towards the child 2015							0.266***	0.241***
							(0.0925)	(0.0909)
Negative attitudes towards the child 2012			-0.702***	-0.730***				
			(0.237)	(0.230)				

Note – This table presents the results for the mediation analysis for child cognitive function outcomes for 2012 and 2015. For each outcome reported, TVIP for 2012 and WISC-Perceptual reasoning for 2015, four columns are presented (keeping the sample constant): i) first column reports the ITT results, ii) second column presents the results including the income-related variables as mediators and iii) third column presents the results including home environment variables as mediators and iv) fourth column presents the results including all income-related and home environment mediators. Panel A reports the treatment coefficients and panel B reports the coefficients on the mediators. All specifications include strata fixed effects and controls for race and gender of the child. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table 12 – Mediation Analysis for Anthropometrics Outcomes

	HFA						WFA					
	2012			2015			2012			2015		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Mean z-score												
Lottery winner	0.146**	0.119*	0.115*	0.093	0.086	0.083	0.192**	0.173**	0.165**	0.139*	0.129*	0.127*
	(0.070)	(0.069)	(0.070)	(0.059)	(0.059)	(0.060)	(0.077)	(0.077)	(0.077)	(0.073)	(0.073)	(0.073)
N	1320	1320	1320	1775	1775	1775	1323	1323	1323	1783	1783	1783
Panel B: Controls												
INCOME RELATED MEDIATORS												
Income 2008		0.000274***	0.000236**		0.000108	0.0000845		0.000118	0.0000558		0.0000250	-0.00000708
		(0.0000995)	(0.0000946)		(0.0000680)	(0.0000699)		(0.000104)	(0.0000974)		(0.0000898)	(0.0000942)
Income 2012		0.0000694*	0.0000527					0.0000976**	0.0000328			
		(0.0000369)	(0.0000424)					(0.0000429)	(0.0000501)			
Income 2015					-0.00000338	-0.0000163			0.0744		0.0000846***	0.0000611*
					(0.0000250)	(0.0000270)			(0.0472)		(0.0000324)	(0.0000329)
Assets 2008			0.0607			0.0215			0.104**			0.0250
			(0.0399)			(0.0351)			(0.0507)			(0.0467)
Assets 2012			0.0710									
			(0.0436)									
Assets 2015						0.0636						0.0947*
						(0.0408)						(0.0511)
Access to credit 2012			-0.0369						0.104			
			(0.0856)						(0.0920)			
Access to credit 2015						0.0567						0.0460
						(0.0711)						(0.0817)
Access to bank account 2012			0.00798						0.0184			
			(0.0725)						(0.0882)			
Acess to bank account 2015						0.0304						0.0284
						(0.0706)						(0.0801)
Expenditures 2012			-0.000188						-0.0000821			
			(0.000127)						(0.000158)			
Expenditures 2015						-0.0000273						0.0000709
						(0.0000902)						(0.000107)

Note – This table presents the results for the mediation analysis for child anthropometrics for 2012 and 2015. For each outcome reported, HFA and WFA, three columns are presented for each year (keeping the sample constant): i) first column reports the ITT results; ii) second column presents the results including the income variables as mediators and iii) third column presents the results all income-related variables as mediators. Panel A reports the treatment coefficients and panel B reports the coefficients on the mediators. All specifications include strata fixed effects and controls for race and gender of the child. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

APPENDIX A. Differential Selective Attrition

Table A.1 – Differential Selective Attrition between all rounds of data collection

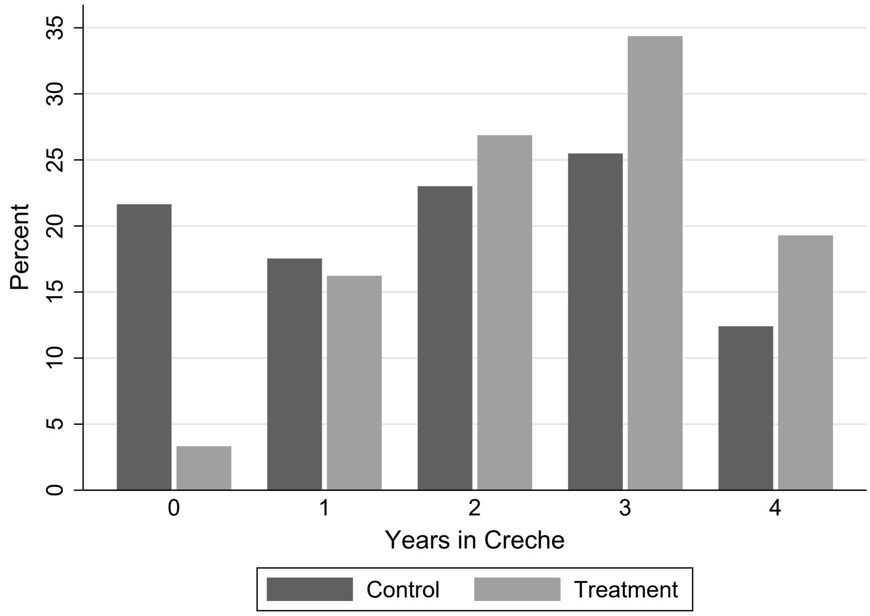
	Male child	White child	Black child	Mixed race child	Other race child	Birthweight in kg	Birth height in cm	Planned Birth	First Born	Age of the Mother at Birth	Prenatal Care
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: 2008 to registry											
Interviewed in 2008* Lottery winner	-0.035	-0.019	0.126**	-0.086	-0.020						
	(0.097)	(0.090)	(0.060)	(0.095)	(0.031)						
N	3897	3887	3887	3887	3887						
Panel B: 2012 to 2008											
Interviewed in 2012* Lottery winner	0.014	0.014	-0.004	0.005	-0.0147	-0.041	0.137	0.016	0.003	0.174	0.004
	(0.034)	(0.032)	(0.021)	(0.034)	(0.011)	(0.042)	(0.285)	(0.032)	(0.034)	(0.329)	(0.016)
N	3774	3764	3764	3764	3764	3742	3722	3770	3764	3767	3765
Panel C: 2015 to 2012											
Interviewed in 2015* Lottery winner	-0.0104	0.0534	-0.0226	-0.0405	0.00974	0.0618	-0.0487	-0.0621	-0.107*	0.0913	-0.0177
	(0.063)	(0.059)	(0.039)	(0.063)	(0.020)	(0.080)	(0.524)	(0.063)	(0.065)	(0.627)	(0.028)
N	1486	1486	1486	1486	1486	1404	1393	1411	1410	1408	1410
Panel D: 2015 to 2008											
Interviewed in 2015* Lottery winner	-0.010	0.042	0.002	-0.038	-0.005	0.004	0.255	-0.038	-0.032	0.025	-0.003
	(0.034)	(0.032)	(0.021)	(0.033)	(0.011)	(0.041)	(0.278)	(0.031)	(0.033)	(0.321)	(0.015)
N	3774	3764	3764	3764	3764	3742	3722	3770	3764	3767	3765

	Natural Birth Delivery	Premature Birth	Breastfed up to 6 Months	HH income	HH size	Age of carer	Carer can read and write	Carer has at least basic education	Carer has at least secondary education	Carer has at least higher education	Highest grade completed of carer
	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
Panel A: 2008 to registry											
Interviewed in 2008* Lottery winner				67.07	-0.212						
				(212.7)	(0.374)						
N				4103	4137						
Panel B: 2012 to 2008											
Interviewed in 2012* Lottery winner	-0.044	0.058**	0.076***	-96.57	-0.413	-0.190	-0.008	-0.018	0.005	-0.000	-0.140
	(0.032)	(0.023)	(0.029)	(154.1)	(0.291)	(0.637)	(0.011)	(0.031)	(0.033)	(0.009)	(0.168)
N	3768	3762	3770	3562	3592	3776	3768	3404	3404	3404	3346
Panel C: 2015 to 2012											
Interviewed in 2015* Lottery winner	0.017	0.020	0.051	478.0	-0.428	2.399*	0.024	-0.034	0.030	-0.002	-0.115
	(0.062)	(0.045)	(0.054)	(346.5)	(0.501)	(1.227)	(0.021)	(0.060)	(0.065)	(0.012)	(0.328)
N	1409	1409	1410	1402	1408	1412	1410	1270	1270	1270	1257
Panel D: 2015 to 2008											
Interviewed in 2015* Lottery winner	0.012	0.018	0.005	-13.30	-0.264	-0.034	-0.001	-0.034	-0.007	0.010	-0.103
	(0.031)	(0.022)	(0.028)	(150.5)	(0.284)	(0.622)	(0.011)	(0.030)	(0.032)	(0.008)	(0.164)
N	3768	3762	3770	3562	3592	3776	3768	3404	3404	3404	3346

Note – This table shows differential selective attrition results for the different waves of our surveys for 22 covariates. Each column reports results of a regression of each covariate on an indicator of whether there was individual data for a given year (relative to registry or a previous wave of survey) on an indicator of winning the lottery, and the interaction between them. All specifications include strata fixed effects. Robust standard errors are in parenthesis. $*p \leq 0.1, **p \leq .05, ***p \leq .01$.

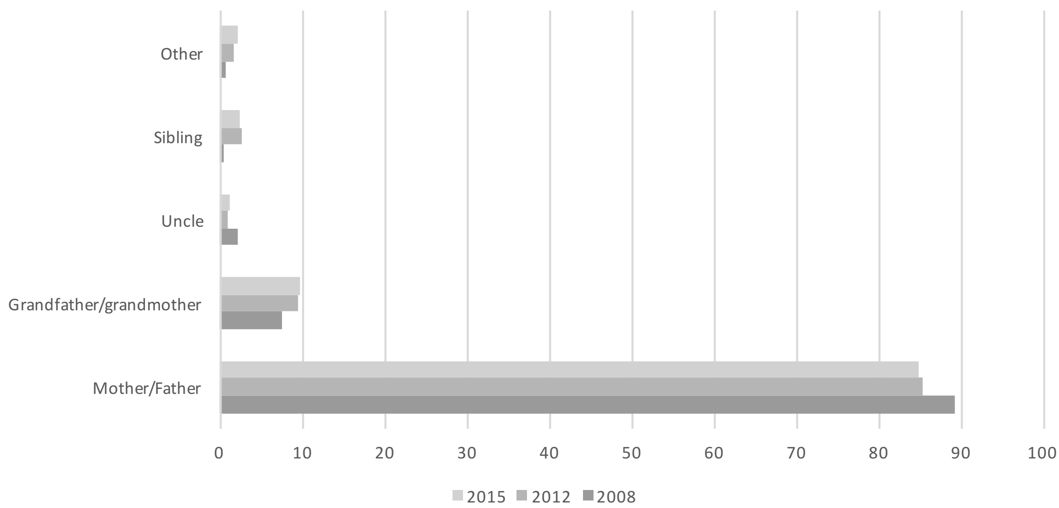
APPENDIX B. Figures

Figure A.1 – Years in Creche by Lottery Status



Note – This figure reports average years in *creches* by lottery status, based on self-reported survey data on *creche* attendance collected in 2012 and 2015

Figure A.2 – Identity of the Main Carer



Note – This figure displays the identity of the person reported as the main responsible for taking care of the child in 2008, 2012 and 2015.

APPENDIX C. Main Outcomes - Instrumental Variables

Table A.2 – IV: Impact of Creche Attendance on Income-Related Variables

	2008	2012	2015
	(1)	(2)	(3)
Panel A: Mean HH Income			
Treatment	100.812***	184.745**	102.291
	(32.430)	(85.245)	(90.278)
N	2287	1486	2049
Panel B: Mean expenditures			
Treatment		45.504*	-8.058
		(27.298)	(25.670)
N		1439	1971
Panel C: Asset index			
Treatment		0.218**	0.063
		(0.088)	(0.054)
N		1486	2049
Panel D: Access to Bank Account			
Treatment		0.119***	0.029
		(0.046)	(0.034)
N		1482	2045
Panel E: Access to Credit			
Treatment		0.032	0.035
		(0.043)	(0.035)
N		1481	2042

Note – This table reports IV estimates of the effect of an additional year of creche attendance (instrumented by lottery status) on household income-related variables (Panels A-E) for years 2008, 2012, and 2015, based on self-reported survey data from these years. All IV estimates are from regressions that include strata dummies and controls for race and gender of the child. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table A.3 – IV: Impact of Creche Attendance on Children's Cognitive Function

	Aggregate Cognitive Score	TVIP	WISC-Perceptual Reasoning
	2012	2012	2015
	(1)	(2)	(3)
Treatment	0.112**	0.191**	0.144**
	(0.054)	(0.090)	(0.070)
N	1486	1466	1999

Note – This table reports IV estimates of the effect of an additional year of creche attendance (instrumented by lottery status) on children's cognitive function for years 2012 (aggregate cognitive z-score and TVIP), and 2015 (WISC-Perceptual reasoning index), based on self-reported survey data from these years. All IV estimates are from regressions that include strata dummies and controls for race and gender of the child. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**\,p \leq .05$, $***p \leq .01$.

Table A.4 – IV: Impact of Creche Attendance on Anthropometrics Z-Scores

	Height for Age		Weight for Age	
	2012	2015	2012	2015
	(1)	(2)	(3)	(4)
Treatment	0.269**	0.170*	0.327***	0.217**
	(0.111)	(0.0873)	(0.125)	(0.109)
N	1433	1938	1436	1946

Note – This table reports IV estimates of the effect of an additional year of creche attendance (instrumented by lottery status) on children's anthropometrics in 2012 and 2015, based on self-reported survey data from these years. All IV estimates are from regressions that include strata dummies and controls for race and gender of the child. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**\,p \leq .05$, $***p \leq .01$.

Table A.5 – IV: Impact of Creche Attendance on Home Environments

	2008	2012	2015
	(1)	(2)	(3)
Panel A: Total time carer spends with child			
Treatment	-19.940***	-1.693	1.212
	(2.558)	(1.571)	(1.451)
N	2287	1482	2049
Panel B: Ever reads or sings for the child			
Treatment		0.109**	0.014
		(0.043)	(0.034)
N		1484	2048
Panel C: Attitudes towards the child			
Positive attitudes			
Treatment		-0.038	-0.001
		(0.026)	(0.018)
N		1484	2034
Negative attitudes			
Treatment		-0.022**	-0.004
		(0.011)	(0.010)
N		1483	1124
Panel D: Stress of the Mother Z-score			
Treatment	-0.120*	0.060	0.103
	(0.067)	(0.088)	(0.065)
N	2287	1486	2048

Note – This table reports IV estimates of the effect of an additional year of creche attendance (instrumented by lottery status) on children's home environments for years 2008, 2012 and 2015, based on self-reported survey data from these years. All IV estimates are from regressions that include strata dummies and controls for race and gender of the child. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**\,p \leq .05$, $***p \leq .01$.

Table A.6 – IV: Impact of Treatment on Labor Market Outcomes for carer and family members

	Carer			Parent or Step-parent		Grandparent		Sibling		Other HH members	
	2008	2012	2015	2012	2015	2012	2015	2012	2015	2012	2015
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Monthly income											
Treatment		101.5*	9.730	85.90*	72.41	404.9**	100.0	289.2	48.29	60.81	131.4
		(52.38)	(56.63)	(49.69)	(51.00)	(164.1)	(121.8)	(186.2)	(46.63)	(122.9)	(86.78)
N		1386	1946	2212	2985	438	478	244	557	615	667
Panel B: Currently employed											
Treatment	0.083**	0.078*	0.028	0.012	0.030	0.344***	0.112	0.400	0.116*	-0.109	0.054
	(0.033)	(0.044)	(0.037)	(0.030)	(0.026)	(0.120)	(0.087)	(0.247)	(0.064)	(0.087)	(0.053)
N	2283	1386	1939	2212	2978	438	475	244	555	615	664
Panel C: Weekly working hours											
Treatment	2.854*	3.902**	-0.046	1.299	0.067	19.80***	5.140	9.675	2.194	-0.309	3.616
	(1.463)	(1.944)	(1.675)	(1.637)	(1.432)	(7.499)	(4.252)	(9.898)	(2.807)	(4.729)	(3.123)
N	2283	1352	1874	2126	2811	415	430	241	527	572	593
Panel D: Contribution to social security											
Treatment		0.035	-0.013	-0.014	0.033	0.357***	0.102	0.362*	0.021	-0.112	0.065
		(0.042)	(0.038)	(0.036)	(0.031)	(0.119)	(0.094)	(0.214)	(0.048)	(0.094)	(0.057)
N		1385	1923	2209	2953	436	471	243	551	610	656

Note – This table reports IV estimates of the effect of an additional year of creche attendance (instrumented by lottery status) on labor market outcomes for carer and family members, based on self-reported survey data from 2008, 2012 and 2015. All IV estimates are from regressions that include strata dummies and controls for race and gender of the child. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

APPENDIX D. Main Outcomes for Balanced Panel

Table A.7 – ITT Estimated Effects on Household Income: Balanced Panel

	2008	2012	2015
	(1)	(2)	(3)
Lottery winner	70.29**	167.0***	90.81
	(29.57)	(59.60)	(82.96)
N	1080	1080	1080

Note – This table considers the impact of winning the lottery on the household income (in current reais) for years 2008, 2012, and 2015, based on self-reported survey data from these years, for the sample for which there is a balanced panel. All ITT estimated effects are from regressions that include strata fixed effects and controls for race and gender of the child. For all years the table displays the control group mean. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table A.8 – ITT Estimated Effects on Children's Cognitive Function: Balanced Panel

		2012 Cognitive measures			
	Aggregate cognitive z-score	TVIP	Executive Function	Memory for Names	Visual Integration
	(1)	(2)	(3)	(4)	(5)
Panel A: 2012 Mean z-scores					
Lottery winner	0.022	0.032	0.008	0.09	-0.039
	(0.037)	(0.060)	(0.060)	(0.061)	(0.060)
N	1105	1105	1105	1105	1105
		2015 Cognitive measures			
	Aggregate IQ z-score	Verbal comprehension	Perceptual reasoning	Working memory	Processing speed
	(1)	(2)	(3)	(4)	(5)
Panel B: 2015 Mean z-scores					
Lottery winner	0.034	-0.058	0.086	0.094	-0.029
	(0.059)	(0.062)	(0.062)	(0.060)	(0.062)
N	1105	1105	1105	1105	1105

Note – This table considers the impact of winning the lottery on the mean z-score for different measures of children's cognitive function in years 2012 (Panel A) and 2015 (Panel B), for the sample for which there is a balanced panel. Panel A displays the aggregate cognitive z-score in column (1), the TVIP vocabulary test in column (2), the aggregate z-score of executive function tests in column (3), the z-score for the Memory for Names Test in column (4) and the z-score for the Visual Integration test in column (5). Panel B displays the aggregate IQ z-score in column (1), and its components in columns (2)-(5), respectively verbal comprehension, perceptual reasoning, working memory and processing speed. All scores have been standardized to have mean zero and σ one within age and within the sample. All ITT estimated effects are from regressions that include strata fixed effects and controls for race and gender of the child. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table A.9 – ITT Estimated Effects on Anthropometrics: Balanced Panel

	Height for Age		Weight for Age	
	2012	2015	2012	2015
	(1)	(2)	(3)	(4)
Lottery winner	0.172**	0.148**	0.196**	0.125
	(0.079)	(0.074)	(0.086)	(0.0933)
N	1050	1050	1050	1050

Note – This table considers the impact of winning the lottery on the mean z-scores of anthropometrics measures, HFA and WFA, using data collected in years 2012 and 2015, for the sample for which there is a balanced panel. All scores have been standardized using the WHO growth standards. All ITT estimated effects are from regressions that include strata fixed effects and controls for race and gender of the child. Robust standard errors are in parenthesis. $*p \leq 0.1$, $**\,p \leq .05$, $***p \leq .01$.

APPENDIX E. Impact of Winning the Lottery on the Distribution of Main Outcomes

Table A.10 – Impact of Winning the Lottery on Probability of being above Quantiles of HH income

	2008	2012	2015
	(1)	(2)	(3)
Probability of being above 20%			
Lottery winner	0.023*	0.011	0.010
	(0.012)	(0.020)	(0.017)
Probability of being above 40%			
Lottery winner	0.041***	0.003	0.039*
	(0.015)	(0.025)	(0.021)
Probability of being above 60%			
Lottery winner	0.032**	0.040	0.033
	(0.015)	(0.025)	(0.021)
Probability of being above 80%			
Lottery winner	0.023*	0.053***	0.015
	(0.012)	(0.020)	(0.017)
P-value joint	0.076	0.062	0.392
N	3762	1486	2049

Note – This table shows the impact of winning the lottery on the probability of being above the 2nd, 4th, 6th and 8th deciles (Panels A-D respectively) of household income, using data collected in 2008, 2012 and 2015. Standard errors are in parenthesis. The estimates are from simultaneous regressions of dummies for being above each decile on lottery status, controls for race and gender of the child and strata dummies. Standard errors are in parenthesis. The p-value at the bottom of the table is for the null hypothesis that all coefficients are equal to zero. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table A.11 – Impact of Winning the Lottery on Probability of being above Quantiles of HFA and WFA Outcomes

	Height for Age		Weight for Age	
	2012	2015	2012	2015
	(1)	(2)	(3)	(4)
Probability of being above 20%				
Lottery winner	0.060***	0.008	0.012	0.032*
	(0.021)	(0.018)	(0.021)	(0.018)
Probability of being above 40%				
Lottery winner	0.028	0.056**	0.061**	0.041*
	(0.025)	(0.022)	(0.025)	(0.021)
Probability of being above 60%				
Lottery winner	0.037	0.046**	0.049**	0.065***
	(0.025)	(0.022)	(0.025)	(0.021)
Probability of being above 80%				
Lottery winner	0.013	0.045***	0.042**	0.010
	(0.020)	(0.017)	(0.020)	(0.017)
P-value joint	0.037	0.019	0.064	0.017
N	1433	1939	1436	1946

Note – This table considers the impact of winning the lottery on the probability of being above the 2nd, 4th, 6th and 8th deciles of the distribution of anthropometrics measures, HFA and WFA, using data collected in years 2012 and 2015. These estimates are of simultaneous regressions of dummies for being above each of these quantiles on the lottery status, controls for race and gender of the child and strata dummies. Standard errors are in parenthesis. The p-value at the bottom of the table is for the null hypothesis that all coefficients are equal to zero. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table A.12 – Impact of Winning the Lottery on the Probability of Being Above Quantiles of 2012 Cognitive Function z-scores

	2012 Cognititive Tests			
	TVIP	Executive Function	Memory for Names	Visual Integration
	(1)	(2)	(3)	(4)
Probability of being above 20%				
Lottery winner	0.038*	-0.015	0.034*	-0.013
	(0.020)	(0.020)	(0.020)	(0.021)
Probability of being above 40%				
Lottery winner	0.039	0.028	0.030	-0.009
	(0.025)	(0.024)	(0.025)	(0.025)
Probability of being above 60%				
Lottery winner	0.025	0.066***	0.030	0.005
	(0.024)	(0.024)	(0.025)	(0.024)
Probability of being above 80%				
Lottery winner	0.030	0.025	0.030	0.021
	(0.020)	(0.018)	(0.020)	(0.020)
P-value joint	0.266	0.020	0.408	0.693
N	1466	1481	1476	1486

Note – This table considers the impact of winning the lottery on the probability of being above the 2nd, 4th, 6th and 8th deciles of the distribution of 2012 cognitive function scores, using data collected in years 2012. These estimates are of simultaneous regressions of dummies for being above each of these quantiles on the lottery status, controls for race and gender of the child and strata dummies. Standard errors are in parenthesis. The p-value at the bottom of the table is for the null hypothesis that all coefficients are equal to zero.$*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table A.13 – Impact of Winning the Lottery on the Probability of Being Above Quantiles of 2015 Cognitive Function z-scores

	2015 Cognitive test: WISC				
	IQ	Verbal agee comprehension	Perceptual reasoning	Working memory	Processing speed
	(1)	(2)	(3)	(4)	(5)
Probability of being above 20%					
Lottery winner	0.010	0.003	0.029*	-0.002	-0.012
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Probability of being above 40%					
Lottery winner	0.031	-0.018	0.046**	0.044**	0.002
	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)
Probability of being above 60%					
Lottery winner	0.027	-0.003	0.063***	0.016	-0.016
	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)
Probability of being above 80%					
Lottery winner	0.009	0.011	0.007	0.013	0.009
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
P-value joint	0.597	0.620	0.013	0.071	0.460
N	1999	1999	1999	1999	1996

Note – This table considers the impact of winning the lottery on the probability of being above the 2nd, 4th, 6th and 8th deciles of the distribution of 2015 cognitive function scores, using data collected in 2015. These estimates are of simultaneous regressions of dummies for being above each of these quantiles on the lottery status, controls for race and gender of the child and strata dummies. Standard errors are in parenthesis. The p-value at the bottom of the table is for the null hypothesis that all coefficients are equal to zero.$*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

Table A.14 – Impact of Winning the Lottery on the Probability of Being Above Quantiles of CBQ z-score

	Child Behavior Questionnaire	
	2012	2015
	(1)	(2)
Probability of being above 20%		
Lottery winner	0.027	0.023
	(0.020)	(0.025)
Probability of being above 40%		
Lottery winner	-0.012	-0.003
	(0.025)	(0.031)
Probability of being above 60%		
Lottery winner	-0.028	-0.001
	(0.025)	(0.031)
Probability of being above 80%		
Lottery winner	-0.003	-0.011
	(0.020)	(0.025)
P-value joint	0.206	0.777
N	1483	923

Note – This table considers the impact of winning the lottery on the probability of being above the 2nd, 4th, 6th and 8th deciles of the distribution of CBQ scores, using data collected in years 2012 and 2015. These estimates are of simultaneous regressions of dummies for being above each of these quantiles on the lottery status, controls for race and gender of the child and strata dummies. Standard errors are in parenthesis. The p-value at the bottom of the table is for the null hypothesis that all coefficients are equal to zero. $*p \leq 0.1$, $**p \leq .05$, $***p \leq .01$.

APPENDIX F. Description of child assessments[19]

TVIP. The Peabody Picture Vocabulary Test (TVIP) is an exercise carried out with children in which the child is shown 4 pictures and is asked to identify an object named by the enumerator. The test has a duration of approximately 10 minutes and is composed of 125 different cards shown in an ascending order of difficulty. The test measures vocabulary development, which relates to cognitive development.

Woodcock-Johnson-Munoz battery of cognitive abilities. In test number 21, children are introduced to a series of space creatures with nonsensical names, and are then shown groups of space creatures from which they are asked to identify the ones previously presented. In test number 22, the purpose is to identify an object which has had its format distorted. These tests aim at measuring cognitive abilities, through the assessment of long-term retrieval and visual-spatial thinking.

WISC-IQ. Seven components of the Wechsler Intelligence Scale for Children, version IV (or WISC-IV), were administered to assess the following IQ dimensions:

- Verbal comprehension: In the Vocabulary subset children are shown pictures or told a word, and they are asked to name the object or define the word. It measures word knowledge and verbal concept formation. In the Similarities subset, children are presented with two similar objects or concepts and are asked to describe how they are similar. It measures logical thinking, verbal concept formation and verbal abstract reasoning.

- Perceptual reasoning: In the Block Design subset, while viewing a picture, the child uses blocks to replicate a printed image or modeled design. It measures the ability to analyze and synthesize an abstract design, and involves spatial visualization and analysis, simultaneous processing and visual-motor coordination. In the Matrix Reasoning subset, children are shown matrices or patterns with a missing part. The child is asked to select the missing piece among a range of options.

- Working memory. In the Digital span subset, several digits are presented in random order and the child is asked to recite the digits in the same order presented, or in reverse order. It measures short term memory and attention.

- Processing speed. In the Coding subset, the children are presented with boxes containing a numeral on the upper side of each box, and a symbol on the lower side. They must then write the symbol corresponding to each numeral in the worksheet. It measures processing speed, visual-motor dexterity, associative nonverbal learning, and nonverbal short-term memory. In the Symbol search subset, the child is asked to identify whether a target symbol appears among the symbols in the search group. It measures processing speed, short-term visual memory, cognitive flexibility and speed of mental operation.

[19] The information presented on this section is based on the information from the different tests' manuals, as well as from the webpages: https://www.helloq.com/overview/the-q-interactive-library/wisc-iv.html and https://www.thinktonight.com/WISC_IV_subtests_s/331.htm

Child Behavior Questionnaire. In this instrument, the enumerator reads out to the responsible for the child several statements that refer to the reactions of the child when facing different situations. For each item, the responsible for the child answers in a 1-7 true/false scale, taking into consideration the reaction of the child in the 6 months previous to the date of the interview. There are five sections to the instrument: i) Frustation, which is associated with the negative reactions to interruption of on-going tasks; ii) Attention focusing, related to the ability of maintaining attention focus upon tasks; iii) Soothability, which is related to rate of recovery from excitement or distress. iv) Impulsivity, which relates to speed of response initiation and v) Inhibitory control, which is related to the capacity to suppress inappropriate responses under instructions.

Executive Function Tests. For children aged 3-6, we applied Luria's tapping test and Day/Night Stroop. For children aged 6 and older, we applied the Heads Toes Knees and Shoulders test.

- **Luria's tapping test** The child is told to tap twice on the table with the pencil when the enumerator taps once, and vice-versa, several times. It measures inhibitory control and working memory.
- **Day and Night Stroop** This test is administered in two stages. In the first one, the examiner shows the child a dark card with a picture of the moon and asks the child to say "Day", then shows a white card with the sun and asks the child to say "Night". Then, sixteen cards are shown in the sequence, showing abstract patterns, but following the same logic that requires that they say the opposite ("Day" or "Night") of what the stimulus cards represent. It measures inhibitory control and working memory.
- **Heads Toes Knees and Shoulders** During this test, the children are given instructions to touch the toes when they are told to touch the head, and touch the knees when they are told to touch the shoulders, and vice-versa. It measures inhibitory control and working memory.

2 Are Public Schools Ready to Integrate Math Classes with Khan Academy?[1]

2.1 Introduction

Primary school enrollment in the different regions of the developing world has substantially increased over the past decades, but evidence shows that converting higher enrollment into improved human capital is a challenge. Overall, learning levels in developing countries remain critically low, with too many children and adolescents leaving school with insufficient literacy and numeracy skills (GLEWWE; MURALIDHARAN, 2016; WORLDBANK, 2018). Among the many different approaches for addressing educational deficiency, the use of technology-enhanced instruction has been growing in popularity as an approach for improving the quality of teaching and learning. Different interventions rely on a range of approaches, such as introducing computers and internet connection in public schools, distributing laptops to students and promoting the adoption of educational softwares that are able to deal with within-class heterogeneity in students' learning levels by delivering content adapted to each students' needs (BULMAN; FAIRLIE, 2016).

One of the most popular online platforms focused on delivering educational content tailored at each students' level is the Khan Academy, which offers free instructional videos and personalized exercises both in math as well as in other subject areas, ranging from kindergarten to college levels. While there are similar alternatives available in the market, Khan Academy stands out for its worldwide popularity, having reached 71 million of individuals in 190 countries since its foundation in 2008. Through partnerships with several organizations in different countries, Khan Academy has increasingly expanded its reach to different audiences in various languages. In this paper, we present the findings of the first randomized evaluation of the *Khan Academy in Schools* program, an effort to promote the use of the platform in Brazilian public schools.

The program was implemented in Brazil as a partnership between Khan Academy and the nonprofit Lemann Foundation, and its main feature was to integrate the Portuguese version of Khan Academy platform into math classes, once a week, in the the schools' computer lab. Our study measures the impacts of the intervention on math proficiency and attitudes towards math based on 5th and 9th grade students from 157 schools located across three different regions of Brazil. We analyze both average treatment effects, and also perform an exercise to estimate the heterogeneous effect of the program based on whether schools faced technology infrastructure challenges to program's implementation and whether they adopted the implementation modality based on an individual or rotational use of the computer during class.

Our findings show that *Khan Academy in Schools* had positive effects on measures of students' attitudes towards math, which were not translated to a positive average treatment effect on math proficiency. However, We find suggestive evidence that such null effect on students' test scores hides a positive effect in schools with better infrastructure to receive the program, but counterbalanced by negative effects in schools with worse infrastructure.

[1] This paper is co-authored with Bruno Ferman and Lucas Finamor.

While this paper is the first randomized evaluation of an effort to integrate the Khan Academy into regular class hours, there has been a series of studies investigating the effects of technology-enhanced instruction interventions in developing countries on learning outcomes. A review by Glewwe and Muralidharan (2016) shows the results are largely varied, with estimates ranging from significantly negative to significantly positive magnitudes. The available evidence suggests the characteristics of the computer-aided learning (*henceforth* CAL) interventions are an important factor to explain the heterogeneity of findings. Positive effects on learning are registered in studies mostly focused on programs that complement traditional teaching with CAL activities, such as Muralidharan et al. (2019), Banerjee et al. (2007), Lai et al. (2015), Linden (2008), Yang et al. (2013) and Mo et al. (2013). One common feature among all of these programs is that they increase the number of hours students are exposed to academic curriculum. However, when we consider the performance of CAL as an alternative for regular teaching, pulling students out of traditional class for classes that integrate CAL sessions, the limited available evidence presents mixed findings (BANERJEE et al., 2007; LINDEN, 2008; CARRILLO et al., 2011). In this context, our results are directly relevant as they shed light on potential reasons for the diverging results found in the literature on the effectiveness of CAL as a substitute for standard math classes. We find that details of program implementation are determinant for the performance of CAL programs as a replacement for traditional teaching pedagogy in developing countries. Therefore, assessing the adequacy of the implementation conditions and the technology infrastructure is crucial before scaling up such programs in a developing country context.

This paper is organized as follows. Section 2.2 describes the background and the program. Section 2.3 presents the experimental design. Section 2.4 describes our data and empirical strategy. Section 2.5 discusses the results and section 2.6 concludes.

2.2 Background and Context: Khan Academy in Schools Program

Khan Academy is an online interactive platform offering free instruction and practice in mathematics as well as other subjects, such as science, computer programming, history, economics, among others. The platform, originally created in the United States, offers contents in a personalized environment, adapting the user's experience to identify strengths and tackle learning gaps. The level of math contents available ranges from basic addition and subtraction to more advanced topics, such as differential equations and multivariable calculus.

Funded by volunteer contributions and partnerships with private sector foundations, the non profit initiative has greatly expanded over the years and currently reaches millions of students in over 190 countries. Khan Academy resources are available in 36 languages, and there are versions of the website in Spanish, French and Brazilian Portuguese. The Brazilian version of the platform was an joint effort between Khan Academy and Lemann Foundation, a Brazilian nonprofit focused at enhancing the quality of public schools in Brazil, which are mostly attended by children coming from lower income families. Focused on math education, the partnership translated the contents into Portuguese and reached 2.6 million students, which registered in the platform in the period of 2012 to 2017.[2]

[2] According to information reported o the Lemann Foundation's website *https://fundacaolemann.org.br/materiais/khan-academy-in-brazil*.

The platform may enhance students' math performance through three main channels. First, it may increase the quality of math content accessed by students by offering quality material developed by specialists. The second potential channel is by increasing students' learning through offering content and exercises adapted to each students' level, addressing students' heterogeneity within class. A third channel through which the platform may have an impact on a students' performance is by shifting the students' perceptions regarding math, turning the studying experience more attractive. By presenting the math content in an interactive and friendly way, designed to promote a fun and exciting learning experience, the platform may change the students' attitudes towards math, which may be ultimately translated into an increased math performance.

The Lemann Foundation has promoted the use of Khan Academy in Brazilian public schools through the program *Khan Academy in Schools*.[3] The program engages Government's Secretaries of Education which, after signing a participation agreement, receive the support from the Lemann Foundation to implement Khan Academy in schools. The 2017 edition of the program, which we evaluate in this paper, had three main pillars: i) delivering a one day training for Math teachers to present the platform and their functionalities; ii) advising teachers to carry out one of their weekly math classes (50 minutes per week) at the school's computer lab using Khan Academy and iii) close monitoring of intervention's implementation by Lemann Foundation staff, which acted as promoters of Khan Academy, providing assistance for solving any potential difficulties schools/teachers were facing. The program also allows teachers to have access to a detailed feedback report on students' performance, indicating their strengths and weaknesses.

The implementation of Khan Academy requires a good technology infrastructure, including a sufficiently high-speed internet connection. To guarantee an adequate implementation of the program, schools that had less than 0.5 computer per student were granted additional ones from the Lemann Foundation. For the evaluation sample, we can observe two different modalities of program implementation: i) individual use of the computer and ii) rotational usage of the computer between two students. In the rotational mode, each student used the computer during half of the class, and was assigned by the teacher other math activities during the remainder of the class. There was also information technology support for schools in the city of Manaus, which had weaker baseline infrastructure, to guarantee that the computers and internet were functioning. Since we are not interested in the effects of such improvements in the computer lab *per se*, all schools, irrespective of treatment status, received these benefits.

2.3 Study Design

2.3.1 Sample Selection

This experiment was conducted in primary public schools of five cities in three different regions of Brazil for the 2017 school calendar year. The cities of Barueri, Mogi das Cruzes and Sao Bernardo do Campo were selected from the Southeast region; Pelotas from the South; and Manaus from the North region. Cities were selected based on previous relationship between the city government and the implementing partner (Lemann Foundation), and conditional on the existence of a satisfactory level of municipal school infrastructure (existence of a computer lab and internet connection).

[3] "*Khan Academy nas Escolas*", later renamed to "*Innovation in Schools*" or "*Inovação nas escolas*".

In the five cities selected, all primary education schools were invited to voluntarily apply to the program. Among all applicants, the Lemann Foundation determined a final list composed of 166 schools that were initially eligible to participate in the treatment randomization. Out of these, before the treatment was assigned, nine schools left the evaluation sample due to lack of the necessary infrastructure or because they did not have a matching pair to compose a stratum. This resulted in 157 schools in the final evaluation sample.

2.3.2 Experimental Design

The study took place in 157 primary education schools.[4] Schools may be of three different types, based on the grades they offer: (a) Cycle I schools, which offer grades 1-5 (students between 6-10 years old); (b) Cycle II schools, which correspond to 6th-9th grades (students between 11-14 years old); and (c) Both cycles schools, which have students from 1st to 9th grades (students aged 6-14 years old).

Schools were initially stratified based on four criteria: i) the municipality they were in; ii) type of school in terms of the grades they offered (cycle I, cycle II or both cycles); iii) whether they had ever received the Khan Academy program in the years preceding the experiment;[5] and iv) whether Math proficiency data for the 2015 national standardized exam was available. For the cases in which the resulting strata were composed of more than 5 schools, further stratification was carried out based on the math scores for the standardized national exam, conditional on data availability.

Every school in our sample was assigned at least one treatment and one control grade, with the purpose of increasing engagement and reducing attrition. This study is based on students from the 5th and 9th grades, since for these grades there is a national standardized exam every two years and math proficiency data would be available for the 2017 academic year. For Cycle I schools, 3rd (or 4th) and 5th grades were eligible to receive the program, and we randomized treatment in the 5th grade. Schools assigned as controls in the 5th grade automatically received treatment in the 3rd or 4th grade. Similarly, for Cycle II schools, 6th and 9th grades were eligible, and treatment in the 9th grade was randomly assigned. For schools assigned 9th grade as control, the 6th grade received the intervention. Schools with both cycles had only the 5th and 9th grades eligible, and similar procedure was followed. Randomization allocated which grade would receive treatment.

The 157 schools in our study were divided into 35 strata (which had from 2 to 11 schools each). Since Cycle II schools had both 5th and 9th grades participating in the study, our sample is composed of a total of 217 grades in 47 strata-grade pairs.

[4] There were 29 schools in Pelotas, 63 schools in Manaus, 21 schools in Barueri, 27 schools in Mogi das Cruzes and 17 in Sao Bernardo do Campo.

[5] In our evaluation sample, only 14 schools in the city of Pelotas had Khan Academy implementation in the previous years. Students in our experiment sample, however, were never exposed to the Khan Academy platform in school. In Section 2.5.1 we check whether control students were even exposed to the platform.

2.4 Data and Empirical Strategy

2.4.1 Data

Data for this study stems from two main sources. First, we use survey data collected over two rounds: a baseline carried out in March 2017, before the beginning of the program, and a follow-up in November 2017, right before the end of the school calendar year. Baseline data was not collected for one municipality (Sao Bernardo do Campo). We collected data for an instrument that measured students' attitudes towards mathematics (BRITO, 1998). This instrument was composed of a questionnaire with 20 questions that presented different statements about an individuals' feelings regarding Math, with Agree/Disagree four point Likert Scale answer options. The different statements express either a positive or a negative connection with Math (such as *"Mathematics is enjoyable and stimulating to me"* or *"Mathematics makes me feel uneasy and confused"*).[6] An index for attitudes towards math was created by summing up all scores for positive statements, and adding the reverse score for negative statements, and then standardized to have zero mean and standard deviation one within the control group, by grade level.[7] We also collected data on students' demographic characteristics, students' self reported access and usage of computer and internet both at home and at school as well as their preference in relation to school subjects. On the follow-up survey, information on the knowledge and usage of Khan Academy was also collected to assess program compliance and contamination in the control group. Survey data is not available for 7 out of the 157 schools, which left the study after treatment assignment.

Our second data source is administrative data from the 2017 Ministry of Education's Basic Education's Evaluation System (*Sistema de Avaliacao da Educacao Basica - SAEB*). Every two years, at the end of the school calendar year, the government implements standardized exams to measure students' academic proficiency in the 5th and 9th grades, compulsory for all Brazilian public schools with 10 or more students. The SAEB exam also collects data on students' characteristics, including demographics, household characteristics, leisure and studying habits, parents' education, employment status and school retention record. Data on teachers' characteristics is also collected, including age and educational level. Although this exam is implemented in all public schools in Brazil, the Ministry of Education only releases proficiency data for those school grades that had at least 80 percent of enrolled students taking the test. We have administrative data for all schools in our sample (including those that left the study after treatment assignment), with the exception of those school grades that did not meet the minimum attendance requirement. Unfortunately, we are not able to link individual level administrative data with survey data because the SAEB dataset is de-identified.

We also use information extracted from the Khan Academy platform on the usage of Khan Academy by treated students. This information is useful for a descriptive view of the implementation of the program, and it is not available for students in the control group.

6 This measure was originally developed by Jr and Dreger (1961) and translated and validated to Portuguese by Brito (1998). See the original papers for the full list of questions.

7 An answer of 4 in a negative statement was recoded into 1 to reflect the reaction to an opposite positive statement, and so on.

2.4.2 Balance and Attrition

2.4.2.1 Survey

Table 1 presents survey student level baseline characteristics for the pooled sample and for the samples of the 5th and 9th grades separately. For each group, the table displays three columns respectively with the control group mean, the regression adjusted differences between treatment and control groups, and number of observations for 27 covariates. We report estimates from a regression for each covariate on an indicator variable for the treatment and strata-grade fixed effects, with standard errors clustered at the school level. The results demonstrate randomization was successful as characteristics are balanced across treatment arms (the *p*-value of a joint test that there is no difference between treatment and control for all baseline covariates is equal to 0.695, 0.453 and 0.720 respectively for the three samples considered).

There are two potential sources of attrition in the survey, school-level and student-level attrition. Our first source of attrition is associated with schools that left the program after treatment assignment. Seven schools out of our sample of 157 schools - both in treatment and control groups - left the study after randomization took place for various reasons, mostly unrelated with treatment assignment. The small number of school dropouts and the different reasons associated with the withdraw minimize our concerns with differential selective attrition. Two out of seven schools left the program after randomization and previously to the communication of treatment assignment. Out of the other 5 schools that dropped out, only 2 dropped out due to problems with the treatment assignment (one school assigned treatment in the 5th grade and one school assigned control in the 5th grade), and one school due to lack of teachers' engagement. The remaining 2 schools left the program due to unavailability of the computer lab and absence of computer lab instructor. Student-level attrition in the survey is related to students either not being present in class during the survey application or failing to complete the answers for the attitudes towards math instrument.

In Table 2 we show attrition results for our different measures of attrition. We report the control group mean, regression adjusted differences between treatment and control groups, the number of observations and number of clusters, for the pooled sample, and for the 5th and 9th grades subsample respectively.[8] In Panel A, we show that survey attrition rate (attrition defined by the absence of data on attitudes towards math) was relatively high, at almost 40% for the pooled sample in the control group. High survey attrition is relatively common in studies that collect data in Brazilian public schools at the end of school year, as it is not atypical for school attendance in Brazil to drop significantly during the last month of classes. Attrition in treatment group is 2.5 percentage points lower than that in the control group (*p*-value=0.027). In Appendix Table A.1, however, we show covariates remain balanced between treatment and control groups even after conditioning on the sample of non attritors in the follow-up survey round. This suggests that the significant differences in attrition rates are unlikely to generate differential selective attrition that could threaten the validity of our results.

---

[8] The dependent variable is an indicator whether there is no outcome data available.

2.4.2.2 SAEB Data

Table 3 shows covariates are also balanced for characteristics reported in the SAEB data set, confirming there are no significant differences between treatment arms in none of the samples considered.

There are two potential sources of attrition in the SAEB dataset: i) school-grade-level attrition, since proficiency data is only released by the Ministry of Education for those school-grades that had at least 80% of student attendance in the exam and ii) student-level attrition for those students that did not take the SAEB exam. In Panel B of Table 2, we show school-grade level attrition results for the SAEB exam. For this dimension, we define attrition as the absence of math proficiency data in the SAEB exam, at the school-grade level. There are no significant differences in attrition rates between treatment and control groups for the math proficiency outcome, for the pooled sample, and for the 5th and 9th grades separately. The results show that the intervention is not correlated with the likelihood of the schools having SAEB data reported. In Panel C, we use student-level data in the SAEB exam to show that there are no differences between treatment and control groups on the proportion of students taking the SAEB test (for those grades that had the results reported).

2.4.3 Empirical Strategy

The experimental design generated random variation on which school $\times$ grades had their teachers assigned to receive a Khan Academy training from the Lemann Foundation, and to use the Khan Academy platform integrated to one math class every week (around 50minutes per week). The assignment to the treated group also involved frequent visits from Lemann foundation staff, which followed up on treated grades' usage of the platform, solved any potential difficulties and acted as promoters of Khan Academy usage. We define the "treatment" as the teacher being assigned to receive this training and follow up from the Lemann Foundation, and the class being assigned to use the Khan Academy platform as recommended in the intervention, which was expected to last for approximately 28 weeks.

It is not possible to guarantee, however, that all teachers followed the exact plan of the intervention (that is, substituting one traditional math class per week for the Khan Academy for the treated grades). Moreover, while every school in the sample had at least one treatment and one control grades, and every school declared they were committed to avoid control grades' usage of the platform, the Khan Academy platform is free and openly available. It is, therefore, possible, although improbable, that control students and teachers were using it. For these reasons, our estimates should be considered as an intention to treat effect (ITT) of the intervention. In subsection 2.5.1 we show that contamination to the control students was minimal, and that the intervention significantly increased the exposure of treated school students to the Khan Academy platform.

Our ITT estimates are based on the following regression:

$$y_{igs} = \alpha + \beta_{\text{ITT}} Z_{igs} + \Gamma \mathbf{X}_{igs} + \epsilon_{igs}, \tag{2.1}$$

where y_{igs} is an outcome of interest for individual i, who belongs to grade g in a school s, Z_{igs} is an indicator variable that takes value 1 if individual i belongs to a treated school-grade, $\mathbf{X}_{igs}$ is a set of baseline controls, which includes strata fixed effects, and ϵ_{igs} is an error term. β_{ITT} is the average treatment effect of the program. We report both results pooling 5th and 9th grades (in which case we interact the

strata fixed effects with grade), and separately for each grade. Standard errors are clustered at the school level.

In this paper, we consider two main outcomes: math proficiency and attitudes towards math.[9] Our math proficiency results are based on the SAEB data, which covers all schools of our sample, including the 7 schools that left the study after treatment assignment (although excluding the school-grades for which data was not released). For attitudes towards math, we rely on survey data, for which we only have information for the subsample of compliers (150 schools). All scores were standardized to have zero meanand standard deviation one within the control group, by grade level.

2.5 Results

2.5.1 Program Implementation and Compliance with Experimental Design

Before presenting the treatment effects on the main outcomes of interest, we present in this section evidence that the students allocated into treatment group were exposed to Khan Academy, and that we find no evidence of contamination in the control group. Table 4 shows results for the follow-up survey which, in addition to collecting data on attitudes towards math, gathered information on other variables, such as student's familiarity with Khan Academy, reported use during school, use of computer and preferences regarding subjects. The table displays, for the pooled sample and 5th and 9th grades separately, the control group mean, the regression adjusted differences between treatment arms and the number of observations for different variables collected on the follow up survey round.

Our results show that around 97% of the students in treated grades report using Khan Academy (around 82% report using it in school). In the control group, only 6.3% of the students report using the platform (4.4% report using in school), so contamination does not raise major concerns. Considering the 5th and 9th grades separately, we observe that proportion of students reporting use of Khan Academy is slightly lower for the 5th grade (96% in the 5th grade as opposed to 98% in the 9th grade).

The intervention increased the probability that students report using the computer lab at schools, both during and outside class. The coefficient for using the computer lab during math classes is very large (0.445) and significant, as expected. There is evidence that the intervention has not substantially crowded out other school activities happening in the computer lab, as the results suggest the probability of using the computer lab in other classes decreased by a very small magnitude (-0.055) relative to the increased use during math class. The intervention also increased the probability that students report using the school computer lab not during classes, which is consistent with treated students using Khan Academy even after school hours. While we do not find an increase in the proportion of students who use computer at home, this does not imply that treated students are not using Khan Academy at home, as the program may have increased the probability of using Khan Academy at home for those who report frequently using computer at home regardless of the treatment status.

While virtually all treated students were exposed to platform, many schools experienced some implementation problems during the program. Lemann Foundation's staff visited all schools five times throughout the school year, and during these visits they collected information on the usage of the Khan

[9] Math proficiency and attitudes towards math were the main outcomes registered in the paper's pre-analysis plan.

Academy platform. In about 31% of those visits, they reported that the implementation was inadequate. In 71% of those cases with inadequate implementation was due to infrastructure problems. Of those cases with infrastructure problem, around 78% was due to internet connectivity problems, while around 15% was due to problems with the computers. Overall, 51% of the schools reported inadequate implementation due to infrastructure problems in at least one month. Around 7% of the cases with inadequate usage were because there were no math teachers during that period, and around 5% of the cases were because teachers were not motivated with the project. Another important information collected by Lemann Foundation's staff was about the modality of implementation in terms of number of students per computer. In around 37% of the schools, there was one computer for each student, so that students could spend the whole math class in the platform. For the other schools, there was a rotation system, in which students would use Khan Academy for half of the class, and work on other math-related activities for the remainder of the class.[10] In only 1% of the cases, more than one student shared the same computer. Teachers were advised not to let that happen, because this would undermine the effectiveness of one of Khan Academy's main feature, which is its adaptive learning nature that tailors the content according to each student's needs.

Such implementation issues had important consequences for the total time of exposure to the platform. In columns 1 to 3 of Table 5, we show how the total number of minutes logged in the platform correlates with infrastructure problems and with the type of implementation. In schools that implemented the program with rotation and had infrastructure problems, 5th graders spent 540 minutes logged in the platform from April to October.[11] When a school did not present internet problems, 5th graders spent approximately 30% more minutes in the platform, while in schools with one computer per student 5th graders spent 42% more minutes. 9th graders spent substantially fewer minutes in the platform relative to 5th graders, spending a total of 386 minutes in schools with infrastructure problems and with rotation. This number was 48% higher in schools with one computer per student, but no higher in schools with no infrastructure problems. Interestingly, even in schools with one computer per student, the total number of minutes for 9th graders is still only about the same as the total number of minutes for 5th graders in schools with infrastructure problems and rotation. We also present in columns 4 to 6 of Table 5 the number of weeks students logged in the platform. We also find that 5th grade students logged in more weeks than 9th graders, and that 5th graders in schools with no infrastructure problems logged in more times. However, there is no significant difference in the number of weeks logged in for schools with one computer per student, suggesting that the larger number of minutes in such schools come mainly from the intensive margin of usage.

2.5.2 Treatment Effects on Main Outcomes

Table 6 shows intent to treat estimates of the program on math proficiency (columns 1-2) and attitudes towards math (columns 3-4) for the pooled sample (Panel A), and for the 5th and 9th grades separately (Panels B and C). The first column for each outcome omits the covariates from the regression specified in Equation 2.1. On average, we find no differences in math proficiency between students

[10] There is no information on the type of implementation for 9 out of 150 schools. For these schools, the staff from the Lemann Foundation did not collect this information during the visits.

[11] We consider usage from the beginning of the implementation until the SAEB exam. If we considered until the end of the school year, then these students would have a total of 687 minutes in the platform.

attending grades assigned to treatment and control groups. In this dimension, there is no effect of the program on average for the pooled sample or for the 5th and 9th grades individually.

Our results also indicate that students attending treatment grades had slightly higher, and significant, scores in the attitudes towards math index (0.060σ for the pooled sample, 0.062σ for the 5th grade and 0.057σ for the 9th grade, for the specification including covariates). Our initial hypothesis was that one of the channels through which the program could foster math proficiency was by improving the students' math learning experience. This hypothesis was based on the assumption that, by learning math in a more exciting and interactive manner, students would have better attitudes regarding math, potentially paying more attention on the exposed content or even spending longer hours studying it, which could ultimately impact proficiency. While we confirm that the intervention has a positive impact on the attitudes towards math, the effects were very small, and our findings suggest the modest gains in attitudes were not translated into higher math proficiency on average.

There are a few factors that may have prevented positive average treatment effects from arising. First, one important aspect to note about the intervention is that, although it exposes students to a potentially more engaging learning experience, it does so by integrating Khan Academy into one of the weekly math classes, so students' total exposure to traditional methods of teaching is reduced. Also, the Khan Academy class was carried out at the schools' computer lab, and there is anecdotal evidence that a significant proportion of class time was wasted moving the students to a different location. Second, the implementation of the program faced some challenges, as 51% of the schools reported infrastructure problems in at least one month of implementation. Lastly, the different types of implementation (individual *vs* rotational use of the computer) may have played an important role. Our data shows that implementation was based in rotation in 59% of the treated schools in the 5th grade and in 55% of the schools treated in the 9th grade.

2.5.3 Treatment Heterogeneity

If the null effect we estimated for students' test scores comes from infrastructure problems and/or from a implementation modality based on rotation of students, then we should expect to find positive effects in schools that had a better implementation. While we do not have experimental variation on whether schools experienced infrastructure problems, or on whether they implemented the program with one student per computer, we take advantage of the fact that all schools implemented Khan Academy in at least one grade and use school-level implementation information that covers our entire sample to perform a heterogeneity exercise. Following our instructions, Lemann Foundation staff visited all schools in our sample, collecting data on implementation in all schools in exactly the same way, irrespective of the grade that received the program.

Given that, within each school, we extrapolate the information on infrastructure problems and type of implementation from the treated to the control grade so that we can use these variables to estimate whether the treatment effect was different depending on these implementation variables. Such empirical strategy relies on the assumption that, within each school, grades that were not assigned to receive treatment would have had the same quality and modality of implementation as grades that were treated. This assumption could be invalid if, for example, school principals put more effort in guaranteeing that the infrastructure is working well when the program is assigned to one of the grades that will be evaluated

in the SAEB exam. Alternatively, the type of implementation may depend on the grade if grades have substantially different number of students.

In Table 7, we provide evidence that this is not the case. In Panel A, we show the results of a school-grade-level regression of a dummy variable that takes value one if the there are no infrastructure problems on the treatment indicator and strata fixed effects. In columns 1-2, we display the results for 5th and 9th grades for all schools. For example, the results presented in column 1 compare the proportion of schools with no infrastructure problem in the 5th grade control schools (so this information come from implementation in the 3rd, 4th, or 9th grades in these schools) to this information for 5th grade treated schools (so this information comes from implementation in the 5th grade). Columns 3-4 and 5-6 show estimates for 5th and 9th grades in two cycle and one cycle schools respectively. In Panel B, we perform the same exercise using an indicator of one computer per student as a dependent variable. None of the estimated coefficients are significant, providing support to the validity of the assumption our extrapolation exercise relies on. Standard errors are not reported for the 9th grade in the subsample of one cycle schools, as the dependent variable reflecting good infrastructure was equal zero for all 14 schools in this group. In Appendix Table A.2, we also show that, controlling for school fixed effects, the number of students per classroom does not significantly vary by grade, providing further evidence that the type of implementation should not be dependent on the grade that received the program in a given school.

Table 8 presents the results for the heterogeneity exercise. Columns 1-2 show the heterogeneity results for math proficiency, while columns 3-4 display the results for attitudes towards math. Our results show that the integration with Khan Academy maybe an effective alternative to traditional curriculum if adequately implemented. Students assigned in treated grades that did not face infrastructure problems had significantly higher math scores (0.058σ), and gains were also registered when the modality of implementation was one computer per student (0.081σ). On the other hand, students assigned to grades that implemented the rotational modality of the program performed worse in the SAEB exam (-0.076σ), which may not be surprising if this type of implementation ultimately leads to a reduction in students' total math exposure.

Our results are mostly driven by the 5th grade subsample, which experienced larger than the average gains both for students assigned to treated grades that faced no infrastructure problems (0.093σ) and for students assigned to the individual use of the computer modality (0.127σ). In the 5th grade, negative effects on math scores were registered for students in the poorer implementation group, but only statistically significant for the group that implemented with rotational use (-0.082σ). For the 9th graders, no significant differences are found, and all estimated coefficients are negative. These findings are consistent with results from Table 5, where we show 9th grades did not have a large exposure to the platform, even in schools with good implementation.

Columns 3-4 of Table 8 present the heterogeneous effects on students' attitudes towards math. In all three panels, standard errors are relatively large, and we cannot reject the null hypothesis that the effects are the same for schools with better and worse implementation (for the pooled sample, p-values equal to 0.478 for the heterogeneity with respect to no infrastructure problems and 0.723 for type of implementation).

It is possible to rationalize the heterogeneous effects on students' math proficiency and the (lack

of) heterogeneous effects on attitudes towards math if we consider that virtually all treated students were exposed to the platform, regardless of the quality and type of implementation. However, students in the rotation implementation had to split one of their weekly classes between studying in the platform and doing other math activities. If there are returns to scale in spending more time in one activity, these math activities are not as effective as standard math classes, and/or there is relevant time wasted in the transition from one activity to the other, then the implementation of the program in these schools may have actually reduced the total amount of math content that these students were exposed to, relative to a setting with no intervention. Moreover, students in schools with the rotation system spent significantly less time in the platform. Likewise, students in schools with infrastructure problems were also exposed to the platform. However, they spent significantly less time in the platform relative to schools with no infrastructure problems. Moreover, it is conceivable that some classes were wasted trying to connect to the internet without success, which again could have reduced the total amount of math content that these students were exposed to. Therefore, these heterogeneous patters can be rationalized in a model in which perceptions about math can be affected by exposing students to a more attractive way to present math content, regardless of whether such exposure comes at the expense of a reduction in standard math classes. Moreover, the extensive margin with respect to exposure to the platform may be more relevant in shaping such views about math relative to the intensive margin of usage. This may explain the lack of heterogeneous effects on attitudes towards math. When we consider the effects on students' math proficiency, however, then this reduction in standard math classes and/or the intensive margin of exposure to the platform may be more relevant, so we find heterogeneous effects depending on the quality and type of implementation.

2.5.4 Discussion

Combining our results with the available evidence on CAL programs suggest that the effectiveness of such programs depend crucially on a series of implementation details. A first important implementation issue regards whether the CAL program complements standard math classes with extra hours of study, or whether it replaces standard math classes. In the second case, the effect of a CAL program depends crucially on the net effectiveness of the CAL program relative to a standard math class. This helps explain why the literature converged in pointing out the benefits of CAL programs in supplementing traditional teaching, while there is mixed evidence on the potential for CAL as effective substitutes (for a review of the literature see, for instance, Glewwe and Muralidharan (2016) or Bulman and Fairlie (2016)).

When we consider the evidence on CAL programs as substitutes for standard math classes, our results help rationalize the mixed evidence found in the literature. We show that the quality and type of implementation are important determinants of whether such programs should have positive or negative effects. Importantly, since in this case the impact of the program depends on the net effectiveness of the CAL program relative to a standard math class, it is possible that the impact of the program is negative when the implementation is inadequate. In our study, we show that this can be the case when students have to rotate between the CAL activity and other math activities, and when infrastructure problems in the school prevents a more extensive usage of the platform. In contrast, CAL programs implemented as complements should be less likely to generate negative results when there are implementation problems.

Overall, these results point out that the external validity of experimental results on CAL programs

should be considered with caution. In this sense, we see our heterogeneity results as an important contribution to the literature in that it provides evidence on some key determinants that are relevant in the extrapolation of experimental results on CAL programs.

Given this discussion, we stress that the results we present on the effects of the Khan Academy platform should be viewed as the effects of this platform integrated to math classes, with a specific type and a given quality of implementation. Given the available evidence, we should expect different results if we considered different types of implementation of the Khan Academy platform, or if we considered a setting with better infrastructure.

2.6 Conclusion and Policy Implications

In this paper, we present novel experimental evidence on the impacts of the Khan Academy platform, through the program *Khan Academy in Schools*, implemented across five cities in three different regions of Brazil. The program aimed at integrating one weekly math class (50 minutes) with a Khan Academy session in the computer lab. We find that the program does not have an impact on average over students' math scores, although we find small but significant effects on attitudes towards math. We also explore treatment heterogeneity by quality of implementation, showing that the program has positive effects when there are no infrastructure problems and when the implementation modality is based on one computer per student. However, it can have negative effects in settings with implementation problems, or in which the implementation modality is based on rotation.

The available evidence points out that computer assisted learning (CAL) programs are very beneficial when they are delivered supplementing the traditional school curriculum. As highlighted by Muralidharan et al. (2019), mode of delivery is important, and effectiveness of CAL programs may vary depending on whether these are implemented in substitute or supplementary manners, in-school or out-of-school. Evidence on the effectiveness of CAL programs as substitutes for teacher delivered curriculum is limited, and the available evidence is not conclusive. Our results contribute to the debate on this issue. We show that implementation challenges may prevent positive treatment effects from arising and that, when adequately implemented, CAL programs may be effective even when it does not increase the total number of hours of exposure to math content. Our conclusion is that details of program implementation matter, and these must be taken into account when considering scaling up of CAL programs as an alternative for traditional teaching pedagogy in developing countries.

TABLES

Table 1 – Baseline Covariates Balance - Survey

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
Attitudes towards math	0.000 [1.000]	0.004 [0.030]	11422	0.000 [1.000]	-0.007 [0.035]	7203	0.000 [1.000]	0.024 [0.059]	4219
Male	0.505 [0.500]	-0.005 [0.009]	12369	0.513 [0.500]	-0.015 [0.010]	7871	0.488 [0.500]	0.012 [0.016]	4498
Year of Birth	2004.614 [2.298]	-0.010 [0.027]	12381	2005.911 [1.396]	-0.053 [0.040]	7872	2001.820 [1.013]	0.066 [0.043]	4509
White	0.327 [0.469]	-0.014 [0.009]	10703	0.364 [0.481]	-0.028 [0.014]	6540	0.256 [0.437]	0.008 [0.015]	4163
Black	0.107 [0.309]	-0.013 [0.006]	10703	0.111 [0.314]	-0.010 [0.008]	6540	0.100 [0.300]	-0.017 [0.012]	4163
Indian	0.038 [0.192]	0.002 [0.004]	10703	0.041 [0.198]	0.004 [0.005]	6540	0.033 [0.180]	0.000 [0.006]	4163
Mixed	0.488 [0.500]	0.026 [0.010]	10703	0.450 [0.498]	0.034 [0.015]	6540	0.563 [0.496]	0.012 [0.014]	4163
Asian	0.039 [0.194]	-0.001 [0.004]	10703	0.034 [0.182]	0.001 [0.005]	6540	0.048 [0.214]	-0.004 [0.007]	4163
Has computer at home	0.580 [0.494]	-0.007 [0.010]	12396	0.572 [0.495]	-0.014 [0.016]	7892	0.596 [0.491]	0.005 [0.018]	4504
Frequently uses computer at home	0.455 [0.498]	-0.003 [0.009]	12380	0.454 [0.498]	-0.007 [0.013]	7884	0.457 [0.498]	0.006 [0.018]	4496
Has internet at home	0.736 [0.441]	-0.008 [0.012]	12360	0.741 [0.438]	-0.022 [0.016]	7867	0.726 [0.446]	0.017 [0.018]	4493
Uses computer at home for school activities	0.520 [0.500]	-0.006 [0.010]	12365	0.518 [0.500]	-0.018 [0.014]	7872	0.526 [0.499]	0.016 [0.017]	4493
Uses computer lab at school	0.367 [0.482]	-0.011 [0.033]	12374	0.419 [0.493]	-0.013 [0.044]	7879	0.255 [0.436]	-0.008 [0.050]	4495
Uses computer lab at school during portuguese classes	0.237 [0.426]	0.023 [0.031]	12403	0.290 [0.454]	0.019 [0.043]	7896	0.123 [0.329]	0.031 [0.052]	4507
Uses computer lab at school during math classes	0.255 [0.436]	0.048 [0.033]	12368	0.318 [0.466]	0.035 [0.041]	7873	0.119 [0.323]	0.071 [0.050]	4495
Uses computer lab at school during other classes	0.332 [0.471]	-0.052 [0.029]	12334	0.335 [0.472]	-0.018 [0.037]	7852	0.327 [0.469]	-0.112 [0.061]	4482
Uses computer lab at school not during class	0.144 [0.351]	-0.013 [0.009]	12377	0.148 [0.355]	-0.018 [0.011]	7878	0.135 [0.342]	-0.005 [0.022]	4499
Has mobile phone	0.715 [0.452]	-0.001 [0.010]	12265	0.683 [0.466]	0.000 [0.014]	7808	0.783 [0.412]	-0.001 [0.014]	4457
Has internet on mobile phone	0.706 [0.455]	-0.003 [0.010]	11286	0.680 [0.467]	-0.004 [0.015]	6925	0.759 [0.428]	-0.003 [0.015]	4361
Lives with mother	0.893 [0.309]	0.005 [0.005]	12362	0.902 [0.298]	0.007 [0.006]	7864	0.874 [0.332]	0.001 [0.009]	4498
Lives with father	0.617 [0.486]	0.003 [0.009]	12360	0.640 [0.480]	-0.002 [0.012]	7861	0.569 [0.495]	0.013 [0.014]	4499
Has books at home	0.767 [0.422]	-0.009 [0.010]	12394	0.740 [0.439]	-0.021 [0.014]	7890	0.826 [0.379]	0.013 [0.013]	4504
Parents talk about school	0.844 [0.363]	-0.001 [0.007]	12394	0.867 [0.339]	-0.012 [0.008]	7891	0.795 [0.404]	0.019 [0.012]	4503
Works outside home	0.082 [0.274]	0.000 [0.005]	12388	0.080 [0.272]	-0.004 [0.007]	7882	0.084 [0.278]	0.008 [0.008]	4506
Has ever repeated a grade	0.238 [0.426]	-0.006 [0.010]	12304	0.186 [0.389]	0.011 [0.013]	7830	0.349 [0.477]	-0.036 [0.016]	4474
Math is the preferred subject	0.428 [0.495]	0.008 [0.013]	12389	0.506 [0.500]	0.007 [0.015]	7894	0.260 [0.439]	0.009 [0.023]	4495
Portuguese is the preferred subject	0.249 [0.432]	0.008 [0.013]	12389	0.267 [0.443]	0.007 [0.014]	7894	0.208 [0.406]	0.010 [0.024]	4495
Other subject is preferred	0.323 [0.468]	-0.016 [0.012]	12389	0.226 [0.418]	-0.014 [0.012]	7894	0.532 [0.499]	-0.018 [0.027]	4495
Participated in Math Olympics	0.192 [0.394]	0.000 [0.009]	11340	0.074 [0.262]	0.005 [0.010]	7192	0.444 [0.497]	-0.009 [0.021]	4148
P value joint		0.695			.453			.720	

Notes – This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the control group mean, ii) the results of student-level regressions of covariates collected in the baseline survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment and strata fixed effects and iii) Number of observations. Standard errors clustered at the school level are in brackets. P-values for a test that all covariates are balanced are reported at the bottom of the table for each of the three samples considered.

Table 2 – Attrition

Pooled sample				5th grade				9th grade			
Mean (control)	Diff	N Obs.	N Clusters	Mean (control)	Diff	N Obs.	N Clusters	Mean (control)	Diff	N Obs.	N Clusters
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Student-level Attrition in Survey											
0.393	-0.025	18065	150	0.377	-0.030	12220	136	0.433	-0.015	5845	72
	[0.011]				[0.015]				[0.027]		
	(0.027)				(0.048)				(0.585)		
Panel B: School-grade-level Attrition in SAEB exam											
0.142	-0.008	217	157	0.099	-0.002	143	143	0.229	-0.020	74	
	[0.038]				[0.045]				[0.089]		
	(0.830)				(0.964)				(0.819)		
Panel C: Student-level Attrition in SAEB exam											
0.868	-0.005	17151	143	0.877	-0.006	11906	129	0.844	-0.002	5245	58
	[0.007]				[0.008]				[0.011]		
	(0.481)				(0.467)				(0.848)		

Notes – This table reports differences in attrition between treatment and control groups in the follow-up survey (Panel A) and in the SAEB exam (school-grade-level in Panel B and student-level in Panel C). We report for the pooled sample and for the 5h grade and 9th grades samples separately: i) the control group mean, ii) the results of regressions of our indicator of attrition (which takes value one if there is no follow-up data available) on a dummy variable indicating treatment assignment and strata fixed effects, iii) Number of observations and iv) Number of clusters. Standard errors, in brackets, are clustered at the school level. P-values are in parenthesis.

Table 3 – Baseline Covariates Balance - SAEB

	Pooled Sample			5th grade			9th grade		
	Mean (control) (1)	Diff (2)	N (3)	Mean (control) (4)	Diff (5)	N (6)	Mean (control) (7)	Diff (8)	N (9)
Male	0.504	-0.008	14411	0.512	-0.010	10072	0.485	-0.001	4339
	[0.500]	[0.009]		[0.500]	[0.010]		[0.500]	[0.019]	
White	0.283	-0.009	14423	0.293	-0.013	10047	0.255	0.002	4376
	[0.450]	[0.009]		[0.455]	[0.013]		[0.436]	[0.015]	
Black	0.073	-0.005	14423	0.070	-0.007	10047	0.082	0.000	4376
	[0.261]	[0.005]		[0.255]	[0.006]		[0.274]	[0.008]	
Mixed	0.527	0.007	14423	0.517	0.015	10047	0.551	-0.014	4376
	[0.499]	[0.010]		[0.500]	[0.014]		[0.497]	[0.018]	
Asian	0.028	0.004	14423	0.023	0.002	10047	0.041	0.007	4376
	[0.166]	[0.003]		[0.151]	[0.003]		[0.198]	[0.007]	
Indigenous	0.025	-0.001	14423	0.025	0.000	10047	0.026	-0.001	4376
	[0.157]	[0.002]		[0.157]	[0.003]		[0.158]	[0.004]	
Race not declared	0.064	0.004	14423	0.071	0.004	10047	0.045	0.006	4376
	[0.244]	[0.005]		[0.257]	[0.007]		[0.207]	[0.006]	
Age	12.007	-0.005	14625	10.821	0.018	10220	15.099	-0.063	4405
	[2.087]	[0.018]		[0.795]	[0.020]		[0.916]	[0.044]	
Mother has completed at least high school	0.625	0.025	9606	0.636	0.019	6034	0.606	0.037	3572
	[0.484]	[0.013]		[0.481]	[0.021]		[0.489]	[0.023]	
Mother literate	0.985	-0.002	14564	0.989	-0.005	10173	0.976	0.006	4391
	[0.120]	[0.002]		[0.106]	[0.002]		[0.152]	[0.004]	
Father has completed at least high school	0.571	0.017	8006	0.565	0.007	4990	0.582	0.034	3016
	[0.495]	[0.013]		[0.496]	[0.020]		[0.493]	[0.023]	
Father literate	0.958	0.001	14373	0.962	0.001	10007	0.948	0.001	4366
	[0.201]	[0.004]		[0.192]	[0.004]		[0.222]	[0.008]	
Teacher younger than 50 years old	0.760	0.008	12805	0.761	0.012	10530	0.752	-0.017	2275
	[0.427]	[0.045]		[0.426]	[0.049]		[0.432]	[0.134]	
2015 Prova Brasil math grade	0.095	0.029	16820	0.090	-0.066	11654	0.107	0.266	5166
	[1.023]	[0.063]		[0.934]	[0.091]		[1.216]	[0.203]	
P value joint		.679			.470			.865	

Notes – This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the control group mean, ii) the results of student-level regressions of covariates available in the SAEB dataset on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment and strata fixed effects and iii) Number of observations. Standard errors clustered at the school level are in brackets. P-values for a test that all covariates are balanced are reported at the bottom of the table for each of the three samples considered.

Table 4 – Follow-up Survey

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Has computer at home	0.622	-0.012	12816	0.631	-0.013	9004	0.595	-0.008	3812
	[0.485]	[0.013]		[0.483]	[0.017]		[0.491]	[0.024]	
		(0.357)			(0.422)			(0.723)	
Frequently uses computer at home	0.472	0.015	12808	0.484	0.020	9004	0.438	0.004	3804
	[0.499]	[0.011]		[0.500]	[0.015]		[0.496]	[0.020]	
		(0.182)			(0.193)			(0.855)	
Has internet at home	0.795	-0.002	12745	0.804	-0.002	8953	0.770	-0.002	3792
	[0.404]	[0.011]		[0.397]	[0.014]		[0.421]	[0.018]	
		(0.865)			(0.911)			(0.893)	
Uses computer at home for school activities	0.519	0.004	12764	0.526	0.001	8962	0.502	0.011	3802
	[0.500]	[0.013]		[0.499]	[0.018]		[0.500]	[0.021]	
		(0.763)			(0.953)			(0.593)	
Uses computer lab at school	0.488	0.285	12820	0.555	0.192	9010	0.300	0.513	3810
	[0.500]	[0.033]		[0.497]	[0.036]		[0.458]	[0.049]	
		(0.000)			(0.000)			(0.000)	
Uses computer lab at school during portuguese classes	0.317	-0.039	12801	0.370	-0.057	8994	0.167	0.003	3807
	[0.465]	[0.029]		[0.483]	[0.039]		[0.373]	[0.042]	
		(0.179)			(0.143)			(0.944)	
Uses computer lab at school during math classes	0.340	0.445	12743	0.398	0.330	8951	0.175	0.728	3792
	[0.474]	[0.035]		[0.490]	[0.039]		[0.380]	[0.042]	
		(0.000)			(0.000)			(0.000)	
Uses computer lab at school during other classes	0.368	-0.055	12703	0.386	-0.066	8923	0.316	-0.027	3780
	[0.482]	[0.026]		[0.487]	[0.030]		[0.465]	[0.060]	
		(0.035)			(0.028)			(0.647)	
Uses computer lab at school not during class	0.151	0.051	12791	0.140	0.037	8985	0.181	0.084	3806
	[0.358]	[0.014]		[0.347]	[0.014]		[0.385]	[0.038]	
		(0.000)			(0.008)			(0.025)	
Uses Khan Academy	0.063	0.903	12673	0.078	0.882	8924	0.022	0.956	3749
	[0.244]	[0.017]		[0.268]	[0.023]		[0.145]	[0.005]	
		(0.000)			(0.000)			(0.000)	
Uses Khan Academy during school	0.044	0.782	12549	0.055	0.707	8833	0.010	0.967	3716
	[0.204]	[0.022]		[0.228]	[0.028]		[0.100]	[0.005]	
		(0.000)			(0.000)			(0.000)	

Notes – This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the control group mean, ii) the results of a student-level regression of different measures collected in the follow-up survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment and strata fixed effects and iii) Number of observations. Standard errors are in brackets and p-values in parenthesis. Standard errors are clustered at the school level.

Table 5 – Descriptive Statistics - Usage of Khan Academy

	Total number of minutes			Total number of weeks logged in		
	Pooled	Grade 5	Grade 9	Pooled	Grade 5	Grade 9
	(1)	(2)	(3)	(4)	(5)	(6)
No infrastructure problem	147.3	169.3	-18.3	2.888	3.979	-2.357
s.e.	[58.0]	[70.3]	[58.4]	[1.330]	[1.498]	[1.475]
p-value	(0.011)	(0.016)	(0.754)	(0.030)	(0.008)	(0.110)
One computer per student	195.0	224.2	183.9	1.669	2.082	1.741
s.e.	[73.3]	[91.0]	[98.0]	[1.402]	[1.586]	[2.190]
p-value	(0.008)	(0.014)	(0.061)	(0.234)	(0.189)	(0.426)
9th grade	-178.3	-	-	-3.206	-	-
	[45.6]			[0.787]		
	(0.000)			(0.000)		
Municipality fixed effects	Y	Y	Y	Y	Y	Y
Mean (with infrastructure problem and rotation)						
5th grade		540.0			13.407	
		[61.2]			[1.154]	
9th grade		386.3			11.359	
		[36.5]			[0.824]	
Number of students	8302	5325	2977	8302	5325	2977
Number of schools	103	65	38	103	65	38

Notes – This table reports, in columns 1-3, results from a student-level regression of the total number of minutes spent in the platform on an indicator of no infrastructure problems, an indicator of modality of implementation based on one computer per student, and municipality fixed effects, for the pooled sample, and 5th and 9th grades subsamples respectively. In column 1 we also include an indicator of the 9th grade. Standard errors are clustered at the school level. In columns 4-6, we report results for the same specifications using the total number of weeks logged in as the dependent variable.

Table 6 – Results on Math Proficiency and Attitudes towards math

	Math test scores		Attitudes towards math	
	(1)	(2)	(3)	(4)
		Panel A: Full sample		
Treatment	-0.023	-0.016	0.056	0.060
s.e.	[0.033]	[0.028]	[0.028]	[0.020]
p-value	(0.490)	(0.566)	(0.043)	(0.003)
N	14846	14846	11157	11157
		Panel B: 5th grade		
Treatment	-0.036	-0.002	0.044	0.062
s.e.	[0.049]	[0.037]	[0.031]	[0.024]
p-value	(0.462)	(0.954)	(0.159)	(0.010)
N	10388	10388	7806	7806
		Panel C: 9th grade		
Treatment	0.011	-0.051	0.086	0.057
s.e.	[0.075]	[0.034]	[0.049]	[0.033]
p-value	(0.883)	(0.129)	(0.080)	(0.086)
N	4458	4458	3351	3351
Includes covariates	No	Yes	No	Yes

Notes – This table reports the results of a student-level regression of math proficiency (columns 1-2) and attitudes towards math (columns 3-4) on an dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment and strata fixed effects. Panels A, B and C refer to the pooled sample, and 5th and 9th grades subsamples separately. For the pooled regressions, we interact the strata fixed effects with grade. The specifications reported in column 2 include the covariates presented in Table 3, while the specifications reported in column 2 include the covariates presented in Table 1. Standard errors are clustered at the school level.

Table 7 – Validity of Measures for Heterogeneity Exercises

	All schools		Two cycle schools		One cycle schools	
	5th grade	9th grade	5th grade	9th grade	5th grade	9th grade
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: No Infrastructure Problem					
T	-0.024	0.019	-0.023	0.023	-0.025	0.000
s.e.	[0.065]	[0.082]	[0.101]	[0.101]	[0.085]	-
p-value	(0.705)	(0.815)	(0.816)	(0.816)	(0.765)	-
Mean (omitted group)	0.551	0.471	0.567	0.571	0.538	0.000
	[0.060]	[0.087]	[0.092]	[0.095]	[0.081]	-
Number of schools	136	72	58	58	78	14
	Panel B: One Computer per Student					
T	0.034	-0.022	0.027	-0.027	0.040	0.000
s.e.	[0.057]	[0.071]	[0.087]	[0.087]	[0.076]	-
p-value	(0.555)	(0.755)	(0.757)	(0.757)	(0.595)	-
Mean (omitted group)	0.403	0.529	0.567	0.643	0.250	0.000
	[0.063]	[0.087]	[0.092]	[0.092]	[0.078]	-
Number of schools	127	72	58	58	69	14

Notes – This table reports, in Panel A, results of a school-grade-level regression of a dummy variable that takes value one if the there are no infrastructure problems on the treatment indicator and strata fixed effects. In columns 1-2, we display the results for 5th and 9th grades for all schools, while columns 3-4 and 5-6 show estimates for 5th and 9th grades in two cycle and one cycle schools respectively. Panel B shows results for the indicator of one computer per student as the dependent variable. The means for the ommitted groups in columns 1 and 2 of Panel B (40% for 5th grade and 53% for 9th grade) are not inconsistent with the number reported in the text, that 37% of schools are based on one computer per student modality. In the table, two cycle schools are accounted twice, since our estimates are at the school-grade level.

Table 8 – ITT Heterogeneity

	Math test score		Attitudes towards math	
	No infrastructure problem	One computer per student	No infrastructure problem	One computer per student
	(1)	(2)	(3)	(4)
		Panel A: Full sample		
T*X	0.058	0.081	0.029	0.036
	[0.041]	[0.047]	[0.043]	[0.041]
	(0.158)	(0.084)	(0.500)	(0.377)
T*(1-X)	-0.056	-0.076	0.072	0.053
	[0.039]	[0.034]	[0.040]	[0.023]
	(0.153)	(0.028)	(0.071)	(0.022)
p-value	0.047	0.011	0.478	0.723
N	13825	13231	11135	10710
		Panel B: 5th grade		
T*X	0.093	0.127	0.066	0.070
	[0.051]	[0.059]	[0.034]	[0.045]
	(0.067)	(0.032)	(0.052)	(0.122)
T*(1-X)	-0.062	-0.082	0.039	0.035
	[0.054]	[0.046]	[0.040]	[0.030]
	(0.253)	(0.074)	(0.329)	(0.243)
p-value	0.037	0.009	0.619	0.516
N	9682	9088	7784	7359
		Panel C: 9th grade		
T*X	-0.064	-0.102	-0.023	-0.031
	[0.060]	[0.065]	[0.079]	[0.073]
	(0.291)	(0.115)	(0.767)	(0.666)
T*(1-X)	-0.009	-0.075	0.076	0.108
	[0.075]	[0.048]	[0.034]	[0.031]
	(0.904)	(0.116)	(0.025)	(0.000)
p-value	0.606	0.743	0.263	0.076
N	4143	4143	3351	3351

Notes – This table reports results for student-level regressions of math proficiency (columns 1-2) and attitudes towards math (columns 3-4) on interaction terms between the treatment dummy and the heterogeneity variable. In columns (1) and (3), X is an indicator variable which takes value one if there were no infrastructure problems; in columns (2) and (4), X is an indicator variable which takes value one if the implementation modality was based on one computer per student. Specifications in columns 1 and 2 include strata fixed effects, the X variable in level, and the covariates reported in Table 3. Specifications in columns 1 and 2 include strata fixed effects, the X variable in level, and the covariates reported in Table 1. Standard errors are clustered at the school level.

APPENDIX A. Appendix Tables

Table A.1 – Balance conditional on non-attritors

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
Attitudes towards math	0.030	0.010	7243	0.049	-0.023	4688	-0.012	0.071	2555
	[1.004]	[0.031]		[1.006]	[0.037]		[0.998]	[0.059]	
Male	0.502	0.004	7761	0.501	-0.001	5056	0.504	0.014	2705
	[0.500]	[0.011]		[0.500]	[0.013]		[0.500]	[0.017]	
Year of Birth	2004.723	-0.022	7764	2005.995	-0.083	5054	2001.894	0.093	2710
	[2.232]	[0.032]		[1.266]	[0.043]		[0.949]	[0.047]	
White	0.336	-0.015	6692	0.369	-0.017	4194	0.269	-0.011	2498
	[0.472]	[0.011]		[0.483]	[0.015]		[0.444]	[0.016]	
Black	0.099	-0.005	6692	0.104	-0.009	4194	0.090	0.001	2498
	[0.299]	[0.007]		[0.305]	[0.009]		[0.286]	[0.011]	
Indian	0.040	0.003	6692	0.043	0.004	4194	0.034	0.002	2498
	[0.196]	[0.005]		[0.203]	[0.006]		[0.181]	[0.007]	
Mixed	0.486	0.019	6692	0.447	0.021	4194	0.563	0.014	2498
	[0.500]	[0.012]		[0.497]	[0.016]		[0.496]	[0.016]	
Asian	0.039	-0.002	6692	0.037	0.001	4194	0.044	-0.007	2498
	[0.194]	[0.005]		[0.188]	[0.005]		[0.205]	[0.007]	
Has computer at home	0.602	-0.016	7772	0.597	-0.020	5065	0.613	-0.009	2707
	[0.490]	[0.013]		[0.491]	[0.017]		[0.487]	[0.024]	
Frequently uses computer at home	0.468	-0.004	7765	0.465	-0.003	5062	0.476	-0.008	2703
	[0.499]	[0.012]		[0.499]	[0.015]		[0.500]	[0.023]	
Has internet at home	0.740	-0.005	7749	0.751	-0.024	5049	0.716	0.030	2700
	[0.439]	[0.013]		[0.433]	[0.017]		[0.451]	[0.021]	
Uses computer at home for school activities	0.531	-0.010	7750	0.528	-0.018	5050	0.539	0.005	2700
	[0.499]	[0.013]		[0.499]	[0.016]		[0.499]	[0.024]	
Uses computer lab at school	0.372	-0.016	7751	0.419	-0.005	5051	0.266	-0.035	2700
	[0.483]	[0.031]		[0.494]	[0.046]		[0.442]	[0.045]	
Uses computer lab at school during portuguese classes	0.245	0.010	7773	0.301	0.002	5065	0.122	0.024	2708
	[0.430]	[0.034]		[0.459]	[0.046]		[0.328]	[0.055]	
Uses computer lab at school during math classes	0.263	0.047	7758	0.333	0.035	5055	0.108	0.070	2703
	[0.440]	[0.034]		[0.471]	[0.043]		[0.311]	[0.054]	
Uses computer lab at school during other classes	0.337	-0.055	7732	0.337	-0.020	5039	0.337	-0.123	2693
	[0.473]	[0.029]		[0.473]	[0.038]		[0.473]	[0.061]	
Uses computer lab at school not during class	0.138	-0.014	7760	0.142	-0.021	5057	0.130	-0.002	2703
	[0.345]	[0.010]		[0.349]	[0.012]		[0.337]	[0.024]	
Has mobile phone	0.711	0.010	7699	0.680	0.008	5018	0.779	0.013	2681
	[0.454]	[0.012]		[0.467]	[0.016]		[0.415]	[0.016]	
Has internet on mobile phone	0.710	0.007	7026	0.689	0.004	4401	0.752	0.013	2625
	[0.454]	[0.012]		[0.463]	[0.017]		[0.432]	[0.016]	
Lives with mother	0.902	0.001	7752	0.908	0.007	5048	0.888	-0.010	2704
	[0.297]	[0.006]		[0.289]	[0.008]		[0.315]	[0.012]	
Lives with father	0.639	0.001	7748	0.658	-0.010	5047	0.595	0.021	2701
	[0.480]	[0.012]		[0.474]	[0.015]		[0.491]	[0.020]	
Has books at home	0.777	-0.009	7771	0.748	-0.013	5064	0.841	0.000	2707
	[0.416]	[0.010]		[0.434]	[0.014]		[0.366]	[0.015]	
Parents talk about school	0.837	0.009	7772	0.859	-0.002	5066	0.787	0.030	2706
	[0.370]	[0.009]		[0.348]	[0.010]		[0.410]	[0.016]	
Works outside home	0.067	0.004	7772	0.064	0.006	5063	0.075	0.000	2709
	[0.251]	[0.005]		[0.245]	[0.007]		[0.263]	[0.010]	
Has ever repeated a grade	0.211	-0.001	7724	0.163	0.011	5033	0.319	-0.025	2691
	[0.408]	[0.011]		[0.369]	[0.014]		[0.466]	[0.021]	
Math is the preferred subject	0.440	0.007	7769	0.521	0.003	5064	0.260	0.015	2705
	[0.496]	[0.015]		[0.500]	[0.017]		[0.439]	[0.026]	
Portuguese is the preferred subject	0.238	-0.001	7769	0.250	0.002	5064	0.212	-0.007	2705
	[0.426]	[0.014]		[0.433]	[0.014]		[0.409]	[0.026]	
Other subject is preferred	0.321	-0.006	7769	0.229	-0.005	5064	0.528	-0.008	2705
	[0.467]	[0.014]		[0.420]	[0.015]		[0.499]	[0.028]	
Participated in Math Olympics	0.182	-0.001	7086	0.063	0.007	4606	0.446	-0.018	2480
	[0.386]	[0.010]		[0.243]	[0.012]		[0.497]	[0.024]	
P value joint		0.845			0.801			0.578	

Notes – This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the control group mean, ii) the results of student-level regressions of covariates collected in the baseline survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment and strata fixed effects and iii) Number of observations. The sample is composed of non-attritors, individuals for which there is follow-up data available. Standard errors clustered at the school level are in brackets. P-values for a test that all variables are balanced are reported at the bottom of the table for each of the three samples considered. Standard errors clustered at the school level are presented in brackets. P-values are presented in parenthesis.

Table A.2 – Number of Students Enrolled per Classroom

	Cycle I schools (1)	Cycle II schools (2)	Two cycle schools (3)
3rd grade	0,526		
	[0,429]		
	(0,220)		
4th grade	-0,588		
	[0,450]		
	(0,192)		
6th grade		2,357	
		[1,474]	
		(0,110)	
9th grade			0,190
			[0,743]
			(0,799)
Mean (omitted group)	28,936	28,936	27,328
	[0,649]	[0,649]	[0,949]
Omitted group	5th grade	9th grade	5th grade
Number of schools	78	14	58

Notes – This table reports results of a regression of maximum number of students enrolled per class in each grade on i) indicator variables of 3rd and 4th grades (in column 1 - Cycle I schools); ii) 6th grade (in column 2 - Cycle II schools) and iii) 9th grade (in column 3 - Two cycle schools) and school fixed effects.

3 Human Capital Investments Among Vulnerable Youth in Rio de Janeiro: Experimental Evidence from the Protejo Program

3.1 Introduction

Traditionally, social policy interventions for youth have tried to improve welfare by investing resources in fostering the academic or vocational skills of disadvantaged young people, changing the long-term benefits and costs associated with crime *vs* work and schooling. Whether these interventions are able to ultimately affect human capital accumulation, labor market prospects, victimization and crime in the long run is an important empirical question.

In this paper, we rely on experimental evidence to evaluate the impacts of the Project of Youth Protection in Vulnerable Territories ("Projeto de Proteção de Jovens em Territórios Vulneráveis", *Protejo*), implemented in Rio de Janeiro in 2010. *Protejo* is a large-scale government-led initiative of the National Program of Public Security with Citizenship, targeting vulnerable young people between the ages of 15 and 24 years old, specifically at-risk individuals that are exposed to domestic or urban violence, have a low educational level and live in a household with total income below two Brazilian minimum wages. Its main goal is to foster cognitive and noncognitive skill accumulation of youth participants to increase labor market attachment, schooling, awareness of rights and, ultimately, prevent crime and victimization. The program comprised a total of 800 hours of activities, divided into around 400 hours of in-class training that included general math, language, information technology and vocational training, and 400 hours dedicated to several community engagement, citizenship building and mentoring activities.

We use both survey and administrative data to evaluate the impacts of *Protejo* on a wide range of dimensions. Our results indicate the program increased community engagement, as expected, but also increased the probability of having children and victimization, which contradicted our initial assumptions. Although we find no average effects on education or noncognitive skills two years after the program, the intervention was likely successful in developing skills of the beneficiaries, since our results show strong impacts of *Protejo* on the probability of formal employment from 3 years after the end of the program onwards (years 2014, 2015 and 2016). When we explore treatment effects heterogeneity, we find that the positive results on formal employment are driven by the outcomes from the 17-19 age group (age at program enrollment in 2010). For the gender-specific subsamples, we find that positive effects for men arise earlier, in the year following the end of the program, and spanning through 2015. For the female subsample, the effects are large and significant for both 2015 and 2016, four and five years after the end of the program.

The findings from our paper augment the evidence on the labor market effects of programs that aim to increase the skills of disadvantaged youth in developing countries. The available evidence mostly indicates positive effects, but it is not conclusive. Positive effects on formal employment are found by Alzúa et al. (2013) in an experimental study of a youth training program that included life-skills and vocational training in Argentina. The Colombian program *Jovenes en Accion*, which provided vocational training for unemployed youth (aged between 18 and 25) from poor households, also registered strong formal

employment effects for women (ATTANASIO et al., 2011), which persisted ten years after the program (ATTANASIO et al., 2015). For Brazil, Camargo et al. (2017) find large labor market and noncognitive skills gains for women as a result of a program that subsidized vocational training courses for current or former students from the public high school system. In the Dominican Republic, however, Ibarraran et al. (2014) find no overall effects on employment rates (although they find positive effects on noncognitive skills) of the *Juventud y Empleo* Program, a youth training program that provided technical and life-skills training targeted at youth that never completed high school.

The intervention we evaluate also has components that resemble mentoring interventions for at-risk youth, which have been associated with gains in schooling, wages and crime (HECKMAN; KAUTZ, 2013). The *Big Brothers/Big Sisters of America* program, for instance, had components of social, cultural, and recreational enrichment, improving peer relationships and improving noncognitive skills. Tierney et al. (1995) find program participants were less likely to use drugs, engage in fights and had improved outcomes in terms of education and attitudes towards work. The *Job Corps* program in the US, a residential program focused on remedial education, counseling and training in social skills, had short run effects on wages and crime, although the effects on earnings were only temporary (SCHOCHET et al., 2008).

The remainder of this paper is organized as follows. 3.2 describes the 2010 edition of the *Protejo* program in Rio de Janeiro and the experimental design. In Section 3.3 we describe our data. Section 3.4 presents our empirical strategy and section 3.5 discusses the results. Section 3.6 concludes with a summary of our findings and discusses potential avenues for future research.

3.2 Background: Program Description and Experimental Design

3.2.1 The Protejo Program

Protejo is one of the core interventions of the National Program of Public Security with Citizenship ("Programa Nacional de Segurança Pública com Cidadania" - *Pronasci*), implemented through a federal law in 2007 to prevent and control crime in Brazil, under the coordination of the Ministry of Justice. *Pronasci*'s main mandate was to promote public safety through integrating federal, state and municipal governments in the articulation of public safety and social policy interventions. Although we don't have data for its overall coverage throughout the years, we found records that, between the years of 2008 and 2013, it benefited roughly 35,000 individuals in 14 Brazilian states, as it can be observed in Figure 1. Its implementation was articulated with another intervention, *Women of Peace*, which trained women in vulnerable communities to promote peace and human rights, acting in social mobilization to prevent domestic violence and protect the community's youth. The *Women of Peace* had the mandate to identify within their community potential beneficiaries to participate in the *Protejo* program. The federal government invested a total of R$205,000,000 in Brazilian Reais (in 2013, approximately 100,000,000 US Dollars) in both programs between 2008 and 2013.[1]

According to the Ministry of Justice, the program had three specific goals: (i) strengthening citizenship with active social participation; (ii) protecting the victimized youth, specially from dying at a

[1] Data for the coverage and expenditure of program between 2008 and 2013 was obtained from the Ministry of Justice's website. Although we don't have information to assess the extent of the program after 2013, the *Protejo* intervention remains active and we find records of its expansion up to 2019.

young age and (iii) promoting education and work. The program combined elements of technical and vocational education and training programs common in Latin America (e.g. Card et al. (2011), Attanasio et al. (2011)), but also had socioemotional development features through its activities that promoted empowerment, citizenship awareness and community (re-)integration.

In total, the 2010 edition of the program in Rio de Janeiro offered 800 hours of activities, divided into 12 different modules that took place between November 2010 and July 2011. of[2] All courses were provided by the National Service of Learning in Commerce (SENAC), a very well-known Brazilian institution that traditionally offers professional education in different market segments. For the elaboration of the "individual development plan", each participant was assigned to an individual mentor. This component involved both group sessions to inform the participants about potential channels for insertion in the labor market, and individual sessions through which the tutor helped each beneficiary to work on their individual strategy.

Additional 400 hours were dedicated to socioemotional development, citizenship building, youth empowerment and community integration. These involved cultural/sports-related activities (200 hours), community development activities (40 hours), community therapy (20 hours), conflict mediation (20 hours) and citizenship building (100 hours) with emphasis on self-knowledge, youth emancipation, youth and minority rights and reflections about violence, value of education and work, family formation, drug abuse and sexuality. This dimension of the program also allocated 20 hours for pedagogical discussions, through which the participants had the opportunity to discuss their progress in the program and potential difficulties. One specific feature of the program is that it payed participating individuals up to twelve monthly stipends of 100 Reais (in 2010, approximately 57 US dollars) conditional on minimum attendance.

3.2.2 Experimental Design

For the 2010 edition of *Protejo* in Rio de Janeiro, the program administration implemented a lottery to select beneficiaries in October 2010. Nineteen low income communities in Rio de Janeiro were selected to receive the program. Within each community, individuals were stratified based on 3 criteria: i) gender; ii) an indicator of social vulnerability (high and low vulnerability) and iii) an indicator of educational level (high and low educational level). Information for stratification was extracted from the form applicants filled out when applying to the lottery. The combination of the three different criteria (with two categories each) generated 8 groups within each community, composing a total of 152 strata. Individuals that never enrolled in high school were classified into the "Low educational level" category, while individuals with some high school attendance were considered "High educational level". The social vulnerability index was composed of five dimensions: i) housing conditions; ii) whether parents were alive; iii) whether the individual had been a victim of violence; iv) whether the individual had been in contact with drugs and v) whether the individual had been indicated to the program by social assistance programs or institutions.

As the goal of the program was to reach the most at-risk individuals, the original design of the

[2] The courses offered were: interior design, handbag manufacturing, makeup artist, hairdresser, hospitality reception, DJ techniques, sales techniques, administrative assistant, administrative services, secretary, reception for food industry, waiter, clothes design, hospitality room organization, manicure and pedicure and letter design.

program was to allocate more vacancies for the most vulnerable groups (men, high social vulnerability, low educational level). However, for the most vulnerable, demand to the program was lower so the final allocation of vacancies between groups in each community did not correspond exactly to the original allocation plan. In general, demand was higher among more educated female applicants. Out of 152 strata, in 24 the lottery was not carried out due to lack of excess demand, so all applicants received an offer to the program. A total of 2,425 individuals were initially offered the program, and 2,432 were allocated in the waiting list. The distribution of vacancies per treatment arm by community is presented in Appendix B.

3.3 Sample and Data

3.3.1 Sample

Due to financial constraints, a balanced sample of 2,248 individuals (in 93 strata) was selected for follow-up survey data collection among the strata that had at least 5 observations in both treatment offer arms. For administrative data outcomes, we consider both the full lottery registry as well as the survey sample.

3.3.2 Survey Data

We conduct our analysis at the individual level using baseline and endline surveys collected respectively in 2010 and 2013, almost two years after the end of the program. The baseline survey stems from the lottery registry and was collected for all lottery applicants at the time of application to the program. It collected information for the entire lottery registry on sociodemographic and housing characteristics, income and assets, parents' education, individuals' educational level and perspectives, family formation, drug use, general labor market information and measures of noncognitive skills. The follow-up survey, which covered the evaluation sample, collected more detailed data for all these dimensions, as well as data on social interaction and community engagement, awareness of rights, victmization, sexually transmitted diseases, exposure to the program and additional measures of noncognitive skills.

Eight dimensions of noncognitive skills were assessed in this study, through four different instruments:

- The "Big Five" dimensions, based on a 44 item instrument developed by John et al. (1991) and adapted to Portuguese by Andrade (2008). These dimensions include i) *Extraversion*, which reflects aspects of the individual's personality such as being friendly, sociable, energetic and enthusiastic; ii) *Agreeableness*, which captures the extent to which individuals are cooperative, tolerant and flexible; iii) *Conscientiouness*, related to an individual's capacity of being efficient, organized and disciplined; iv) *Neuroticism*, which is related to the consistency of a person's emotional reactions and v) *Openness to Experiences* which reflects the individual's abilities of being imaginative, curious and interested in a wide range of topics.
- Rosenberg self-esteem scale, developed by Rosenberg (1965) and validated to Brazilian Portuguese by Dini et al. (2001), which measures global self-esteem.

- Barratt impulsiveness scale, developed by Patton et al. (1995) and adapted to Portuguese by Malloy-Diniz et al. (2010), which assesses impulsiveness based on the theoretical model proposed by Ernst Barratt.
- Leary Brief Fear of Negative Evaluation scale, developed by (LEARY, 1983) and translated to Portuguese by Correia et al. (2014), which measures symptoms of social anxiety.

3.3.3 Administrative Data

The formal employment data is based on information from the Brazilian Ministry of Employment and Labor's Annual Social Information Report ("Relação Anual de Informações Sociais", RAIS) for years 2010-2016, a confidential longitudinal data set of compulsory administrative records reported by every employer in the formal sector. It does not include workers without signed work cards, such as the self-employed, elected officials, domestic workers and other smaller categories. This information must be sent each year to the Ministry of Labor by all operating firms. Details on the matching performed to link this data with our full lottery registry follow in Appendix C.

Measures of program participation used for instrumental variables estimation stem from program's administration database containing the number of monthly stipends paid to each individual that was officially enrolled in the program.

3.3.4 Descriptive Statistics and Balance

Random assignment of the offer to participate in the *Protejo* program appears to have been carried out successfully, as suggested in Table 1. Columns (1)-(4) present means across offer arms and p-values for a test of difference between groups for the full registry data that includes all applicants to the 2010 edition of the program. We report estimates from a regression for each covariate on an indicator variable for the treatment and strata fixed effects, with heteroskedasticity-robust standard errors. Similarly, columns (5)-(8) consider the evaluation sample selected for the follow-up interview. Among the 32 differences we show, only three (contact with cocaine in the first group of columns and contact with crack in the first and second group) are significant at the 10% level. Because we are looking at so many statistics, it is not surprising that a small fraction is significant at the 10% level. Perhaps more importantly, differences suggest that those that received the offer were "worse" (e.g., 2 p.p.s more likely to have had any contact with cocaine) than those who did not, minimizing concerns about upward bias in the estimates. Finally, p-values for the test of joint significance are close to 0.6, suggesting that there is no systematic correlation between observables and the offer to participate in *Protejo*.

The average age of women and men in the sample at the time of application to the program was around 17 years old, and the age distribution displayed in Figure 2 shows that the program followed its initial focalization plan to target youth aged 15-24. The sample was composed of 62% of female participants. Almost 3 in 4 individuals were black or mixed race. Only a little more than 10% had a job, around 22% were not attending school and almost 20% already had a child.

3.3.5 Attrition

The follow-up interviews were conducted between March and June 2013, with the goal of interviewing the evaluation sample composed of $2,248$ individuals. In total, $1,779$ individuals were successfully interviewed in the follow-up survey, which corresponds to 80% percent of the total initial sample. This attrition rate compares favorably to the attrition rates registered in similar types of studies. We test whether there was differential attrition between the treatment and control groups. Table 2 shows results from regressions of whether the individual leaves the sample on an indicator of whether the person was a lottery winner. Column (1) reports a simple difference-in-means estimate, while columns (2) and (3) subsequently add strata dummies and the baseline characteristics controls listed in Table 1. The results show that treatment individuals were less likely to leave the sample from baseline to endline (approximately 2.4 p.p. or 10% of the lottery losers mean) and the coefficients are not significant.

3.4 Empirical Strategy

We use variations of the following specification to compare outcomes of individuals across treatment offer arms:

$$y_{is} = \alpha + \tau_{\text{ITT}} L_{is} + \Gamma \mathbf{X}_{is} + \mu_s + \epsilon_{is} \tag{3.1}$$

where y_{is} denotes an outcome for individual i in strata s, L_{is} is an indicator variable that takes on the value one if individual i won the lottery for participation, $\mathbf{X}_{is}$ is a set of baseline individual level controls, μ_s are strata fixed-effects and ϵ_{is} is an error term. The intention to treat effect captures the effect of being offered the chance to participate in the program and is given by τ_{ITT} in equation (3.1). Our preferred specification does not include controls, because missing values lower the sample size considerably. We present robust standard errors, mainly because randomization was conducted at the individual level.

Winning the lottery increased the likelihood of an individual taking up the program but is not equivalent to the treatment. Both administrative and survey data reflecting participation in the program show there was imperfect compliance. We use IV estimation to recover the effect of treatment on the population of compliers (LATE). The first stage of our estimation, which captures the causal effect of L_i on T_i is given by:

$$T_{is} = \gamma + \theta L_{is} + \phi \mathbf{X}_{is} + \mu_s + \xi_{is} \tag{3.2}$$

where T_{is} is the treatment status of the individual i in strata s, L_{is} is an indicator variable that takes on the value one if individual i won the lottery for participation, $\mathbf{X}_{is}$ is a set of baseline individual level controls, μ_s are strata fixed-effects ξ_{is} is an error term. The second stage of our LATE estimation is given by the following equation:

$$y_{is} = \alpha + \beta_{\text{LATE}} \widehat{T_{is}} + \Gamma \mathbf{X}_{is} + \mu_s + \epsilon_{is} \tag{3.3}$$

where y_{is} denotes an outcome for individual i in strata s, $\widehat{T_{is}}$ is the predicted value of the individual $i's$ treatment status, $\mathbf{X}_{is}$ is a set of baseline individual level controls, μ_s are strata fixed-effects and ϵ_{is} is an error term. β_{LATE} is the coefficient that captures the local average treatment effect in our empirical model. The treatment status for the IV specifications is based on two measures of program's participation, presented in detail in the next section.

3.5 Results

This section presents the results for the main groups of outcomes on which the program was designed to have an effect. For all outcomes, we report results from an intention-to-treat specification following equation (3.1) without controls and using heteroskedasticity-robust standard errors, unless specified otherwise. For our main outcomes, we also report in the appendix LATE estimates following the IV specification from equation (3.3) without controls and using heteroskedasticity-robust standard errors, with the lottery as an instrument for program participation. For the instrumental variables estimation, we consider two measures of program participation, both based on administrative data: in Panel A, we consider whether an individual ever received a stipend from *Protejo* program and, in Panel B, we consider the fraction of total potential stipends (12) each individual received (this variable takes value zero if never received a stipend). We also report in all IV tables the Kleibergen-Paap F-statistic for weak instruments test to confirm the validity of our instrument, following Kleibergen (2007). All estimates are based on survey data collected in 2013 for the evaluation sample, with the exception of the formal labor market outcomes, for which we have data for the full registry for a longer period of time.

3.5.1 Program Participation and Compliance with Experimental Design

No Winning the lottery significantly increased the probability of participating in the program but was not a perfect predictor of participation. Table 3 displays the first stage results, i.e results of a regression following intent-to-treat (ITT) specification from equation (3.1) without controls and using heteroskedasticity-robust standard errors, based on four different measures of program participation for the impact evaluation sample. We rely on two different data sources to measure participation: i) self-reported follow-up survey data, from which we construct an indicator of having ever participated in *Protejo* (reported in column (a)) and ii) administrative data containing information on monthly stipends received by program beneficiaries, from which we construct three different variables. In column (b) we report results for an indicator variable of ever receiving at least one program's stipend; in column (c) we report results for a variable reflecting the number of stipends received by program's participants, ranging from 0-12 and, in column (d), we report results for a variable that reflected the fraction of total potential stipends (which is the variable in column (c) divided by 12).

Figure 3 displays a histogram that plots the distribution of stipends received by lottery status, considering the evaluation sample. Our results show that lottery winners were 20 p.p. more likely to have reported they ever participated in the *Protejo* program (lottery loser's mean was 0.516). The results for administrative data are roughly consistent with this finding, indicating that those who won the lottery were 22 p.p. more likely to having ever received a program's stipend (which corresponds to an increase of 70 percent when compared to the lottery losers' mean of 0.327). Lottery winners on average received stipends for approximately 4.5 months, in contrast with the lottery losers' mean of 2.87 (considering zero for those who never received a stipend).

These results indicate that the program faced challenges with compliance with treatment assignment, as many people who received the offer never participated in the program. This explains the participation in the program of lottery losers, who were originally placed in the waiting list. Our results also suggest difficulties with program evasion as, conditional on ever receiving a stipend, irrespective

of lottery status, only around one third of individuals received all the 12 months of program's stipend. Although participants received a stipend to participate in the program, its value corresponded to 20 percent of a minimum wage, so retaining participants may have been a challenge specially for those that needed to generate higher income and hence were likely to leave the program if they found higher paying jobs. Also, the target population is composed of vulnerable groups for which engagement in training programs may be difficult.

3.5.2 Family Formation and Social Interactions

One aspect of the program was to promote citizenship building, which covered topics involving the reflection about relationships, sexuality and family formation. In Table 4 we report the ITT results for dimensions associated with family formation and social interactions. Column (a) reports the results for an indicator variable of having any child or being pregnant at the time of the survey; column (b) reports results for an indicator of using at least one in 12 contraceptive methods with the current partner at the time of the survey, if applicable; column (c) displays results for an indicator of being married, cohabiting or having a partner; column (d) shows results for "Number of close friends", a categorical variable with the number of current friends considered "close friends" and, column (e), we report results for the "Family proximity index", which is the fraction of six groups of family members (mother, father, grandparents, brother/sisters, offspring and others) with a "very close" relationship with the person interviewed.

Lottery winners were 3.5 p.p. more likely to have reported having children or being expecting, which corresponds to a 12 percent increase when compared to the control group mean of 0.285. LATE estimates in Appendix Table A.3 show those that participated in the program were 15 p.p. more likely to have children (control group mean=0.310). At first, this result seems to go in the opposite direction of our expectation, as usually such types of programs aim at reducing teenage pregnancy. However, if the program fostered family relations and family building, promoting socioemotional stability, there is a possibility that this may have been a desired outcome. For all other dimensions of family formation and social interaction, there are not noticeable differences between the treatment offer arms.

3.5.3 Conflict and Victimization

The program's module of citizenship building and conflict mediation promoted awareness of youth rights, and fostered a discussion about the different types of violence in society, including domestic violence, homophobia, racism, among others. In Table 5 we show ITT results for indicator variables reflecting whether: (a) the individual has been a victim of a physical aggression, or verbal aggression with physical aggression threat; (b) the individual has been a victim of bullying, suffered defamation of isolation by anyone; (c) the individual has been discriminated due to racism; (d) the individual has been discriminated due to homophobia; (e) the individual has been in a fight with physical aggression and (f) the individual has engaged in a fight with close friends. Our results show that lottery winners are more likely to report having been victims of bullying or defamation (2.6 p.p. or 30 percent of control group mean) and the coefficient on being victim of homophobia is marginally significant and equal to 1.1 p.p, corresponding to 65 percent of the control group mean. For the LATE estimates in Panel A of Table A.4, the coefficients for these dimensions are, respectively, 11.5 p.p. and 5.0 p.p.

These findings contradicted our initial assumptions, as one of the goals of the program was to reduce victimization. However, this result may be explained by the fact that the program also invested in promoting awareness of rights and different forms of violence. So one possibility is that these findings reflect increased awareness of violence episodes, and not necessarily increased incidence. Other papers in the literature, such as Green et al. (2018), have also registered increases in reporting of violence episodes as a result of programs that promoted awareness and empowerment.

3.5.4 Engagement with Community, Religion, Politics and Sports

Community engagement was an important feature of the program, which dedicated around half of its total time to group activities (sports, culture, community development, citizenship building and community therapy). We collected very detailed data on community participation, including dimensions of political engagement, sports and religion, to understand whether lottery winners remained engaged in such types of activities after the program. We aggregate data for each of these dimensions to build four constructs for which we report results in Table 6. The estimates for each construct are from an ITT specification, and we follow Kling et al. (2007) to calculate average mean standardized effect size across multiple outcomes using the seemingly-unrelated regression framework to account for covariance across estimates. The variables used for each construct are as follows. In column (a), 10 outcomes: current member of a community council (education, health, etc.), a worker's union, an association of neighbors, a voluntary group, a youth group or other community group, participated in a community task-force to construct houses or clean the streets, participated in a solidarity campaign, participated in other community activities, in reunions or partied in the community in the last year; in column (b), 3 outcomes: current member of a religious group, participated in a mass, cult or religious ceremony, participated in other religious activities; in column (c), 5 outcomes: current member of a political party, voted in the last election, has voting documentation, correctly answered the president's name, correctly answered the governor's name; in column (d), 2 outcomes: current member of a sports club or association, practiced sports in the last year.

Our results show that the program had long lasting results in community engagement. Two years after the program, lottery winners were more likely to be involved in community activities by 0.039σ, as expected, although no differences between lottery winners and losers were registered for engagement in religion, politics or sports.

3.5.5 Education

In addition to 290 hours of Portuguese, math and technical formation course, the program also stimulated a reflection about the value of work and education in the citizenship building module. For these reasons, we would expect the program to have an effect on educational enrollment and educational perspectives.

For our ITT analysis of educational enrollment outcomes, displayed in Table 7, we present the results separately for three groups, divided according to their baseline educational status: i) those who were neither enrolled in school nor working at baseline (columns a-c); ii) those not enrolled in school at baseline

(columns d-f) and iii) those enrolled in school at baseline (columns g-i).[3] For each group, three variables were considered: (a), (d) and (g) display results for "In school in 2012 and/or 2013", an indicator variable for having reported attendance in school after the end of the program, in 2012 and/or 2013; (b), (e) and (h) display results for "In school in 2012 and 2013", an indicator variable for having reported attendance in school in 2012 and in 2013 and (c), (f) and (i) show results for an indicator variable for having reported no attendance in school in 2012 and attendance in 2013. Panel A reports results for the full sample of each group, while Panels B and C display results for the female and male subsamples respectively.

For all the combinations between these variables and subsamples, the only significant difference we find between treatment offer arms is for the variable reflecting the group of students enrolled in school at baseline, that were not in school in 2012 and returned to school in 2013, driven by the results for the female subsample. This measure is not particularly interesting as the program ended in 2011, so outcomes measured in both 2012 and 2013 were both affected by program participation. Although the program was designed to foster school engagement, on average it did not have an effect on mean school enrollment after the program.

In Table 8, we investigate the effects of the program offer on individuals' educational perspectives. Panel A displays the mean results, while Panels B and C display results for the female and male subsamples respectively. The definition of the variables in each column is the following: (a) "Dissatisfied with attainment" is an indicator of not being satisfied with the current educational level; (b), (c) and (d) are indicators of wanting to attain high-school, VET (vocational education and training) and college, respectively, conditional on being dissatisfied with the current educational level. Our results show that, conditional on being dissatisfied with current educational level, the offer to the program reduced individuals' desire to pursue a technical degree, by 3.2 p.p., which corresponds to 43 percent of the lottery losers' mean. This might be related to the fact that the program itself had a component of vocational training, which offered 160 hours of technical training for several occupations.

3.5.6 Noncognitive Skills

The *Protejo* program had several features that could potentially foster the development of different dimensions of noncognitive skills. The different modules of the program invested in community and social integration, building self-confidence and discussing the importance of hard work to reach individuals' future objectives. In principle, these dimensions may be associated with the "Big Five" dimensions, as well as with self-esteem, impulsivity and reaction to critic. In Table 9 we describe the impacts of program offer on these dimensions of noncognitive skills (described in more detail in section 3.3). For each construct reported in columns (a)-(e), computed using a 44-item inventory that measures an individual on the "Big Five" dimensions of personality, we follow Kling et al. (2007) to calculate average mean standardized effect size across multiple outcomes using the seemingly-unrelated regression framework to account for covariance across estimates. As previously specified, we report in columns (f) - (h) results for the "Rosemberg Self-esteem Scale", "Barratt Impulsiveness Scale" and "Leary Brief Fear of Negative Evaluation Scale".

[3] Group i is a subgroup of group ii.

Our results indicate that the offer to participate in the program had no significant effects on noncognitive skills based on the four different instruments we applied.

3.5.7 Labor Market

Table 10 presents evidence of the impacts of the program offer on various measures of labor market attachment, based on self-reported data from the 2013 survey. Panel A considers all individuals in the sample and Panel B considers only the currently employed individuals. Column (d) uses mean imputed work earnings for individuals who did not want to reveal exact earnings but only reported an income range, and in column (e) the definition of formality is to have a working contract, conditional on being an employee (i.e., self-employed and employers were discarded). The results we find indicate there are no differences between lottery winners and losers in the various measures of labor market outcomes we consider.[4]

In Table 11, we investigate the effects of the program offer on formal employment, based on administrative data from the Ministry of Labor (RAIS) from 2010 to 2016, more than five years after the end of the program. Details on this dataset and on the matching performed follow in section 3.3 and Appendix C respectively. Panel A considers the full registry and Panel B refers to the evaluation sample. Variables in columns (a) - (g) are indicator variables of whether an individual was found in the RAIS dataset for the years 2010-2016 respectively. We test pre-program balance between treatment arms based on the 2009 RAIS data and find no significant differences between the two groups.[5]

Consistently with our survey findings, we also register no average short run effects of the program offer on formal employment (for years 2010-2013). However, we find significant differences between treatment offer arms on formal labor market attachment from 2014 onwards, for both the full registry and the evaluation sample. In 2014, for the full registry, differences correspond to 5 percent of the lottery losers' mean (2.7 p.p.), and are marginally significant (p-value=0.115). For 2015 and 2016, winning the lottery increased the probability of being formally employed respectively by 5.7 p.p. and 3.2 p.p. (12 and 7 percent of the no offer group mean). For the evaluation sample, coefficients are even higher (0.029 for 2014, 0.059 for 2015 and 0.050 for 2016). Our LATE estimates in Appendix Table A.7 indicate ever participating in the program increases the likelihood of formal employment on average by 12 p.p. in 2014, 25 p.p. in 2015 and 14 p.p. in 2016. When we consider the fraction of total stipends received as a measure of program participation, these coefficients are even higher (ranging from 18 p.p. to 37 p.p.). The size of these effects are consistent with the impacts on employment of other successful Latin American youth training programs, such as *Jovenes en Accion*, evaluated by Attanasio et al. (2011).

Appendix Table A.1 describes treatment effects heterogeneity by gender. For the female subsample, the formal employment effects are not present in 2014, but are higher than the mean results for 2015 and 2016 (0.062 and 0.037 respectively). For the male subsample, results are significant from 2012 (in the year following the end of the program) to 2015 (marginally significant for 2013), with treatment effects ranging from 10 to 15 percent of the no offer group mean. When we explore gender heterogeneity in the survey

[4] We consider the following variables: i) Any paid job in the last year, ii) Current employment, iii) Current hours worked, iv) Current earnings, v) Current formal employment.

[5] This result is available upon request.

data, we also find a positive coefficient (0.020) in formal employment for the male subsample in 2013, although not significant, which diverges from the findings from the administrative data.[6] However, this finding is based on a much smaller sample of only 591 individuals, so we may not have enough statistical power to identify a significant treatment effect.

Appendix Table A.2 presents results for age-groups heterogeneity. We divide the sample in three age-groups (Panels A, B and C), according to the age distribution from the registry data, collected on the application to the program: i) Ages 13-16; ii) Ages 17-19 and iii) Ages 20 and above. Our findings show that the formal employment effects are being driven by the 17-19 year old group. Between 2012 and 2016, the years for which we find significant results for either the full sample or the gender-specific subsamples, the age of this group ranged from 19-23 years old.

These results suggest that, since we do not capture significant mean effects on education enrollment or on the noncognitive skills measured in the 2013 survey, these are probably not the intermediate outcomes through which *Protejo* impacts formal employment. The channel through which *Protejo* likely increases the probability of beneficiaries finding a job in the formal labor market may be associated with the improvement of skills as a result of the program's 400 hours of in-class training. The hours of training were dedicated to both the development of general human capital (130 hours of math and portuguese and 30 hours of "individual development plan") and specific human capital skills (80 hours of information technology class and 160 hours of vocational training). Unfortunately, we do not have data on the technical course selected by each program participant to further explore whether these individuals are being attached to the specific sectors of the economy for which they were trained.

3.6 Discussion and Conclusion

In this paper, we present novel experimental evidence of the impacts of *Protejo*, one of the pillars of the Brazilian's National Program of Public Security with Citizenship. *Protejo* combined elements of technical and vocational education and training programs alongside with basic math and language classes, mentoring components and activities to promote community engagement, empowerment and citizenship awareness.

We show that the program increased community engagement, as expected, but also increased the probability of having children, going in the opposite direction of our initial assumptions. We hypothesize that, since *Protejo* had components that fostered family relations and family building, this may have played a role in explaining these findings. Also contrary to our expectations, we find a higher probability of lottery winners to have been victims of bullying or defamation and homophobia. This result does not necessarily reflect increased incidence, but may be explained by the fact that the program had a module that promoted awareness of youth rights and the different types of violence in society, including domestic violence, homophobia, and racism.

We find no average effects on education, noncognitive skills or labor market outcomes based on the 2013 survey data, but find strong longer run impacts on formal employment based on administrative data from the Ministry of Labor. The significant results hold for both the male and female subsamples,

[6] This result is available upon request.

and are driven by the outcomes from the 17-19 age group. The positive effects for the male subsample arise one year after the end of the program and fade out four years later, while large and significant effects for females emerge four years after the program. The size of these effects are comparable to those from interventions aimed at increasing skills and labor market outcomes for disadvantaged youth in developing countries.

As we find no average effects on education, we speculate that the program may be changing youth trajectories by enhancing, through its 400 hours of in-class training, a set of skills that are important predictors of labor market attachment and crime. Although we do not have data to investigate the program's impact on crime and youth mortality, there is extensive empirical evidence showing a negative association between crime and employment (HELLER, 2014; GELBER et al., 2015). A promising avenue for future research is to measure whether *Protejo*'s formal employment effects persist in the long run and whether crime and mortality rates are lower for beneficiaries.

FIGURES AND TABLES

Figure 1 – *Protejo* Program by State - 2008-2013

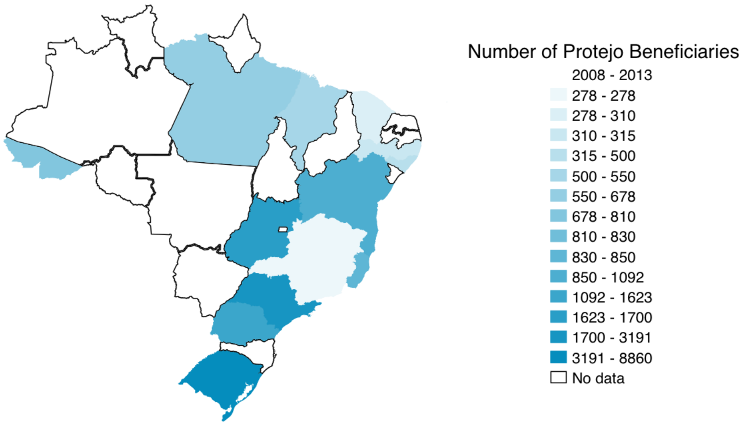
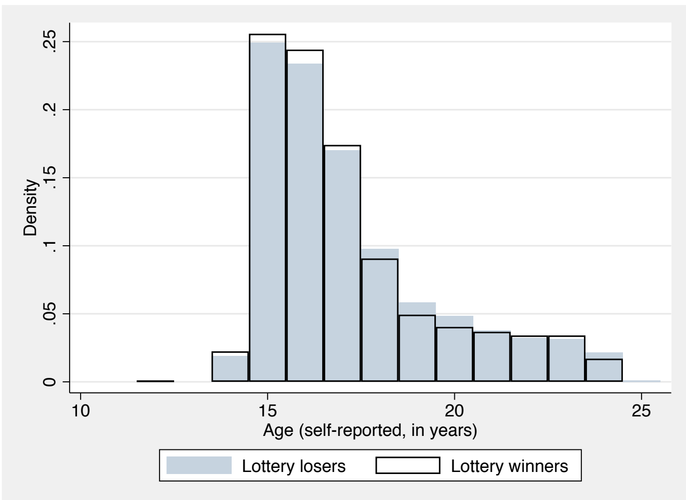


Figure 2 – Age at Baseline by Lottery Status



Notes: This histogram is based on baseline data for the evaluation sample

Table 1 – Tests of Randomization Balance, by Treatment Offer

	Full registry data				*Sample*			
	Mean (Loser) (1)	Mean (Winner) (2)	p-value ((1)=(2)) (3)	Obs. (4)	Mean (Loser) (5)	Mean (Winner) (6)	p-value ((5)=(6)) (7)	Obs. (8)
Age	17.56 (6.94)	17.36 (11.10)	0.916	4,837	17.57 (8.01)	17.40 (7.83)	0.603	2,241
"Pardo" or Black	0.73 (0.44)	0.73 (0.45)	0.830	4,857	0.73 (0.44)	0.75 (0.43)	0.359	2,248
Household size	5.43 (22.33)	5.20 (14.15)	0.342	4,464	5.86 (28.78)	5.40 (17.70)	0.618	2,083
Has a child	0.15 (0.36)	0.14 (0.35)	0.346	4,730	0.18 (0.39)	0.18 (0.38)	0.824	2,195
Attended school	0.79 (0.41)	0.78 (0.41)	0.994	4,819	0.76 (0.43)	0.78 (0.42)	0.267	2,231
Literate	0.99 (0.08)	0.99 (0.11)	0.959	4,806	0.99 (0.09)	0.99 (0.09)	0.987	2,223
Mother was alive	0.96 (0.20)	0.94 (0.23)	0.827	4,836	0.95 (0.22)	0.94 (0.24)	0.251	2,239
Mother is literate	0.90 (0.30)	0.89 (0.31)	0.440	4,568	0.88 (0.33)	0.88 (0.33)	0.975	2,110
Worked in August 2010	0.13 (0.34)	0.18 (0.38)	0.271	4,725	0.15 (0.36)	0.16 (0.36)	0.591	2,188
Satisfied with current educational level	0.60 (0.49)	0.58 (0.49)	0.702	4,800	0.56 (0.50)	0.60 (0.49)	0.108	2,217
Self-esteem indicator (from -4 to 4)	0.64 (0.43)	0.59 (0.47)	0.981	4,845	0.62 (0.47)	0.61 (0.45)	0.560	2,243
Impulsivity indicator (from -4 to 4)	0.46 (0.40)	0.49 (0.39)	0.736	4,845	0.48 (0.43)	0.49 (0.39)	0.748	2,243
Victim of violence in the last year	0.09 (0.28)	0.17 (0.37)	0.854	4,704	0.13 (0.34)	0.13 (0.33)	0.793	2,183
Ever had contact with alcohol	0.36 (0.48)	0.40 (0.49)	0.742	4,810	0.39 (0.49)	0.39 (0.49)	0.850	2,230
Ever had contact with cocaine	0.01 (0.10)	0.03 (0.17)	0.072	4,780	0.01 (0.11)	0.02 (0.13)	0.371	2,215
Ever had contact with crack	0.01 (0.08)	0.02 (0.14)	0.075	4,768	0.01 (0.08)	0.02 (0.12)	0.071	2,208
p-value joint (F-statistic)			0.591				0.552	

Notes – An observation is an eligible individual found in the registry data collected in 2010 (columns 1- 4) or a sampled eligible individual (columns 5-8). Each line presents basic descriptive statistics for a baseline variable, separately for lottery winners and losers. Standard deviations are in parentheses and p-values in columns (3) and (7) are for a test of difference between groups computed using regressions with strata dummies and heteroskedasticity-robust standard errors.

Table 2 – Impact of Lottery on Likelihood of Attrition in the Endline Sample

	(a) Attriter	(b) Attriter	(c) Attriter
Lottery winner dummy	-0.024	-0.024	-0.028
(p-value)	(0.161)	(0.151)	(0.131)
Lottery loser's mean	0.221	0.221	0.216
Number of observations	2,248	2,248	1,772
Controls			
Strata dummies		✓	✓
Baseline characteristics			✓

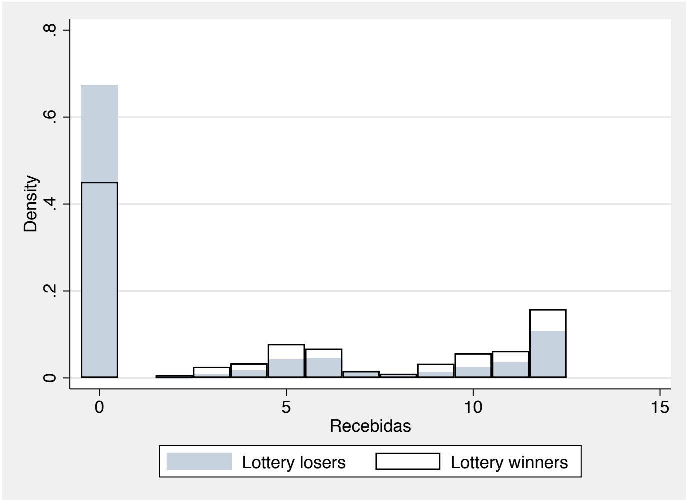
Notes – An observation is a sampled eligible individual. The table reports the coefficient of an indicator of whether the person leaves the sample in the follow-up survey on the treatment offer indicator, using the controls marked in the bottom part. Baseline characteristics are listed in Table 1. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table 3 – Impact of Lottery on Program Participation

	(a) Self-reported participation in Protejo	(b) Ever received a monthly stipend	(c) Number of Monthly Stipends Received	(d) Fraction of total potential stipends
Lottery Winner	0.195	0.221	1.792	0.149
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)
Lottery loser's mean	0.516	0.327	2.87	0.239
Number of Strata	93	93	93	93
Number of Observations	1641	2248	2248	2248

Notes – An observation is a sampled eligible individual. This table displays the first stage results; i.e results of a regression following intent-to-treat (ITT) specification (3.1) without controls, using different measures for program participation. Description of variables: (a) "Self reported participation in Protejo" is an indicator of having participated in Protejo, according to self-reported survey data, collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program); b) "Ever received a monthly stipend" is an indicator of being present in the administrative program's dataset which contains the names of beneficiaries that ever received a program's stipend; c) "Number of Monthly Stipends" is the total number of monthly stipends received by the program's participants. This variables ranges from 1-12 for program's participants, and takes value 0 if never received a stipend and hence not present in the program's database. d) "Fraction of stipends" is the variable described in column c) divided by 12, which is the total potential number of months a beneficiary could have received the program's stipends. This variable takes value zero if never received a stipend. p-values are in parentheses, and heteroskedasticity-robust standard errors in brackets.

Figure 3 – Months receiving the Program's Stipend by Lottery Status



Notes: This histogram is based on administrative data reflecting number of stipends received by lottery participants in the evaluation sample

Table 4 – Intention-to-treat Estimates of Lottery Status on Family Formation and Social Interactions

	(a) Has children	(b) Uses prevention	(c) In a relationship	(d) Number of close friends	(e) Family proximity index
Lottery winner	0.035	0.016	0.006	-0.298	-0.007
(p-value)	(0.083)	(0.424)	(0.799)	(0.121)	(0.519)
Lottery loser's mean	0.285	0.840	0.593	4.090	0.672
Number of strata	93	93	93	93	93
Number of observations	1,779	1,242	1,779	1,779	1,778

Notes – An observation is a sampled eligible individual. All variables are self-reported and were collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program). The estimates in each column are from the intent-to-treat (ITT) specification (3.1) without controls. Description of variables: (a) "Has children" is an indicator of having any child or pregnancy at the time of the survey; (b) "Uses prevention" is an indicator of using at least one in 12 contraceptive methods with the current partner at the time of the survey, if applicable; (c) "In a relationship" is an indicator of being married, cohabiting or having a partner; (d) "Number of close friends" is a categorical variable with the number of current friends considered "close friends"; (e) "Family proximity index" is the fraction of six groups of family members (mother, father, grandparents, brother/sisters, offspring and others) with a "very close" relationship with the interviewed. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table 5 – Intention-to-treat Estimates of Lottery Status on Conflict

	(a) Victim of physical aggression	(b) Victim of bullying or defamation	(c) Victim of racism	(d) Victim of homophobia	(e) Engaged in fight with physical aggression	(f) Engaged in fight with close friends
Lottery Winner	0.014	0.026	0.012	0.011	0.002	0.024
(p-value)	(0.437)	(0.068)	(0.311)	(0.120)	(0.919)	(0.170)
Lottery loser's mean	0.180	0.086	0.063	0.017	0.24	0.145
Number of Strata	93	93	93	93	93	93
Number of Observations	1751	1764	1765	1764	1748	1779

Notes – An observation is a sampled eligible individual. All variables are self-reported and were collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program). The estimates in each column are from the intent-to-treat (ITT) specification (3.1) without controls. Description of variables: (a) An indicator of whether the individual has been a victim of a physical aggression, or verbal aggression with physical aggression threat; (b) An indicator of whether the individual has been a victim of bullying, suffered defamation of isolation by anyone; (c) An indicator of whether the individual has been discriminated due to racism; (d) An indicator of whether the individual has been discriminated due to homophobia; (e) An indicator of whether the individual has been in a fight with physical aggression; (f) An indicator of whether the individual has engaged in a fight with close friends. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table 6 – Intention-to-treat Estimates of Lottery Status on Engagement with the Community, Religion, Politics and Sports

	(a) Community	(b) Religion	(c) Politics	(d) Sports
Lottery winner	0.039	0.014	-0.030	0.030
(p-value)	(0.095)	(0.708)	(0.268)	(0.430)
Number of strata	93	93	93	93
Number of observations	1,779	1,779	1,779	1,779

Notes – An observation is a sampled eligible individual. All variables are self-reported and were collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program). The estimates for each construct in the columns are from intent-to-treat specifications (3.1) without controls. We follow (KLING et al., 2007) to calculate average mean standardized effect size across multiple outcomes using the seemingly-unrelated regression framework to account for covariance across estimates. Description of variables used in each column: (a) 10 outcomes: current member of a community council (education, health, etc.), a worker's union, an association of neighbors, a voluntary group, a youth group or other community group, participated in a "mutirão" to construct houses or clean the streets, participated in a solidarity campaign, participated in other community activities, in reunions or partied in the community in the last year; (b) 3 outcomes: current member of of a religious group, participated in a mass, cult or religious ceremony, participated in other religious activities; (c) 5 outcomes: current member of a political party, voted in the last election, has voting documentation, correctly answered the president's name, correctly answered the governor's name; (d) 2 outcomes: current member of a sports club or association, practiced sports in the last year. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table 7 – Intention-to-treat Estimates of Lottery Status on Education

	Conditional on ...								
	...neither being in school or working at baseline			*...not being in school at baseline*			*...being in school at baseline*		
	(a) In school in 2012 and/or 2013	(b) In school in 2012 and 2013	(c) Not in school in 2012 but in school in 2013	(d) In school in 2012 and/or 2013	(e) In school in 2012 and 2013	(f) Not in school in 2012 but in school in 2013	(g) In school in 2012 and/or 2013	(h) In school in 2012 and 2013	(i) Not in school in 2012 but in school in 2013
Panel A: Full sample									
Lottery Winner	-0,062	0,015	0,019	-0,063	0,003	0,016	0,020	-0,015	0,022
(p-value)	(0,329)	(0,725)	(0,497)	(0,247)	(0,930)	(0,526)	(0,396)	(0,554)	(0,023)
Lottery loser's mean	0,338	0,089	0,039	0,346	0,097	0,041	0,711	0,381	0,023
Number of Strata	86	86	86	89	89	89	93	93	93
Number of Observations	312	312	312	384	384	384	1.381	1.381	1.381
Panel B: Female subsample									
Lottery Winner	-0,078	-0,002	0,028	-0,065	-0,011	0,031	0,013	-0,013	0,040
(p-value)	(0,229)	(0,966)	(0,349)	(0,271)	(0,790)	(0,281)	(0,685)	(0,692)	(0,006)
Lottery loser's mean	0,344	0,108	0,030	0,355	0,117	0,034	0,710	0,378	0,022
Number of Strata	54	54	54	56	56	56	56	56	56
Number of Observations	252	252	252	303	303	303	790	790	790
Panel C: Male subsample									
Lottery Winner	0,055	0,131	-0,044	-0,055	0,071	-0,057	0,030	-0,017	-0,001
(p-value)	(0,805)	(0,235)	(0,497)	(0,700)	(0,185)	(0,260)	(0,410)	(0,656)	(0,938)
Lottery loser's mean	0,289	-0,016	0,089	0,314	0,011	0,080	0,713	0,386	0,022
Number of Strata	32	32	32	33	33	33	37	37	37
Number of Observations	60	60	60	81	81	81	591	591	591

Notes – An observation is a sampled eligible individual. All variables are self-reported and were collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program). Panel A displays the mean results, while Panels B and C display results for the female and male subsamples respectively. The estimates in each column are from the intent-to-treat specification (3.1) without controls. Columns a-c present results for the subsample of individuals who reported not being in school or working at baseline; columns d-f present results for the subsample of individuals who reported not being in school at baseline and columns g-i present results for the subsample of individuals who reported being in school at baseline. Description of variables: (a), (d) and (g) "In school in 2012 and/or 2013" is an indicator variable for having reported attendance in school in 2012 and/or 2013; (b) "In school in 2012 and/or in 2013" is an indicator variable for having reported attendance in school in 2012 and in 2013; (b), (e) and (h) "In school in 2012 and/or 2013" is an indicator variable for having reported attendance in school in 2012 and in 2013 and (c), (f) and (i) is an indicator variable for having reported no attendance in school in 2012 and attendance in 2013. *p*-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table 8 – Intention-to-treat Estimates of Lottery Status on Educational Perspectives

		Conditional on being dissatisfied...		
	(a) Dissatisfied w. attainment	(b) ... would like HS	(c) ... would like VET	(d) ... would like College
Panel A: Mean results				
Lottery winner	0.017	0.030	-0.032	0.042
(p-value)	(0.454)	(0.245)	(0.051)	(0.218)
Lottery loser's mean	0.483	0.162	0.073	0.479
Number of strata	93	93	93	93
Number of observations	1,779	875	875	875
Panel B: Results for the Female Subsample				
Lottery Winner	-0.013	0.021	-0.037	0.029
(p-value)	(0.667)	(0.492)	(0.064)	(0.485)
Lottery loser's mean	0.523	0.16	0.072	0.498
Number of Strata	56	56	56	56
Number of Observations	1101	569	569	569
Panel C: Results for the Male Subsample				
Lottery Winner	0.066	0.046	-0.023	0.067
(p-value)	(0.077)	(0.309)	(0.439)	(0.26)
Lottery loser's mean	0.419	0.165	0.074	0.442
Number of Strata	37	37	37	37
Number of Observations	678	306	306	306

Notes – An observation is a sampled eligible individual. All variables are self-reported and were collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program). Panel A displays the mean results, while Panels B and C display results for the female and male subsamples respectively. The estimates in each column are from the intent-to-treat specification (3.1) without controls. Description of variables: (a) "Dissatisfied with attainment" is an indicator of not being satisfied with the current educational level; (b), (c) and (d) are indicators of wanting to attain high-school, VET (vocational education and training) and college, respectively, conditional on being dissatisfied with the current educational level. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table 9 – Intention-to-treat Estimates of Lottery Status on Noncognitive Skills

	"Big Five"							
	(a) Extraversion	(b) Agreeableness	(c) Conscientiousness	(d) Neuroticism	(e) Openness to experience	(f) Self-esteem	(g) Impulsivity	(h) Reaction to critic
Lottery winner	-0.017	-0.001	-0.008	0.025	-0.027	0.008	0.052	0.008
(p-value)	(0.346)	(0.955)	(0.739)	(0.315)	(0.246)	(0.857)	(0.277)	(0.868)
Number of strata	93	93	93	93	93	93	93	93
Number of observations	1,779	1,779	1,779	1,779	1,779	1,779	1,779	1,779

Notes – An observation is a sampled eligible individual. All variables are self-reported and were collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program). The estimates for each construct in the columns are from intent-to-treat specifications (3.1) without controls. We follow (KLING et al., 2007) to calculate average mean standardized effect size across multiple outcomes using the seemingly-unrelated regression framework to account for covariance across estimates. Description of variables used in each column: (a), (b), (c), (d), (e) are computed using a 44-item inventory that measures an individual on the "Big Five" dimensions of personality; (f) is computed using the standardized "Rosemberg Self-esteem Scale"; (g) is computed using the standardized "Barratt Impulsiveness Scale"; (h) is computed using the standardized "Leary Brief Fear of Negative Evaluation Scale". More information on these survey instruments can be found in the Section 3.3. p-values computed using heteroskedasticity-robust standard errors are in parentheses

Table 10 – Intention-to-treat Estimates of Lottery Status on Labor Market Outcomes - self reported 2013 survey data

	(a) Any paid job, last year	(b) Employed, currently	(c) Hours worked, currently	(d) Earnings currently	(e) Formally em-, ployed currently
Panel A: Unconditional					
Lottery winner	-0.001	-0.002	-0.686	6.125	0.020
(p-value)	(0.972)	(0.942)	(0.496)	(0.731)	(0.347)
Lottery loser's mean	0.178	0.496	18.6	242.5	0.245
Number of strata	93	93	93	93	93
Number of observations	1,779	1,779	1,779	1,779	1,621
Panel B: Conditional on employment					
Lottery winner dummy			-1.155	18.724	0.053
p-value			(0.329)	(0.511)	(0.158)
Lottery loser's mean			37.4	486.5	0.545
Number of strata			93	93	92
Number of observations			881	881	723

Notes – An observation is a sampled eligible individual. All variables are self-reported and were collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program). The estimates in this table are from the intent-to-treat specification (3.1) without controls. Panel A considers all individuals in the sample and Panel B considers the only currently employed individuals. Column (d) uses mean imputed work earnings for individuals who did not want to reveal exact earnings but only reported an income range, and in column (e) the definition of formality is to have a working contract ["carteira assinada"], conditional on being an employee (i.e., self-employed and employers were discarded). p-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table 11 – Intention-to-treat Estimates of Lottery Status on Formal Employment (RAIS) 2010 - 2016

	Present in RAIS						
	(a) In 2010	(b) In 2011	(c) In 2012	(d) In 2013	(e) In 2014	(f) In 2015	(g) In 2016
Panel A: Full registry							
Lottery Winner	-0.013	0.005	0.022	0.011	0.027	0.057	0.032
(p-value)	(0.288)	(0.747)	(0.178)	(0.534)	(0.115)	(0.001)	(0.066)
Lottery loser's mean	0.157	0.238	0.338	0.446	0.512	0.479	0.446
Number of Strata	152	152	152	152	152	152	152
Number of Observations	4.841	4.841	4.841	4.841	4.841	4.841	4.841
Panel B: Sample							
Lottery Winner	-0.020	0.005	0.013	-0.010	0.029	0.059	0.050
(p-value)	(0.169)	(0.782)	(0.501)	(0.642)	(0.165)	(0.005)	(0.016)
Lottery loser's mean	0.150	0.218	0.327	0.428	0.499	0.458	0.422
Number of Strata	93	93	93	93	93	93	93
Number of Observations	2.241	2.241	2.241	2.241	2.241	2.241	2.241

Notes – An observation is a sampled eligible individual. All variables are based on administrative data from the Ministry of Labor, RAIS, from years 2010-2016, matched to the Protejo dataset. Details on the matching performed follow in appendix. The estimates presented are from the intent-to-treat specification (3.1) without controls. Panel A considers the full registry and Panel B the evaluation sample. Variables in columns (a-g) are indicator variables of whether an individual was found in the RAIS dataset for the years 2010-2016 respectively. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

APPENDIX A. ITT Heterogeneity

Table A.1 – Intention-to-treat Estimates of Lottery Status on Formal Employment by Gender (RAIS) 2010 - 2016

	Present in RAIS						
	(a) In 2010	(b) In 2011	(c) In 2012	(d) In 2013	(e) In 2014	(f) In 2015	(g) In 2016
Panel A: Female subsample							
Lottery Winner	-0.024	0.002	0.007	-0.010	-0.002	0.062	0.037
(p-value)	(0.110)	(0.894)	(0.752)	(0.642)	(0.944)	(0.005)	(0.088)
Lottery loser's mean	0.162	0.242	0.341	0.444	0.511	0.459	0.428
Number of Strata	76	76	76	76	76	76	76
Number of Observations	2.793	2.793	2.793	2.793	2.793	2.793	2.793
Panel B: Male subsample							
Lottery Winner	0.006	0.008	0.048	0.044	0.075	0.050	0.023
(p-value)	(0.751)	(0.723)	(0.075)	(0.112)	(0.008)	(0.075)	(0.417)
Lottery loser's mean	0.144	0.232	0.323	0.434	0.493	0.508	0.475
Number of Strata	76	76	76	76	76	76	76
Number of Observations	2.048	2.048	2.048	2.048	2.048	2.048	2.048

Notes – An observation is a sampled eligible individual. All variables are based on administrative data from the Ministry of Labor, RAIS, from years 2010-2016, matched to the Protejo dataset. Details on the matching performed follow in appendix. Panel A reports results for the female subsample, and Panel B for the male subsample. The estimates presented are from the intent-to-treat specification (3.1) without controls. Variables in columns (a-g) are indicator variables of whether an individual was found in the RAIS dataset for the years 2010-2016 respectively. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table A.2 – Intention-to-treat Estimates of Lottery Status on Formal Employment by Age Group (RAIS) 2010 - 2016

	Present in RAIS						
	(a) In 2010	(b) In 2011	(c) In 2012	(d) In 2013	(e) In 2014	(f) In 2015	(g) In 2016
Panel A: Ages 13-16 in 2010							
Lottery Winner	-0,022	-0,017	-0,009	-0,012	0,011	0,010	0,013
(p-value)	(0,073)	(0,304)	(0,684)	(0,598)	(0,661)	(0,690)	(0,597)
Lottery loser's mean	0,089	0,137	0,248	0,399	0,492	0,488	0,454
Number of Strata	150	150	150	150	150	150	150
Number of Observations	2.474	2.474	2.474	2.474	2.474	2.474	2.474
Panel B: Ages 17-19 in 2010							
Lottery Winner	-0,034	-0,001	0,054	0,035	0,062	0,136	0,036
(p-value)	(0,139)	(0,977)	(0,088)	(0,261)	(0,051)	(0,000)	(0,262)
Lottery loser's mean	0,186	0,331	0,433	0,499	0,535	0,471	0,456
Number of Strata	151	151	151	151	151	151	151
Number of Observations	1.601	1.601	1.601	1.601	1.601	1.601	1.601
Panel C: Ages 20 and above in 2010							
Lottery Winner	0,038	0,059	0,031	0,020	0,002	0,026	0,042
(p-value)	(0,377)	(0,194)	(0,503)	(0,664)	(0,961)	(0,578)	(0,346)
Lottery loser's mean	0,348	0,403	0,460	0,505	0,541	0,492	0,428
Number of Strata	137	137	137	137	137	137	137
Number of Observations	768	768	768	768	768	768	768

Notes – An observation is a sampled eligible individual. All variables are based on administrative data from the Ministry of Labor, RAIS, from years 2010-2016, matched to the Protejo dataset. Details on the matching performed follow in appendix. Panels A, B and C report results for the subsamples of age groups (13-16), (17-19) and (20 and above). Age based on self-reported information from the 2010 registry, before the beginning of the program. The estimates presented are from the intent-to-treat specification (3.1) without controls. Variables in columns (a-g) are indicator variables of whether an individual was found in the RAIS dataset for the years 2010-2016 respectively. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

APPENDIX B. Instrumental Variables - LATE

Table A.3 – LATE Estimates of Protejo Participation on Family Formation and Social Interactions

	(a) Has children	(b) Uses prevention	(c) In a relationship	(d) Number of close friends	(e) Family proximity index
Panel A: Ever					
Participation measure	0.154	0.074	0.026	-1.305	-0.032
(p-value)	(0.084)	(0.410)	(0.794)	(0.116)	(0.507)
F-statistic	103.53	65.18	103.53	103.53	103.10
Panel B: Fraction					
Participation measure	0.236	0.119	0.039	-1.999	-0.049
(p-value)	(0.090)	(0.413)	(0.794)	(0.119)	(0.508)
F-statistic	66.71	39.82	66.71	66.71	66.64
Control complier mean	0.310	0.852	0.584	3.892	0.677
Number of observations	1,779	1,242	1,779	1,779	1,778

Notes – An observation is a sampled eligible individual. All variables are self-reported and were collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program). The estimates in this table are from the IV specification from (3.3) without controls. Lottery is an instrument for program participation. We consider two measures of program participation, both based on administrative data: in Panel A, we consider whether an individual ever received a stipend from Protejo program and, in Panel B, we consider the fraction of total potential stipends (12) each individual received (this variable takes value zero if never received a stipend). Variables in each column are the same as those used in the ITT specification. Kleibergen-paap F-statistic for weak instruments test is reported. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table A.4 – LATE Estimates of Protejo Participation on Conflict

	(a) Victim of physical aggression	(b) Victim of bullying or defamation	(c) Victim of racism	(d) Victim of homophobia	(e) Engaged in fight with physical aggression	(f) Engaged in fight with close friends
Panel A: Ever						
Participation measure	0.064	0.115	0.053	0.050	0.009	0.104
(p-value)	(0.425)	(0.066)	(0.299)	(0.116)	(0.917)	(0.162)
F-statistic	95.324	98.476	102.231	97.272	97.539	102.453
Number of Observations	1751	1764	1765	1764	1748	1779
Panel B: Fraction						
Participation measure	0.100	0.177	0.082	0.076	0.014	0.159
(p-value)	(0.426)	(0.068)	(0.300)	(0.118)	(0.917)	(0.164)
F-statistic	59.195	61.860	64.472	61.159	60.574	65.245
Control complier mean	0.174	0.084	0.61	0.020	0.235	0.138
Number of Observations	1751	1764	1765	1764	1748	1779

Notes – An observation is a sampled eligible individual. All variables are self-reported and were collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program). The estimates in this table are from the IV specification from (3.3) without controls. Lottery is an instrument for program participation. We consider two measures of program participation, both based on administrative data: in Panel A, we consider whether an individual ever received a stipend from Protejo program and, in Panel B, we consider the fraction of total potential stipends (12) each individual received (this variable takes value zero if never received a stipend). Variables in each column are the same as those used in the ITT specification. Kleibergen-paap F-statistic for weak instruments test is reported. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table A.5 – LATE Estimates of Protejo Participation on Education

	Conditional on ...								
	...neither being in school or working at baseline			*...not being in school at baseline*			*...being in school at baseline*		
	(a) In school in 2012 and/or 2013	(b) In school in 2012 and 2013	Not in school in 2012 but in school in 2013	(d) In school in 2012 and/or 2013	(e) In school in 2012 and 2013	(f) Not in school in 2012 but in school in 2013	(g) In school in 2012 and/or 2013	(h) In school in 2012 and 2013	(i) Not in school in 2012 but in school in 2013
Panel A: Ever									
Participation measure	-0.322	0.077	0.099	-0.429	0.021	0.112	0.087	-0.063	0.095
(p-value)	(0.286)	(0.677)	(0.430)	(0.234)	(0.920)	(0.476)	(0.377)	(0.542)	(0.023)
F-statistic	9.202	9.202	9.202	7.509	7.509	7.509	84.427	84.427	84.427
Number of Observations	312	312	312	384	384	384	1.381	1.381	1.381
Panel B: Fraction									
Participation measure	-0.636	0.152	0.195	-0.866	0.042	0.226	0.127	-0.092	0.139
(p-value)	(0.318)	(0.676)	(0.446)	(0.276)	(0.920)	(0.492)	(0.376)	(0.543)	(0.025)
F-statistic	4.325	4.325	4.325	3.484	3.484	3.484	56.128	56.128	56.128
Control complier mean	0.310	0.088	0.026	0.324	0.108	0.028	0.685	0.369	0.026
Number of Observations	312	312	312	384	384	384	1,381	1,381	1,381

Notes – An observation is a sampled eligible individual. All variables are self-reported and were collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program). The estimates in this table are from the IV specification from (3.3) without controls. Lottery is an instrument for program participation. We consider two measures of program participation, both based on administrative data: in Panel A, we consider whether an individual ever received a stipend from Protejo program and, in Panel B, we consider the fraction of total potential stipends (12) each individual received (this variable takes value zero if never received a stipend). Variables in each column are the same as those used in the ITT speficiation. Kleibergen-paap F-statistic for weak instruments test is reported. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table A.6 – LATE Estimates of Protejo Participation on Labor Market Outcomes

	(a) Any paid job, last year	(b) Employed, currently	(c) Hours worked, currently	(d) Earnings currently	(e) Formally em-, ployed currently
Panel A: Ever					
Participation measure	-0.003	-0.007	-3.005	26.819	0.088
(p-value)	(0.971)	(0.940)	(0.486)	(0.724)	(0.334)
F-statistic	102.453	102.453	102.453	102.453	90.145
Number of Observations	1779	1779	1779	1779	1621
Panel B: Fraction					
Participation measure	-0.004	-0.011	-4.601	41.069	0.134
(p-value)	(0.971)	(0.940)	(0.486)	(0.724)	(0.337)
F-statistic	65.245	65.245	65.245	65.245	58.527
Control complier mean	0.196	0.465	17.272	215.158	0.236
Number of Observations	1779	1779	1779	1779	1621

Notes – An observation is a sampled eligible individual. All variables are self-reported and were collected in the first semester of 2013 (i.e., approximately 2 years after the end of the program). The estimates in this table are from the IV specification from (3.3) without controls. Lottery is an instrument for program participation. We consider two measures of program participation, both based on administrative data: in Panel A, we consider whether an individual ever received a stipend from Protejo program and, in Panel B, we consider the fraction of total potential stipends (12) each individual received (this variable takes value zero if never received a stipend). Variables in each column are the same as those used in the ITT specification. Kleibergen-paap F-statistic for weak instruments test is reported. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

Table A.7 – LATE Estimates of Lottery Status on Formal Employment (RAIS) 2010 - 2016

	Present in RAIS						
	(a) In 2010	(b) In 2011	(c) In 2012	(d) In 2013	(e) In 2014	(f) In 2015	(g) In 2016
Panel A: Ever							
Participation measure	-0,054	0,020	0,095	0,045	0,118	0,247	0,137
(p-value)	(0,277)	(0,744)	(0,173)	(0,527)	(0,109)	(0,001)	(0,061)
F-statistic	191.046	191.046	191.046	191.046	191.046		
Number of Observations	4.841	4.841	4.841	4.841	4.841	4.841	4.841
Panel B: Fraction							
Participation measure	-0,081	0,030	0,143	0,068	0,177	0,372	0,206
(p-value)	(0,276)	(0,744)	(0,175)	(0,527)	(0,110)	(0,001)	(0,061)
F-statistic	129.984	129.984	129.984	129.984	129.984		
Control complier mean	0.192	0.269	0.360	0.446	0.502	0.471	0.433
Number of Observations	4.841	4.841	4.841	4.841	4.841	4.841	4.841

Notes – An observation is a sampled eligible individual. All variables are based on administrative data from the Ministry of Labor, RAIS, from years 2010-2016. The estimates in this table are from the IV specification from (3.3) without controls. Lottery is an instrument for program participation. We consider two measures of program participation, both based on Protejo's administrative data: in Panel A, we consider whether an individual ever received a stipend from Protejo program and, in Panel B, we consider the fraction of total potential stipends (12) each individual received (this variable takes value zero if never received a stipend). Variables in each column are the same as those used in the ITT specification. Kleibergen-paap F-statistic for weak instruments test is reported. p-values computed using heteroskedasticity-robust standard errors are in parentheses.

APPENDIX B. Lottery Results by Community

Figure A.1 – Lottery Results by Community

	Lottery Results by Community		
	Community	Lottery loser	Lottery Winner
		N	N
1	Complexo do Alemão	99	150
2	Cantagalo	15	75
3	Caxias	71	150
4	Itaboraí	188	150
5	Manguinhos	138	150
6	Niterói	113	100
7	Providência	51	100
8	Rocinha	193	75
9	São João do Meriti	52	150
10	Tavares Bastos	38	50
11	Vila Kennedy	98	150
12	Belford Roxo	186	150
13	Itaguaí	152	150
14	Mesquita	176	150
15	Nilopolis	173	150
16	Nova Iguacu	238	150
17	Pavao-Pavaozinho	58	75
18	Queimados	129	150
19	São Goncalo	264	150
	Total	2,432	2,425

APPENDIX C. Matching with Administrative RAIS dataset

The lottery records had information on full name for all individuals, and date of birth and government identity number (cpf) for only a portion of the sample. For the cases of which we had *cpf*, the matching with the Ministry of Labor dataset was straightforward. For the remainder of the observations, we followed the same procedure used by Arabage (2018) to merge data with string variables. The first stage was to match by name, and matches with low scores were discarded. In the second stage, we compared the date of birth and cpf for each matched pair (when this information was available) and constructed the sample of exact matches (34% of the cases)

For the remaining observations, when the only information we had available was the name of the individual, and for the cases in which there was more than one person with the same name and different cpfs in the database, a randomization of individuals within the same name group was carried out, maintaining only one observation per name per year.

The following step was to perform a visual analysis of each pair of non-exact names and classify them as correct, incorrect, or uncertain matches. Matching pairs of names differing only by typos (e.g, “SANTOS” and “SNTOS”) were considered correct matches. Unlikely combinations were classified as incorrect (eg “RAMAL” and “CARVALHO”). The fourth step was to classify the combinations as uncertain if there was a high similarity between them, but without the certainty that they actually referred to the same individual (e.g. “SOUZA” and “DE SOUZA”). The combinations considered incorrect were discarded. The percentage of observations matched by each method follow below.

Table A.8 – Percentage of Observations Matched by Each Method

Method	Percent of Observations
Match by CPF	0.295
Match by Date of Birth	0.047
Match by Name (Single obs)	0.191
Match by Name (Multiple obs - random choice)	0.094
Slight Difference Match	0.127
Not Matched	0.246
Total	1.000

References

ALFONSO, V. C.; FLANAGAN, D. P.; RADWAN, S. The impact of the cattell-horn-carroll theory on test development and interpretation of cognitive and academic abilities. *Contemporary intellectual assessment: Theories, tests, and*, n. 2nd, p. 185–202, 2005.

ALMOND, D.; CURRIE, J. *Human Capital Development Before Age Five*. [S.l.], 2010. Available from Internet: <http://www.nber.org/papers/w15827>.

ALZÚA, M. L.; CRUCES, G.; ERAZO, C. L. Youth training programs beyond employment. evidence from a randomized controlled trial. *La Plata, Buenos Aires: CEDLAS.(Mimeo.)*, 2013.

ANDRADE, J. M. de. *Evidências de validade do inventário dos cinco grandes fatores de personalidade para o Brasil*. Tese (Doutorado) — Ph. D. thesis, Universidade de Brasília, 2008.

ANGELES, G. et al. The impact of daycare on maternal labor supply and child development in mexico. *Impact Evaluation Report*, v. 6, 2014.

ATTANASIO, O. et al. *Long term impacts of vouchers for vocational training: experimental evidence for Colombia*. [S.l.], 2015.

ATTANASIO, O.; KUGLER, A.; MEGHIR, C. Subsidizing vocational training for disadvantaged youth in colombia: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, American Economic Association, v. 3, n. 3, p. 188–220, 2011.

ATTANASIO, O.; VERA-HERNANDEZ, M. Medium-and long run effects of nutrition and child care: evaluation of a community nursery programme in rural colombia. Institute for Fiscal Studies. Working Paper EWP04/06, 2004.

ATTANASIO, O. P. The determinants of human capital formation during the early years of life: Theory, measurement, and policies. *Journal of the European Economic Association*, Oxford University Press, v. 13, n. 6, p. 949–997, 2015.

ATTANASIO, O. P.; MARO, V. D.; VERA-HERNÁNDEZ, M. Community nurseries and the nutritional status of poor children. evidence from colombia. *The Economic Journal*, Wiley Online Library, v. 123, n. 571, p. 1025–1058, 2013.

BAKER, M.; GRUBER, J.; MILLIGAN, K. Universal child care, maternal labor supply, and family well-being. *Journal of political Economy*, The University of Chicago Press, v. 116, n. 4, p. 709–745, 2008.

BAKER, M.; GRUBER, J.; MILLIGAN, K. *Non-cognitive deficits and young adult outcomes: The long-run impacts of a universal child care program*. [S.l.], 2015.

BANERJEE, A. V. et al. Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*, MIT Press, v. 122, n. 3, p. 1235–1264, 2007.

BARROS, R. et al. The impact of access to free childcare on women's labor market outcomes: evidence from a randomized trial in low-income neighborhoods of rio de janeiro. In: *World Bank Economists' Forum*. [S.l.: s.n.], 2011.

BAUERNSCHUSTER, S.; SCHLOTTER, M. Public child care and mothers' labor supply—evidence from two quasi-experiments. *Journal of Public Economics*, Elsevier, v. 123, p. 1–16, 2015.

BEHRMAN, J. R.; DEOLALIKAR, A. B. Health and nutrition. *Handbook of development economics*, Elsevier, v. 1, p. 631–711, 1988.

BERLINSKI, S.; GALIANI, S.; EWAN, P. J. M. Preschool and maternal labor market outcomes: evidence from a regression discontinuity design. *Economic Development and Cultural Change*, University of Chicago Press Chicago, IL, v. 59, n. 2, p. 313–344, 2011.

BERLINSKI, S.; GALIANI, S.; MANACORDA, M. Giving children a better start: Preschool attendance and school-age profiles. *Journal of public Economics*, Elsevier, v. 92, n. 5-6, p. 1416–1440, 2008.

BERNAL, R.; FERNÁNDEZ, 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 & Medicine*, Elsevier, v. 97, p. 241–249, 2013.

BETTENDORF, L. J.; JONGEN, E. L.; MULLER, P. Childcare subsidies and labour supply—evidence from a large dutch reform. *Labour Economics*, Elsevier, v. 36, p. 112–123, 2015.

BHUTTA, Z. A. et al. What works? interventions for maternal and child undernutrition and survival. *The lancet*, Elsevier, v. 371, n. 9610, p. 417–440, 2008.

BLACK, S. E. et al. Care or cash? the effect of child care subsidies on student performance. *Review of Economics and Statistics*, MIT Press, v. 96, n. 5, p. 824–837, 2014.

BLAU, D.; CURRIE, J. Who is minding the kids. *The Handbook of Education Economics*, New York: North Holland, 2006.

BRILLI, Y.; BOCA, D. D.; PRONZATO, C. D. Does child care availability play a role in maternal employment and children's development? evidence from italy. *Review of Economics of the Household*, Springer, v. 14, n. 1, p. 27–51, 2016.

BRITO, M. R. F. Adaptação e validação de uma escala de atitudes em relação à matemática). *Zetetiké: Revista de Educação Matemática*, v. 6, n. 9, 1998.

BULMAN, G.; FAIRLIE, R. W. Technology and education: Computers, software, and the internet. In: *Handbook of the Economics of Education*. [S.l.]: Elsevier, 2016. v. 5, p. 239–280.

CAMARGO, J.; LIMA, L.; PORTELA, A. *Vocational Education and Socio-economia Outcomes: Experimental Evidence from Brazil*. [S.l.], 2017.

CARD, D. et al. The labor market impacts of youth training in the dominican republic. *Journal of Labor Economics*, University of Chicago Press Chicago, IL, v. 29, n. 2, p. 267–300, 2011.

CARNEIRO, P. *Human capital policy. Inequality in America: What Role for Human Capital Policy? eds Heckman JJ, Krueger A*. [S.l.]: MIT Press, Cambridge, MA, 2003.

CARRILLO, P. E.; ONOFA, M.; PONCE, J. Information technology and student achievement: Evidence from a randomized experiment in ecuador. 2011.

CORREIA, L.; VAGOS, P.; RIJO, D. A versão específica da escala de medo de avaliação negativa: Dimensionalidade e características psicométricas numa amostra de adolescentes. *Laboratório de Psicologia*, ISPA-Instituto Universitário, v. 12, p. 31–40, 2014.

CUNHA, F. et al. Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, Elsevier, v. 1, p. 697–812, 2006.

CURRIE, J.; ALMOND, D. Human capital development before age five. In: *Handbook of labor economics*. [S.l.]: Elsevier, 2011. v. 4, p. 1315–1486.

DIAMOND, A.; TAYLOR, C. Development of an aspect of executive control: Development of the abilities to remember what i said and to "do as i say, not as i do". *Developmental psychobiology*, Wiley Online Library, v. 29, n. 4, p. 315–334, 1996.

DINI, G. et al. Adaptação cultural e validação da versão brasileira da escala de auto-estima de rosenberg. *Revista Brasileira de Cirurgia Plástica*, Revista Brasileira de Cirurgia Plástica, v. 19, n. 1, p. 41–52, 2001.

DUNN, L. M. et al. *TVIP: Test de vocabulario en imágenes Peabody: adaptación Hispanoamericana*. [S.l.]: American Guidance Service, 1986.

FERNALD, L. C.; GERTLER, P. J.; NEUFELD, L. M. Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of mexico's oportunidades. *The Lancet*, Elsevier, v. 371, n. 9615, p. 828–837, 2008.

GELBER, A.; ISEN, A.; KESSLER, J. B. The effects of youth employment: Evidence from new york city lotteries. *The Quarterly Journal of Economics*, Oxford University Press, v. 131, n. 1, p. 423–460, 2015.

GLEWWE, P.; MURALIDHARAN, K. *Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications*. 1. ed. Elsevier B.V., 2016. v. 5. 653–743 p. ISSN 15740692. ISBN 9780444634597. Available from Internet: <http://dx.doi.org/10.1016/B978-0-444-63459-7.00010-5>.

GREEN, D. P.; WILKE, A.; COOPER, J. Silence begets violence: A mass media experiment to prevent violence against women in rural uganda. *Newhaven, CT: Innovations for Poverty*, 2018.

HAVNES, T.; MOGSTAD, M. No child left behind: Subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy*, v. 3, n. 2, p. 97–129, 2011.

HECKMAN, J. J. Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, American Association for the Advancement of Science, v. 312, n. 5782, p. 1900–1902, 2006. ISSN 0036-8075. Available from Internet: <http://science.sciencemag.org/content/312/5782/1900>.

HECKMAN, J. J. SCHOOLS, SKILLS, AND SYNAPSES. *Economic Inquiry*, v. 46, n. 3, p. 289–324, 2008. Available from Internet: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1465-7295.2008.00163.x>.

HECKMAN, J. J.; KAUTZ, T. *Fostering and measuring skills: Interventions that improve character and cognition*. [S.l.], 2013.

HECKMAN, J. J.; PINTO, R. Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs. *Econometric Reviews*, v. 34, p. 6–31, 2015. ISSN 15324168.

HECKMAN, J. J.; STIXRUD, J.; URZUA, S. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, The University of Chicago Press, v. 24, n. 3, p. 411–482, 2006.

HELLER, S. B. Summer jobs reduce violence among disadvantaged youth. *Science*, American Association for the Advancement of Science, v. 346, n. 6214, p. 1219–1223, 2014.

HERBST, C. M. Universal child care, maternal employment, and children's long-run outcomes: Evidence from the us lanham act of 1940. *Journal of Labor Economics*, University of Chicago Press Chicago, IL, v. 35, n. 2, p. 519–564, 2017.

HODDINOTT, J. et al. Effect of a nutrition intervention during early childhood on economic productivity in guatemalan adults. *The lancet*, Elsevier, v. 371, n. 9610, p. 411–416, 2008.

HUIZEN, T. V.; PLANTENGA, J. Universal child care and children's outcomes: A meta-analysis of evidence from natural experiments. *Discussion Paper Series/Tjalling C. Koopmans Research Institute*, UU USE Tjalling C. Koopmans Research Institute, v. 15, n. 13, 2015.

IBARRARAN, P. et al. Life skills, employability and training for disadvantaged youth: Evidence from a randomized evaluation design. *IZA Journal of Labor & Development*, Springer, v. 3, n. 1, p. 10, 2014.

JOHN, O. P.; DONAHUE, E. M.; KENTLE, R. L. *The big five inventory—versions 4a and 54*. [S.l.]: Berkeley, CA: University of California, Berkeley, Institute of Personality . . . , 1991.

JR, L. R. A.; DREGER, R. M. The effect of attitudes on performance in mathematics. *Journal of Educational Psychology*, American Psychological Association, v. 52, n. 1, p. 19, 1961.

KLEIBERGEN, F. Generalizing weak instrument robust iv statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics*, Elsevier, v. 139, n. 1, p. 181–216, 2007.

KLING, J. R.; LIEBMAN, J. B.; KATZ, L. F. Experimental analysis of neighborhood effects. *Econometrica*, Wiley Online Library, v. 75, n. 1, p. 83–119, 2007.

KNUDSEN, E. I. et al. Economic, neurobiological, and behavioral perspectives on building america's future workforce. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 103, n. 27, p. 10155–10162, 2006.

LAI, F. et al. Does computer-assisted learning improve learning outcomes? evidence from a randomized experiment in migrant schools in beijing. *Economics of Education Review*, Elsevier, v. 47, p. 34–48, 2015.

LEARY, M. R. A brief version of the fear of negative evaluation scale. *Personality and Social Psychology Bulletin*, Sage Publications Sage CA: Thousand Oaks, CA, v. 9, n. 3, p. 371–375, 1983.

LEFEBVRE, P.; MERRIGAN, P. Low-Fee ($5/Day/Child) Regulated Childcare Policy and the Labor Supply of Mothers with Young Children: A Natural Experiment from Canada. *Ssrn*, v. 26, n. 3, p. 519–548, 2005. ISSN 0734-306X.

LEROY, J. L.; GADSDEN, P.; GUIJARRO, M. The Impact of Daycare Programs on Child Health, Nutrition and Development in Developing Countries: A Systematic Review. p. 1–44, 2011. Available from Internet: <http://www.tandfonline.com/doi/abs/10.1080/19439342.2011.639457>.

LIMA, R. A. S. C. Tradução, adaptação e validação do test de vocabulario en imágenes peabody (tvip). Universidade Federal de Pernambuco, 2007.

LINDEN, L. L. *Complement or substitute?: The effect of technology on student achievement in India*. [S.l.]: InfoDev Working Paper, Columbia University, 2008.

LØKEN, K. V. Family income and children's education: Using the norwegian oil boom as a natural experiment. *Labour Economics*, Elsevier, v. 17, n. 1, p. 118–129, 2010.

LUFT, C. D. B. et al. Versão brasileira da escala de estresse percebido: tradução e validação para idosos. *Revista de Saúde Pública*, SciELO Public Health, v. 41, p. 606–615, 2007.

MACOURS, K.; SCHADY, N.; VAKIS, R. Cash transfers, behavioral changes, and cognitive development in early childhood: Evidence from a randomized experiment. *American Economic Journal: Applied Economics*, JSTOR, p. 247–273, 2012.

MALLOY-DINIZ, L. F. et al. Tradução e adaptação cultural da barratt impulsiveness scale (bis-11) para aplicação em adultos brasileiros. *J Bras psiquiatr*, v. 59, n. 2, p. 99–105, 2010.

MANLEY, J.; VÁSQUEZ, F. et al. Childcare availability and female labor force participation: An empirical examination of the chile crece contigo program. *Towson University, Department of Economics*, 2013.

MANNING, M.; PATTERSON, J. Lifetime effects: The high/scope perry preschool study through age 40. *Childhood Education*, Taylor & Francis Ltd., v. 83, n. 2, p. 121, 2006.

MARTÍNEZ, C.; PERTICARÁ, M. Childcare effects on maternal employment: Evidence from chile. *Journal of Development Economics*, Elsevier, v. 126, p. 127–137, 2017.

MCGREW, K. S. The cattell-horn-carroll theory of cognitive abilities: Past, present, and future. Guilford Press, 2005.

MO, D. et al. *The persistence of gains in learning from computer assisted learning: Evidence from a randomized experiment in rural schools in Shaanxi Province*. [S.l.], 2013.

MORRISSEY, T. W. Child care and parent labor force participation: a review of the research literature. Available from Internet: <https://link.springer.com/content/pdf/10.1007{\%}2Fs11150-016-9331-3.>

MURALIDHARAN, K.; SINGH, A.; GANIMIAN, A. J. Disrupting education? experimental evidence on technology-aided instruction in india. *American Economic Review*, v. 109, n. 4, p. 1426–60, April 2019. Available from Internet: <http://www.aeaweb.org/articles?id=10.1257/aer.20171112>.

NOLLENBERGER, N.; RODRÍGUEZ-PLANAS, N. Full-time universal childcare in a context of low maternal employment: Quasi-experimental evidence from spain. *Labour Economics*, Elsevier, v. 36, p. 124–136, 2015.

OREOPOULOS, P.; BROWN, R. S.; LAVECCHIA, A. M. Pathways to Education: An Integrated Approach to Helping At-Risk High School Students. *Journal of Political Economy*, v. 125, n. 4, p. 947–984, 2017. ISSN 0022-3808. Available from Internet: <http://www.journals.uchicago.edu/doi/10.1086/692713>.

OREOPOULOS, P.; PAGE, M.; STEVENS, A. H. The intergenerational effects of worker displacement. *Journal of Labor Economics*, The University of Chicago Press, v. 26, n. 3, p. 455–483, 2008.

PATTON, J. H.; STANFORD, M. S.; BARRATT, E. S. Factor structure of the barratt impulsiveness scale. *Journal of clinical psychology*, Wiley Online Library, v. 51, n. 6, p. 768–774, 1995.

PONITZ, C. E. C. et al. Touch your toes! developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly*, Elsevier, v. 23, n. 2, p. 141–158, 2008.

ROSENBERG, M. Rosenberg self-esteem scale (ses). *Society and the adolescent self-image*, Princeton University Press Princeton, NJ, 1965.

ROTHBART, M. K. et al. Investigations of temperament at three to seven years: The children's behavior questionnaire. *Child development*, Wiley Online Library, v. 72, n. 5, p. 1394–1408, 2001.

SCHOCHET, P. Z.; BURGHARDT, J.; MCCONNELL, S. Does job corps work? impact findings from the national job corps study. *American Economic Review*, v. 98, n. 5, p. 1864–86, 2008.

SCORZA, P. et al. Towards clarity in research on "non-cognitive" skills: Linking executive functions, self-regulation, and economic development to advance life outcomes for children, adolescents and youth globally. *Human development*, NIH Public Access, v. 58, n. 6, p. 313, 2016.

SHEA, J. Does parents' money matter? *Journal of public Economics*, Elsevier, v. 77, n. 2, p. 155–184, 2000.

SHEKAR, M. et al. *Repositioning nutrition as central to development: A strategy for large scale action*. [S.l.]: World Bank Publications, 2006.

STROOP, J. R. Studies of interference in serial verbal reactions. *Journal of experimental psychology*, Psychological Review Company, v. 18, n. 6, p. 643, 1935.

TIERNEY, J. P. et al. Making a difference. an impact study of big brothers/big sisters. ERIC, 1995.

VERA, P. M. Public day care and female labor force participation: Evidence from chile. Universidad de Chile, Facultad de Economía y Negocios, 2009.

WALKER, S. P. et al. Nutritional supplementation, psychosocial stimulation, and growth of stunted children: the jamaican study. *The American journal of clinical nutrition*, Oxford University Press, v. 54, n. 4, p. 642–648, 1991.

WECHSLER, D. *Wechsler intelligence scale for children-WISC-IV*. [S.l.]: Psychological Corporation, 2003.

WOODCOCK, R. W. et al. *Bateria III Woodcock-Munoz: pruebas de habilidades cognitivas*. [S.l.]: Riverside Publishing Company, 2005.

WORLDBANK. World development report 2018 : Learning to realize education's promise. 2018.

YANG, Y. et al. Computers and the academic performance of elementary school-aged girls in china's poor communities. *Computers & Education*, Elsevier, v. 60, n. 1, p. 335–346, 2013.