HCEO WORKING PAPER SERIES

The Effects of a Project and Play-Based Early Education Program on Medium Term Developmental Trajectories of Young Children in a Low-Income Setting

Raquel Bernal
Michele Giannola
Milagros Nores

Working Paper

2022-027

09/2022

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

"The effects of a project and play-based early education program on medium term developmental trajectories of young children in a low-income setting" [§,†]

August 2022

Raquel Bernal, Ph.D. *

Economics Department and Centro de Estudios sobre Desarrollo Económico, Universidad de Los Andes, Bogotá, Colombia

Michele Giannola, Ph.D.

University of Naples Federico II, CSEF, and the Institute for Fiscal Studies

Milagros Nores, Ph.D.

National Institute for Early Education Research, Graduate School of Education, Rutgers, The State University of New Jersey

---

[§] Acknowledgments: This research was supported by the Jacobs Foundation (Grant No. 209-805-1), the UBS Optimus Foundation, the LEGO Foundation and the Inter-American Development Bank (PO 209-805-1). We are very thankful to aeioTU for their commitment to early childhood and opening the doors to our evaluation team; iQuartil for their excellent work managing data collection on site; and our data collectors who worked under very difficult conditions. We also gratefully acknowledge the valuable research assistance provided by Santiago Lacouture and Diana Pérez. Any views expressed are those of the authors and do not necessarily represent those of the funders. The authors confirm contribution to the paper as follows: study conception and design: RB & MN. data collection: RB & MN. Analysis, interpretation of results, and draft manuscript preparation: RB, MG & MN. All authors reviewed the results and approved the final version of the manuscript.

[†] Trail Registry # AEARCTR-0001903 https://www.socialscienceregistry.org/trials/1903

[*] Corresponding author: Raquel Bernal; rbernal@uniandes.edu.co; +57 1.332.4383

.

Study Highlights

- Longitudinal evaluation of a randomized-control trial of a high-quality (project and play-based learning, rich adult-child interactions, intentional integration of multiple domains, intense professional development, among others) program in Colombia.
- Large effects in cognitive development and health, and null effects on socioemotional development.
- Sustained effects are observed in health over time.
- Convergence is observed in cognition by the end of the study for a subgroup of children and various potential mechanisms are discussed.
- Differences by gender, age and household vulnerability are discussed.
- This study points towards the importance of quality in programming, and of attending to nutrition and child development comprehensively.

Abstract

Extensive research has shown comprehensive early interventions can improve the developmental outcomes of disadvantaged children. However, the evidence on the effectiveness of high-quality center-based programs for young children in developing countries is still scarce, where programs are typically of low quality and only short-term impacts have been assessed. This paper reports short and medium-run effects from a high-quality early education intervention characterized by key elements of process quality such as project and play-based learning and rich adult-child interactions, on children younger than four years of age in two communities in northern Colombia. We find strong positive effects on cognitive development and health, and no significant impacts on socioemotional development.

Keywords: early childhood development, early education, poverty, impact evaluation

JEL codes: J13, I10, I20, H43.

I. Introduction

Interest in early childhood care and education (ECCE) investments to improve the development of disadvantaged young children has globally risen in recent decades (Britto et al., 2017; Berlinski & Schady, 2016; Nores & Barnett, 2010; Yousafzai & Aboud, 2014). Poverty compromises the development of hundreds of millions of children in the developing world at great cost to individuals and their countries (Black et al., 2016). Cognitive and language gaps by socio-economic status have been assessed to be present as early as age three in a wide range of countries (Barnett & Lamy, 2013; Duncan & Magnuson, 2013; Fernald, Kariger, Hidrobo, & Gertler, 2012; Heckman, 2008; Rubio-Codina, Attanasio, Meghir, Varela & Grantham-McGregor, 2015). Early intervention has shown its potential to alter children's developmental trajectories (Almond, Curie & Duque, 2018; Berlinski & Schady, 2016; Black & Dewey, 2014; Daelmans et al., 2017; Elango, García, Heckman & Hojman, 2015; Nores & Barnett, 2010). There is unparalleled interest in high-quality and comprehensive programming for young children: early childhood development is now a target (goal 4.2) under the sustainable development goals for 2030 (United Nations, 2016) and was highlighted as a priority in the G20 2018 communiqué (point 14). However, questions remain about how best this might be done.

Over the last two decades, various center-based early childhood programs have been evaluated in Latin America, Asia, and Africa, adding to the extensive literature for North America and Europe.[1] For the most part, these programs had positive impacts on children's emerging skills (Camilli, Vargas,

[1] See Arbour, et. al, 2016; Behrman, Cheng & Todd, 2004; Hojman & López Bóo, 2019; Nores, Bernal & Barnett, 2019 for research on Latina America, and Aboud, 2006; Arora, Bharti, & Sharma, 2007; Armecin et al., 2006; Brinkman et al., 2016; Nair et al., 2017; Noboa-Hidalgo & Urzua, 2012 for research on Asia, and Bago, Ouédraogo, Akakpo, Lompo & Ouédraogo, 2019; Malmberg, Mwaura & Sylva, 2011; Martinez, Naudeau & Pereira, 2017; Woldehanna, 2016; Wolf, 2019 for studies in Africa. See Camilli, et. al, 2010; Duncan, et. al, 2011 for North America and (Drange & Havnes, 2019; Felfe & Lalive, 2018; Melhuish, et al., 2004 for Europe.

Ryan & Barnett, 2010; Del Boca, Martino, Meroni & Piazzalunga, 2019; Duncan et al., 2011; Duncan & Magnuson, 2013; Elango at al., 2015; Nores & Barnett, 2010). Benefits appear to be present across both developed and developing countries, and various social contexts (Nores & Barnett, 2010); the strongest evidence coming from well-known programs in the U.S. following children into their teens and adulthood (Barnett, 2011; Englund, White, Reynolds, Schweinhart & Campbell, 2014; Reynolds, Ou, Mondi & Hayakawa, 2017). This evidence also suggests that even when there is some initial convergence in academic outcomes, long-term effects are still likely to persist across many types of outcomes including health, schooling, behavior (Nores & Prayag, forthcoming).

However, some programs have been found to have neutral or negative effects on children, as is the case of studies in Denmark, Italy, Norway and Quebec, as well as some programs in Latin America (Baker, Gruber & Milligan, 2008; Bernal, Attanasio, Peña & Vera-Hernandez, 2019; Carta & Rizzica, 2018; Fort, Ichino & Zanella, 2017; Gupta & Simonsen, 2010; Havnes & Mogstad, 2015; Kottelenberg & Lehrer, 2017; Noboa-Hidalgo & Urzua, 2012; Rosero & Oosterbeek, 2011). Such neutral or negative effects appear to be influenced by the higher socio-economic status of attending children, low-quality programming and/or the comparable quality of alternative early childhood programs (Cornelissen, Dustmann, Raute & Schönberg, 2018; Hojman & López Bóo, 2019).

Few recent studies have investigated the longitudinal benefits of early childhood education for children within the early years, the formal education years or beyond, with somewhat mixed findings.[2] These studies support the argument of the critical importance of quality of the early childhood programming to produce long-term gains.

[2] See Bastos, Bottan & Cristia, 2017; Berlinski, Galiani & Gertler, 2009; Bach, Koebe & Peter, 2019; Biroli, et al., 2018; Brutti & Montolio, 2019; Drange & Telle, 2017; Felfe, Nollenberger, & Rodríguez-Planas, 2015; Fort, Ichino & Zanella, 2017; Baker, Gruber & Milligan, 2019; Havnes & Mogtad, 2015; Kottelenberg & Lehrer, 2017; Melhuish, et al., 2004; Özler et al., 2016; Woldehanna, 2016; Wolf, 2019

Consequently, the field has moved from a strong focus on access, to an equally strong emphasis on quality (Araujo et al., 2015; López Bóo et al., 2019; Nores, Figueras-Daniel, López & Bernal, 2018). The field has started to focus more strongly on quality, transitions to and the subsequent quality of the formal education experience to understand when convergence does and does not occur (Dietrichson, Kristiansen & Nielsen, 2018; Duncan et al., 2015; Jenkins et al., 2018; McCoy, Gonzalez & Jones, 2019; Schweinhart, 2019; Bassok, Gibbs & Lathman, 2019; Weiland, 2016).

Several international studies have estimated long-term effects taking advantage of expansions of center-based care or admissions thresholds to estimate intention-to-treat effects (Bago et al., 2019; Bastos, Botan & Cristia, 2017; Brinkman et al., 2016; Biroli et al., 2018; Felfe & Lalive, 2018; Gupta & Simonsen, 2010; Hojman & López Bóo, 2019; Martinez, Naudeau & Pereira, 2017; Noboa-Hidalgo & Urzua, 2012, among others). Many of these, albeit not all, have found long-term benefits on children.

Only a subset of studies outside of the U.S. followed a randomized set of children for more than a year. Bernal et al. (2019) followed children transitioned from home-based care to center-based care over an 18-month period and report neutral or negative effects. Malmberg, Mwaura and Sylva (2011) shows positive cognitive effects for preschool children over a 36-month period. Özler et al. (2016) studied a preschool teacher and parent training program finding positive effects 18 months later but not at the 36-month mark. Armecin et al. (2006) found program impacts for children exposed to more than 12 months of integrated programing having followed a large cohort of children for 2 years. We build on this literature to understand children's development within the early childhood years in relation to a high-quality intervention.

In this paper, we investigate the short and medium-term effects of the aeioTU program, a Reggio inspired comprehensive, high-quality early education intervention in Colombia characterized by key elements of process quality such as project and play-based learning, rich adult-child interactions, and intentional integration of multiple domains (i.e. children's cognitive, socio-emotional, and health development). We assess the program's effectiveness through a five-year randomized control trial of disadvantaged children in two communities in northern Colombia from 2010 to 2014. We focus on their early childhood trajectory in the years before they enter the formal education system. We measure health, cognitive and socio-emotional development, as well as parenting, using age-appropriate instruments. The key features of quality embedded in the aeioTU model, and its growth to over 13 thousand children nation-wide via public-private partnerships make of it a noteworthy case study. More so given the financial and human capital constraints in Colombia, which often predominate in low and middle-income countries.

We find strong positive effects on cognitive development, of 0.36 of a standard deviation. We also report positive effects on health (measured using nutritional status) with effect sizes between 0.08 and 0.16 of a standard deviation and find no effects on socioemotional development nor home investments. We find suggestive evidence of differences in treatment effects by gender, age and household vulnerability.

We also find that the treatment effects on child cognitive development are mitigated in the fifth year of the program. We investigate possible mechanisms and conclude that this is likely due to a combination of several factors: an earlier enrollment in the formal education system by children in the control group, the crossover of control children who enter the program over time, and the plausible improvements in the counterfactual in the context of a national early childhood strategy that increased access and quality of alternative chilcare services nation-wide. All these mechanisms provide an opportunity to reflect on the strengths and limitations of long-term experimental designs, such as crossover as a function of intervention length, changes in the counterfactual over time, and the importance of interpreting results with attention to the context (e.g. Angrist & Pischke, 2010; Banerjee, Banerjee & Duflo, 2011; Deaton & Cartwright, 2018; Deaton, 2010; Leamer, 2010).

The results presented in this study contribute to the existing literature in various ways. First, we follow four cohorts of chidren randomized to a treatment and control groups, and document their developmental outcomes over a period of five years. Longitudinal studies of this length are more the exception than the rule in developing countries (Nores & Barnett, 2010; Tanner, Candland & Odden, 2015). Second, the study is embedded within a program that grew to serve over 13 thousand children per year in Colombia, which speaks to its scalability potential (Banerjee, Britto, Daelmans, Goh & Peterson, 2019). Finally, the study speaks to the importance of quality of early education given the relevant factors related to center-based quality (including a strong pedagogical content) that define the aeioTU program.

This paper is organized as follows. We start by briefly describing the early childhood policy context in Colombia, and how the aeioTU program fits in this framework. In Section III we describe the study design, sample, measures, and empirical strategy. Section IV presents the results of the study. Section V discusses the results and mechanisms, and Section VI concludes.

II. Background and intervention

A. Early childhood education in Colombia

Colombia started the decade with a growth rate of close to 7%. However, this has consistently declined over time, reaching 1.7% in 2018. At the same time, inequality has decreased over the last decade or so from a Gini coefficient of 0.56 in 2008 to 0.50 in 2018 (Banco de la República, 2020).

About 65% of the 4.3 million children in Colombia younger than five years of age (3.4 million younger than four) are born to socioeconomically disadvantaged families.[3] This population is the focus of this study.

Colombian children are *eligible* for public early childhood education from six months of age through age five. At this point, they are expected to transition to the formal education system; although enrollment in the grade *Transicion*, the first grade in the formal system, is not universal. Early childhood and family support policies in Colombia aim to ameliorate inequality and have been led by the Colombian Institute of Family Welfare (*Instituto Colombiano de Bienestar Familiar*, ICBF). While not low for the Latin American context, and gradually expanding, public early childhood care and education enrollment for socioeconomically vulnerable children under age five in Colombia remains at 38% (Bernal & Camacho, 2014; World Bank, 2013). Children in low-income households show language and cognition developmental gaps as early as 12 months of age, relative to high-income children, and these gaps increase to about one standard deviation by age five (Bernal, Martínez, and Quintero 2015).

Enrollment in public childcare programs is close to 31% for children younger than four years of age, and 63% for children between four and five (Bernal & Camacho, 2014). About 20% of children nationwide attend home-based care known as *Hogares Comunitarios de Bienestar* (HCB). Despite HCB being originally conceived as a program to promote female labor supply, and the low quality of the service provided, recent evaluations have found positive impacts of HCB on children's health and cognition of about 0.15 of a standard deviation (SD) (Attanasio, Di Maro & Vera-Hernández, 2013; Bernal & Fernández, 2013).

[3] Socioeconomic disadvantage is measured in Colombia using SISBEN scores (a proxy means-indicator based on a household socio-demographic survey). We refer exclusively to this vulnerable population throughout the study.

In this context, the government launched in 2011 "*De Cero a Siempre*" (DCAS), a national early childhood strategy aimed at improving the quality of existing early childhood services as well as increasing enrollment in comprehensive services for 1.2 million children (Bernal & Ramírez, 2018; Comisión Inter-Sectorial para la Primera Infancia, 2013).[4] As a result, center-based care grew from 125,000 children in 2011 to about 380,000 children by 2016. Bernal and Ramírez (2019) report an improvement of 0.35 SD in receptive vocabulary of young children because of the increased access to integrated center-based care, associated with the introduction of DCAS. This study therefore needs to be interpreted in relation to this context of national large-scale investments in ECCE.

B. The aeioTu program

The aeioTU program makes part of this national strategy. The program provided comprehensive early childhood education to about 13,300 low-income children aged zero to five throughout urban Colombia by 2016, operating 28 centers. The program is delivered by an NGO through a public-private partnership with DCAS. During the initial years (*circa* 2010), the government provided a stipend that amounted to USD 1,500 per child per year, and aeioTU supplemented this with an additional 20–30% of own resources.[5,6] Starting in 2014, aeioTU stopped supplementing the government's stipend after careful planning and monitoring of its cost-structure.

The program offers full-day (9 hours per day) educational care 11 months of the year. When this study began in 2010, it was characterized by relatively low child-to-teacher ratios (4:1 for infants, 6:1 for toddlers), high teacher qualifications (32% had a BA and the rest had a vocational degree in

[4] Comprehensive childcare services embed pedagogical content aimed at stimulating cognitive and socio-emotional development and do not simply provide a safe environment for the child while the mother works. Moreover, comprehensive services would concurrently offer nutrition, health, care, and early education.
[5] Using the average COP/USD exchange rate in 2010, at the time this study began.
[6] The additional funding provided by aeioTU was used for teacher training and the nutritional component of the program, which was underfunded by the governmental stipend, and provided nutritional supplementation over the holidays.

early childhood education) and concerted pre- and in-service training (120 hours pre-service, over 130 hours in-service). These are all characteristics recommended for program quality (Friedman-Krauss, et al., 2020). Bernal et al. (2019) report that teacher training and coaching strategies in most comparable public center-based programs were not common and varied significantly across service providers at the time this study began. They also report lower educational levels of teachers, and a child-to-teacher ratio of 25:1 for toddlers in these programs.

Teacher training, curricular supports, classroom materials and quality monitoring have increased since the study began. In their analyses of quality and improvement over time, Nores, et. al (2018) report increasing levels of process quality comparable to other strong programs in the region, as well as high levels of process quality as reported by teachers. In contrast, teacher-child ratios and teacher qualifications have been reduced over time, constrained by public requirements and funding, e.g. the government required service providers to hire the paraprofessionals working in the HCB as some of these were being phased off.

The aeioTU program also provides 70% of children's daily nutritional requirements through breakfast, two snacks and lunch, and provides regular nutritional monitoring.[7,8]

The program is inspired by the Reggio Emilia approach (Malaguzzi, 1993).[9] It features project-and play-based learning, rich adult-child interactions, intentional integration of learning

[7] This daily nutritional requirement is mandatory in all public child-care centers. However, this component was underfunded by the government's stipend in about 20-25%, which was covered with aeioTU's own resources, and the program also provided nutritional supplementation (micronutrients) during holidays. We did not collect any data to monitor how the nutritional component was actually implemented in centers.

[8] An on-site nutritionist periodically monitors children's nutritional status. Children found to be at risk are referred to public health services and the center would adjust the nutritional supplement as recommended by the nutritionist. This was not mandatory for all public child-care services at the beginning of this study but was implemented across all government partners after 2012.

[9] The Reggio Emilia Approach is an education philosophy for pre-school and primary education. It is based on the notion that children are capable of constructing their own learning process through their innate curiosity to understand the world. The basic principle is that children learn about themselves and their context through explorations with

across multiple domains, and a balance of teacher-directed and child-initiated activities. The program has a heavy emphasis on projects and themes of study, which emerge through the children's play and investigations. Daily activities are guided and structured through specific pedagogical guidelines and group planning sessions and incorporate play and art into the learning experience. In line with the educational content, the program staffed centers with a team of professionals including the atelier (in-site artist) and a pedagogical coordinator (director), who plays a critical role in pedagogical planning.[10] In line with the Reggio Emilia approach, aeioTU also emphasizes the active participation of families, holding regular workshops for parents, keeping close communication with families on their children's experiences and progress, and engaging in an open-door policy.[11]

The program characteristics are aligned with critical aspects of high-quality programs highlighted in the literature, such as ample opportunities for play, engaging materials, intentional teacher support of learning during play, and strong curriculum that enhances children's development in all domains.[12] This stands in contrast with the "*De Cero a Siempre*" strategy, which emphasizes curricular freedom and intentionally defines broad national standards. Consequently, comparable public center-based programs were not mandated to use structured curricula or clear pedagogical guidelines for teachers' daily activities (Bernal et al., 2019). In addition, aeioTU included strong pre- and in-service training, and used data on children and classrooms to support quality

---

others and their environment. These explorations belong spontaneously to children's everyday experiences, their play, and their speaking, thinking, and negotiating (Malaguzzi, 1993). Thus, adults are mentors and guides of this process rather than mere caregivers or providers of knowledge, in the sense of providing opportunities for children to explore their own interests. The approach recognizes many ways to understand the world and express thoughts and aims at promoting these communication channels within the educational experience, including art, music, dance, movement, play and exploration.

[10] The hiring of a pedagogical coordinator was implemented across all government partners starting in 2012 as part of the national large scale investments in early care and education.

[11] Including encouraging parents to come into the center to breastfeed their infants.

[12] See Barker et al., 2014; Cavanaugh, et. al, 2017; Jacob & Crowley, 2007; Jensen, et. al, 2019; Pyle, Poliszczuk & Daniels, 2018; Singer, Golinkoff & Hirsh-Pasek, 2006; Yoshikawa et al., 2018.

improvement; features often critical for ECCE quality (Barnett & Boocock, 1998; Bernal, 2015; Bowman, Donovan & Burns, 2001; Yoshikawa, Weiland & Brooks-Gunn, 2016).[13]

III. Methods

A. Study timeline

We conducted a randomized controlled trial with families of young children assigned to a treatment and control groups in two early care centers in a northern-coastal city in Colombia. Figure 1 depicts the study's timeline and shows each yearly assessment. We conducted baseline assessments in late 2010, prior to random assignment and to the beginning of the intervention. We assessed children and families every subsequent year (approximately every 10 to 12 months) until 2014. Data collection was carried out each year around mid-school year, which runs from January through November. We conducted baseline data collection in the first center in July-September 2010 (Y1 henceforth), and the program started in November 2010. We conducted baseline data collection in the second center in October-December 2010, with the program starting in March 2011. We collected the first follow-up (Y2) after 8 months, a second follow-up (Y3) 20-22 months later, a third follow-up (Y4) at 32 months, and the last follow-up (Y5) approximately 41 months since the start of the intervention. We subsequently use the term wave to refer to these.

B. Sampling, randomization and masking

The evaluation was designed as a two-site randomized controlled trial based on an oversubscription model. The evaluation sites were selected from centers being opened around the

[13] We do not have comparative assessments of quality to confirm that aeioTU had better quality than comparable center-based child-care programs targeting the same population. We can only qualify that the inputs (ratio, training, qualifications, among others) often linked to center-based quality as described above were higher, on average.

time the study was planned. Centers had to fulfill two criteria for inclusion: (i) be large enough, so that we would have a sufficient power to detect progamme impacts, and (ii) have enough demand (i.e., oversubscription), so that a lottery could be drawn among applicants. Two centers under construction and opening in 2010 in two disadvantaged communities in northern Colombia were deemed suitable, particularly given the relative absence of alternative early childcare services in these communities when the study begun.[14]

We identified all children under age five residing in these communities through a door-to-door census and the help of community leaders, for a total of 1,288 children. All families were income-eligible for the program by SISBEN scores and expressed interest in enrolling in aeioTU centers if offered a slot. Seventy children were excluded from randomization: 66 children were offered a center slot for reasons including being related to center staff, 2 children moved out of the communities prior to randomization, and 2 children outgrew the required program age (zero to five) between the census and the first day of services. This rendered a final sample of N=1,218 children under age five at baseline. All families consented to participate in the study through active consent.[15]

Each of the two centers in the study had a capacity of 320 children ages 0–5, with just over half of that for children up to age 3. Children were randomly assigned to treatment and control. The randomization was stratified by age groups (which we refer to as cohorts), gender and location (four neighborhoods). We used computer generated random lists to assign children to the treatment and control groups. We did this in a public event, within the community, and with all families present.

The centers then followed-up with lottery winners' parents for effective registration. Some lottery winners opted out of enrollment. With a census on the universe of age-eligible children in

[14] Location choices for the centers also included the political will of local mayors who often supported with funding for the infrastructure, as well as ICBF's approval, which prioritized underserved areas.
[15] Ethics Committees at participating institutions approved the study's protocol in 2009.

these communities, centers resorted to children in the control group to fill the available slots. Enforcement and monitoring over the study's assignment protocol was relaxed over the years due to ethical and programmatic reasons, allowing eligible children to enroll. This increased crossover from the control to the treatment group over the years. Details on compliance and crossover are discussed in section IV.D, while implications for our results are discussed in section V.

Child assessors and parent interviewers were blind to treatment status of participants. Realistically, parents could have communicated their status at post-testing, so it is plausible that assessors learnt this information over time.

C. Study sample

In this paper, we focus on the effects of aeioTu on the development of pre-school age children and leave the study of program impacts on longer run outcomes for future research. We therefore restrict our analysis to children who are *eligible* for early childhood education in a give year. As explained above, a child is age-eligible for early education in a given school year if she is younger than five years of age by March 31st of that year. Since data collection took place in the second half of each school year, we recorded a child as age-eligible for the program in a given year if she was eligible for at least 30% of the time between wave *t*'s data collection and the third week of January of the previous year.[16]

The analytical sample consists of 1,073 children between zero and four years of age at baseline, so that we can observe at least one follow-up in which the child is still age-eligible for the program. Given the longitudinal nature of our data, we progressively exclude older cohorts as time goes by and children lose eligibility for early childhood programming and are expected to move on

[16] As we report below, the main results remain if this threshold is increased to eligibility for up to 50% of the time.

to formal education. Consequently, if we were to pool *eligible* and *non-eligible children*, results for the impact of the program would be mingled with the impact of the early elementary grades on children that have moved onto the formal education system.

Table 1 describes the sample by treatment group and shows how the sample size changes at each yearly assessment (which we hereby describe as waves), as children aged and therefore lost eligibility for early childhood education. A total of 1,073 children aged zero to four at baseline were randomized into a treatment and a control group: 471 were offered slots in the centers (the treatment group) and 602 allocated to the control group. By Y3 4%-7% of the oldest cohort (three to four at baseline) had lost eligibility, and in Y4 39% (control) and 57% (treatment) of children in this cohort lost eligibity. In Y5, all children in the oldest cohort lost eligibility, and 40% (control) and 56% (treatment) of children in the two-to-three cohort would also have been expected to progress to the formal education system based on their age. Children in the two youngest cohorts remained eligible throughout the five years period of this study.

Power analyses indicate a power of 0.85 for sample sizes of 1,073, with α=0.05 and an expected effect size of 0.20 SD, allowing for an attrition rate of 15%.

D. Theory of Change

We hypothesize that exposure to the program would have an impact on children’s health and education outcomes, particularly given the limited supply of early childhood center and home-based services available in the two low-income communities in the study.[17] Specifically, we expected that the components of program quality, and the strong infrastructure and operational supports would translate into enhanced learning opportunities for children, thus improving their language, motor,

[17] At baseline, only 14.4% of children 0 to 4 years of age in the sample had used childcare services during the previous year. The rest of the children (85.6%) were being cared at home by parents, other relatives or non-relatives.

cognitive and socio-emotional development. Moreover, we anticipated that the nutritional component of the program would result in improvements in children's health status. As we discuss later, changes in the counterfactual over time could affect our expected results.

Finally, we predicted that the program could affect the home learning environment and the engagement of parents with their children. However, the direction of these changes is ambiguous a priori. On the one hand, the program's emphasis on active participation and education of families could affect parental knowledge and influence parenting. Similarly, if the marginal return of parental investments increases with improvements in the quality of early education (because complementarities exist between these two in the child human capital production function), parents could invest additional resources in their children. In contrast, it is possible that parents switch resources (including time) to non-treated children (or themselves) as a response to the intervention, to equalize the allocation of resources across their children (or household members). This would mitigate the program's impacts. We therefore also assessed the home environment.

E. Measures and outcomes

The instruments used in this study have adequate psychometric properties and have been used extensively in evaluations of early care and education programs, including studies in developing countries (Fernald, Prado, Kariger & Raikes, 2017). Child assessments were collected by graduates and seniors in psychology trained to reliability standards (100% agreement with the trainer) by experienced staff in a two-week training which included live reliability with young children.[18]

Data collection was conducted in spaces rented and adapted for that purpose every year, under identical conditions for treatment and control children, with parental informed consent. Instruments

[18] Assessores involved in child assessment at any given wave were offered a refresher training every year, and new assessors (if any) were fully trained in similar conditions.

that relied on parent surveys were collected through direct interviews carried out in a separate room alongside the child's assessment. Families and children were provided small incentives for participation and an on-site snack.

Table 2 summarizes the child assessments used. In Appendix A, we present a detailed description of each of these measures. These assessments changed over time to accommodate the fact that children aged over the course of the study, and different assessments are developmentally appropriate at different ages.

Health: We measure health outcomes, using height, weight, and arm circumference. These were collected following World Health Organization (WHO) standards and procedures (WHO, 2006; WHO, 2007).

Cognitive skills: Cognitive development was measured using the cognitive, motor, and language scales from the Bayley Scales of Infant Development III (BSID) for children up to 42 months of age (Bayley, 2005). We administered the Peabody Picture Vocabulary Test in Spanish (TVIP) which measures receptive vocabulary for children over 30 months of age in all waves (Padilla, Lugo & Dunn, 1986). We used the Early Literacy Skills Assessment (ELSA) to measure early literacy in children over 36 months of age in all waves except Y5 (DeBruin-Parecki, 2005).

Emerging math and literacy skills for children older than three were measured using three subtests of the Woodcock-Muñoz III Tests of Achievement (WM-III): subtests #1 (letter-word identification), #9 (text comprehension) and #10 (applied problems) (Muñoz-Sandoval, Woodcock, McGrew and Mather, 2005). The applied problems subtest was used every year, while the literacy subtests were included in Y4. Executive functions were measured using the Head-Toes-Knees and Shoulders (HTKS) which assesses self- regulation in children older than four (Ponitz, McClelland, Jewkes, Connor, Farris & Morrison, 2008; Ponitz, McClelland, Matthews & Morrison, 2009).

Socio-emotional skills: Socio-emotional development was assessed using The Ages and Stages Questionnaire: Socio-Emotional (ASQ:SE), a parent-completed questionnaire for children 6–60 months old measuring self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interactions with others (Squires, Bricker & Twombly, 2009).[19] We also used the Behavior Assessment System for Children, Second Edition (BASC-II) for children older than 36 months. The BASC measures adaptive functioning and behavioral problems (Bracken, Keith, & Walker, 1998; Doyle, Ostrander, Skare, Crosby & August, 1997).

Starting in Y4, we incorporated the Vineland Adaptive Behavior Scales for children aged three and older. The Vineland is a parent questionnaire on personal and social skills, daily living skills, socialization, and motor skills (Sparrow, Balla, Cicchetti & Harrison, 1985).

Home environment: To assess the home environment we surveyed parents on: (1) discipline strategies used at home, (2) nutritional and feeding habits, and (3) parental engagement with children. Discipline strategies are measures with an 8-item scale asking parents to rate how often they use certain types of discipline strategies that range from physical and verbal punishments to positive alternatives, adapted from the Fragile Families Study (Westat, 2011). The scale was collected for all children across all waves. Starting in Y2, nutritional and feeding habits were assessed with questions about meal contents from which we constructed a measure of balanced diets (i.e., all nutritional elements in each meal or each day), and a measure of food insecurity (i.e. the child skipped at least one meal due to lack of resources). Parental engagement with the child was assessed using questions regarding the number of hours devoted by parents to childcare during weekdays and weekends, and

[19] Higher scores on the ASQ:SE represent higher levels of socio-emotional risk or negative behaviors.

the frequency of activities with children such as reading, feeding them, playing with them, and walking with them or visiting places. These questions were collected in all waves.

In addition to the outcome measures described above, we included a household survey answered by mothers or the household head inquiring on educational attainment, maternal age at birth of the child, race, income and expenditures, employment, household assets, health insurance, number of children in the household, and childcare experiences.

Outcome variables: To keep the number of outcome variables contained, thus allowing for greated statistical power, we follow Heckman, Pinto and Savelyev (2013) and implement a factor analytic approach. This approach has the advantages of (a) summarizing the wealth of information contained in the different developmental measures in a lower dimensional and interpretable construct that we define as a *skill* or *developmental domain*, (b) correct for measurement error in the measures, and (c) compare program effects over time although assessments changed throughout the study. We construct three latent factors, one for each domain of child development that we study: cognitive skills, socio-emotional skills, and health.[20] Appendix B provides additional details on the estimation of the factor model.

Because of the longitudinal nature of this study, a key question of interest is how to compare treatment effects over time. As pointed out by Little (2013), when developmental measures change over the course of a study, the scores across different measures should be made comparable to refer to a unique underlying developmental domain changing over time. Ignoring this recommendation, longitudinal studies often compare scores on different measures over time, even if it is not clear how

[20] We also report results using individual tests in our analysis.

to translate score points between different assessments.[21] When age-appropriate measures for younger children are completely replaced with age-appropriate measures for older ones, the possibility to compare program effects over time is limited.

We exploit the fact that a subset of measures for younger and older children are jointly available in one of the waves, to "link" the measures and compare treatment effects over time. We specify and estimate a dedicated measurement system in which each measure is associated with at most one developmental domain (Gorsuch, 1983 and 2003). We then estimate factor scores for each child at each wave using the Bartlett scoring method (Bartlett, 1937).[22]

F. Statistical Strategy

We estimate intention-to-treat (ITT) effects on child development for age-eligible children based on the following ordinary-least-squares (OLS) specification:

$$Y_{i,s,t} = \alpha + \sum_{t=2}^{5} \beta_t (T_i \times W_t) + \sum_{t=2}^{5} \zeta_t W_t + \delta_s + \Gamma' X_i + \varepsilon_{i,s,t} \qquad (1)$$

where $Y_{i,s,t}$ is the outcome of child *i*, in strata *s* at time *t*. W_t are survey wave indicators (*t=2, 3, 4, 5*). $T_i = 1$ if the child was randomly assigned to treatment in the initial lottery and 0 otherwise. δ_s are randomization strata, and X_i is a vector of background characteristics included to improve efficiency and to capture baseline imbalances between the treatment and control group (these are discussed below). Unless otherwise specified, we always control for tester effects in the analysis (we

---

[21] For example, Ramey et al. (2000) compare treatment effects on different cognitive tests over time. Whilst this comparison is valid to study the treatment effect on each test *individually*, it is not clear whether the magnitude of these coefficients is interpretable with respect to a common underlying developmental domain.

[22] Appendix Figure G1 plots the distributions of the three latent factors. In Appendix Table G1, we report the correlations between the latent factors and sociodemographic characteristics at baseline. The distributions are well behaved and correlations show expected signs. Cognition is positively correlated with maternal education, household wealth and the number of books at home. The correlation between cognition and maternal education increases with child's age. Health is correlated only with maternal education. The correlations between socio-emotional skills and household characteristics are weaker, with number of books for children at home being the most relevant variable.

also report results without tester effects as a robustness check). The error term is clustered at the individual level, the unit of randomization. We exclude from the analysis children with scores lower than three standard deviations below the mean of the relevant domain at baseline. We consider this degree of delay to be a proxy for the presence of a disability.[23]

β_t identifies the intention-to-treat program impact at each survey wave (i.e. the yearly assessments carried out). To benchmark the magnitude of the effects, we report the impacts in terms of standard deviation units of the outcome variable of the control group at baseline throughout. Variations of this model inquire into heterogeneous effects by age, gender, baseline development, and household socioe-economic status.

Because of the comprehensive nature of the programme and the longitudinal nature of the evaluation, we report program impacts on a large number of outcomes. To deal with multiple inference, we compute Romano and Wolf (2005) step-down p-values.[24] We report both adjusted and unadjusted p-values in our analysis. The p-values correspond to one-tailed test for the impact of the intervention, reflecting the presumption that the intervention could not harm the children.[25]

Given crossover between treatment and control groups (dicussed in section IV.D), we also report instrumental variable estimates of the impact of effective enrollment in aeioTu centers on child outcomes, using the initial random assignment as an instrument for actual enrollment. Initial assignment to the progamme is a valid instrument as it was random, and significantly explains enrollment. A regression of enrollment on random assignment yields a statistically significant positive coefficients with an F-statistic of about 30 (see Table 7). Information on effective enrollment

[23] This resulted in the exclusion of 26 (2.48%) observations of cognitive skills in Y1, 35 (3.55%) in Y2, 15 (1.89%) in Y3, 13 (1.31%) in Y4 and 9 (0.89%) in Y5, as well as 14 (1.16%) observations of health in Y1 and 1 in Y2.
[24] Romano and Wolf (2005) step-down adjustment for multiple testing were done across waves for a given developmental outcome extracting t-statistics of effect sizes from Stata and using a Stata algorithm based on the Matlab algorithm written by D Wunderli (University of Zurich).
[25] We also report two-sided p-values below.

in aeioTu centers is directly obtained from administrative records. A child is recorded as enrolled if she was registered in center rosters for at least one month during the period between wave *t* and wave *t-1*. Using this variable, we constructed cumulative enrollment at each wave *t* and estimate the effect of cumulative participation *t* years into the program. This is a local average treatment effect that applies to the "compliers", that is, those whose program participation was affected by the lottery (Imbens and Angrist, 1994).

IV. Results

A. Baseline characteristics

Table 3 and 4 describes the sample in terms of socio-demographic characteristics of children and their families and children's development at baseline, and report balance by treatment assignment.[26] At baseline, children were on average 25 months old. They lived in households with an average of 2.6 children under age five, and 27% lived in single-headed households. Their mothers had on average 8.5 years of education, and only 36% had a high school degree. Children were highly nutritionally vulnerable, with average height-for-age one standard deviation below average WHO standards (WHO, 2006; WHO, 2007). About 21.6% of children in the sample were stunted, compared to a national rate of 15.2% in rural areas and 12% in urban areas (Colombian Longitudinal Household Survey, 2010).

To assess the degree of vulnerability of children in our sample in terms of cognitive development, we computed standardized BSID III scores at baseline (these are not reported in Table

[26] When we re-create the same table for the subsample of children re-interviewed at each wave, the main stylized facts are replicated, including potential imbalances at baseline.

4, which reports raw scores).[27] These were 90.4 (SD=13.3) for cognition, 88.9 (SD=13.2) for language, and 93.6 (SD=13.6) for motor development. That is, children in the sample were about 0.7 of a standard deviation below published norms (Feinstein, 2003), and slightly below scores reported on a recent study of low-income children aged 12–24 months in rural areas in Colombia, which reports scores of 92.0 for cognition and 91.6 for language (Attanasio et al., 2018). In contrast, children between 18 and 36 months of age in Bogota (Colombia's capital) from middle income households scored only 0.1 of a standard deviation below the norming sample (Rubio-Codina et al., 2015). This indicates that children in our sample scored significantly below higher income peers in Colombia. Average socio-emotional (ASQ:SE) scores were slightly above the validation sample and quite comparable to children from low socio-economic urban households in the ELCA (Colombian Longitudinal Household Survey, 2013).

There are some statistically significant differences between groups at baseline. Table 3 shows that treatment families have on average a slightly higher number of children under the age of five (2.78 vs 2.63), and children in the treatment group are more likely to have attended childcare the year prior to baseline (14% vs. 11%). There were fewer babies than toddlers and preschoolers in the treatment by design, as class sizes for younger children were smaller.

In terms of baseline development, Table 4 reports that children in the treatment group scored higher in BSID receptive vocabulary and TVIP than children in the control group. As a result, the same difference emerges in latent cognitive skills.[28] On the other hand, parents of children in the treatment group reported a higher prevalence of problematic behaviors in compliance and autonomy than parents

[27] Bayley III composites computed based on published norms provided by test developers. Standardized scores have mean 100 and standard deviation 15. These are reported here only for comparison purposes but are otherwise not used in the analysis.
[28] Given that children in the treatment group are older than children in the control group, we also compute mean differences in developmental outcomes conditional on children's age (see Appendix Table G2). Even after controlling by age, differences in cognitive measures at baseline in favor of the treatment group are still observed.

of children in the control group. Finally, children in the control group exhibited slightly better baseline BMI scores and weight-for-age than children in the treatment group. All analyses control for baseline imbalances between groups.

C. Attrition

Detailed attrition analysis is reported in Appendix C. We have low levels of attrition. Throughout the study period, it ranged from 3% to 10%, remarkable for a five-year longitudinal study (see Appendix Table C1).[29] Appendix Table C2 estimates the probability of not attriting as a function of treatment status and household background characteristics. This shows that: (i) attrition rates do not differ between the treatment and control groups, (ii) slightly wealthier families are less likely to be surveyed in all waves, (iii) beneficiaries of Colombia's conditional cash transfer (CCT) program are less likely to leave the sample. Appendix Table C3 estimates the difference in our main outcome variables at baseline between attriters and non-attriters, revealing non-significant differences between these two groups.

D. Compliance

Records of daily attendance were incomplete or nonexistent across the years. For this reason, we define children as enrolled in an aeioTu center in any given wave if they are registered in center rosters for at least one month between two data collections.[30]

Figure 2 reports enrollment rates by randomization status (similar results by cohort are reported in Appendix Figure D1). Compliance varied significantly throughout the study. In Y2, 72% of lottery winners enrolled in centers, and 81% of lottery losers did not. Enforcement and monitoring over the

[29] Children missing in each wave varied slightly, as some children not found in a given wave could be found in the following one, for example.
[30] To code this variable, we combined information from different administrative systems tracking enrollment during the study period.

study's assignment protocol was relaxed over the years due to ethical and programmatic reasons, allowing eligible children in the control group to enroll. As a result, compliance rates decreased over the years. By Y5, a significant fraction of children in the treatment group moved to preschools in the formal education system, as preschool classrooms in schools expanded during this period in line with the national initiative. At the same time, enrollment of control children in the program increased from 19% in Y2 to 35% in Y5.[31]

These results document an important cross-over from control to treatment that occurred throughout the course of the study, particularly for children enrolled as babies or toddlers. This would attenuate measured treatment effects and has important implications in the interpretation of the results. We use two approaches to attend to this issue. First, we use an intent-to-treat approach and consider initial random assignment to treatment. Second, we use initial random assignment as an instrument for effective program enrollment and estimate an average treatment effect of the program on the compliers (Imbens and Angrist, 1994).

Detailed compliance analysis is reported in Appendix D. In Appendix Table D1 we study the determinants of compliance by regressing compliance with initial random assignment on household characteristics both for the whole sample and separately for the treatment and control groups. We show that: (i) compliance is not related to treatment assignment, and (ii) on average compliers and non compliers are not different in terms of observable characteristics. In Appendix Table D2 we look at determinants of enrollment in programme centers and show that: (i) as expected, children in the treatment group are more likely to enroll in aeioTu compared to control children, and (ii) children with more siblings, children with working mothers, and CCT beneficiaries are more likely to enroll.

[31] This varied by cohort, with enrollment rates being higher for the youngest cohorts and lower for the oldest cohorts (see Appendix Figure D1), consistent with older children gradually moving to the formal education system.

E. Intention-to-treat effects on child development

Figure 3 and Table 5 present our main results. Table 5 shows the intention-to-treat effect by domain of child development, including Romano-Wolf p-values to account for the fact that we presents results over four periods for each developmental domain.[32] The estimates reported correspond to equation (1), and each coefficient corresponds to the impact of aeioTu at each wave *t* for children who are still age-eligible for early childhood education in that particular year. The effects are reported as fractions of a standard deviation (SD) of the outcome in the control group at baseline.

We find a positive and significant impact of the program on child cognitive skills. At the first follow-up, just eight months into the program, cognitive skills are 0.19 SD higher for treated children compared to similar children in the control group. These effects increase to 0.36 SD in Y3 and 0.27 SD in Y4. The effect in Y5 for eligible children is non statistically different from zero, although the coefficient has a negative sign (-0.05 SD). We discuss potential mechanisms for this convergence in section V.D. The effect in Y2, Y3 and Y4 are not statistically different to each other, while the effect in the last year is statistically different from the effect in Y3 at the 10% level.

We also find positive and significant effects on child health throughout the study. The effect size is 0.08 SD in Y2 and stabilizes at around 0.15 SD thereafter. Here the effects over time are never statistically different from each other. Finally, we do not find any significant effects on child socio-emotional development.

We report several robustness checks for these results in Appendix E. Appendix Tables E1 show that the coefficients are robust to the exlusion of tester effects. Appendix Table E2 shows the sensitivity

[32] The p-values we report in Table 5 are one sided p-values. We report both unadjusted and Romano Wolf p-values. These are adjusted for multiple hypotheses testing within each dimension of child development. That is for each domain (cognition, health, and socio-emotional development), we tested for four hypotheses i.e., across four years. The confidence intervals reported in Figure 3 do not take into account multiple hypothesis testing.

of the results to the exclusion of the household characteristics that were unbalanced at baseline. Again, we confirm our main results. To check whether the changes in treatment effects over time for child cognitive skills are due changes in the estimation sample, in Appendix Tables E3 we report the treatment effects on the sample of children who remain age-eligible until the last study year. The estimated coefficients follow the same pattern as in Table 5 and Figure 3. Appendix Tables E4 and E5 report the effects for: (ii) children that are age-eligible for early childhood education for 50% of the time between two data collections (instead of 30% as in our baseline specification), and (iii) the full sample of children regardless of their age-eligibility for early childhood education in each wave. This battery of robustness checks reveals largely similar results. If anything, the estimated impacts on child health are larger if we use the more conservative 50% eligibility threshold (the effect in Y5 is 0.21 SD vs. 0.16 SD in our baseline specification). Finally, in Appendix Table E6 we report the effects on each individual tests included in the construction of the cognitive factor.

F. Heterogeneous effects

In Figures 4 and 5 and Appendix Tables G3-G6 we present heterogeneous treatment effects on child cognition and health along the following four dimensions: child age (younger 0-2 vs older children 2-4), gender, baseline development (splitting the sample at the median level of the outcome), and baseline household wealth (splitting the sample at the median level of household wealth).[33,34]

We report positive effects on child cognitive development for both younger and older children from Y2 to Y4. We also find suggestive evidence that the effects sizes are larger for older children

[33] There are no meaningful heterogeneous effects on socioemotional development. The results are reported in Appendix Figure G2 and Appendix Tables G3-G6.
[34] The wealth index is the principal component resulting from a factor analysis on characteristics of the dwelling and ownership of durable goods.

(although we cannot reject the null hypothesis of equality of coefficents across the two groups, see Appendix Table G3).

Effects on cognitive skills are particularly strong for girls, with effect sizes close to 0.50 SD in Y2 and Y3 and positive throughout the study. Here we reject the null hypothesis that girls' and boys' coefficients are identical in Y2 (p-value = 0.005) but not in later years (Appendix Table G4). Effects for boys are never statistically different from zero although positive and above 0.2 SD in Y3 and Y4. The effect on boys' cognitive skills is negative in Y5 (-0.36, p-value=0.913), while it stays positive for girls (p-value=0.174).

In terms of socio-economic background, we find suggestive evidence that the effects on cognitive development are stronger for children from relatively less disadvantaged households in comparison to children from the most disadvantaged households in the sample (although we can reject that the coefficients are identical only in Y2, see Appendix Table G5). The effects on children from less vulnerable households are close to 0.50 SD in Y3 and Y4. On the other hand, the effects for the most vulnerable children in the sample are small and not statistically different from zero. Finally, we find suggestive evidence of larger effects for children with lower cognitive developmental levels at baseline (although we cannot reject the null hypothesis of equality of coefficents across the two groups, see Appendix Table G6).

Figure 4 and Appendix Tables G3-G6 show similar results for health. Older children benefitted the most from the intervention. The coefficients are consistently larger for older children (with effects sizes above 0.30 SD in Y4 and Y5) than for younger children (we can only reject the equality of the coefficients in Y4, see Appendix Table G3). Interestingly, in terms of gender we observe stronger effects for boys, with effect sizes up to 0.27 SD in Y5. This contrasts with the results on cognitive skills.

We also find suggestive evidence of the fact that the positive effects on health are driven by the most socioeconomically vulnerable children, with effects of 0.26 SD in Y4 and 0.44 SD in Y5. On the other hand, the effects on the health for children from less vulnerable households are small in magnitude and not statistically different from zero (Appendix Table G5). Finally, we find that children with lower baseline health benefitted the most from the intervention in all waves but the last (we reject the null hypothesis of equality of coefficents across the two groups in Y2 and Y3 but not in later waves, see Appendix Table G6). We discuss these results in greater detail in Section V.

G. Effects of program enrollment on child development

Table 6 reports the two stages least squares (2SLS) results, where we instrument cumulative program enrollment with initial random assignment. Appendix Figure G3 plots the distribution of cumulative enrollment by randomization status. As previously discussed, compliance decreased over time, with 47% of control children enrolled for at least one year, and 19% percent enrolled for three or more years throughout the study period. Nonetheless, cumulative enrollment was significantly higher for childred in the treatment group, so that we can use initial random assignment as an exogenous shifter for program participation.

The 2SLS results in Table 6 mirror those presented in Table 5, but the magnitude of the effects is usually larger, as expected. The effect of an additional year of enrollment in an aeioTu center on child cognitive skills is 0.36 SD in Y2, 0.39 SD in Y3, 0.23 SD in Y4 and -0.04 SD in Y5. The effects on health of an additional year are between 0.11 and 0.14 SD every year. We find no effects on child socio-emotional skills.

H. Effects on intermediate outcomes

We report the program's impacts on intermediate outcomes in Table 7. We find no effects on parental time, nutritional investments, or discipline strategies. Importantly, the point estimates are always close to zero and not statistically significant, so that we can rule out meaningful impacts of the program on these intermediate outcomes.

These results imply that the treatment neither crowded-in nor crowded-out parental investments. Beaton and Ghassemi (1982) and Jacoby (2002) discuss the concerns over in-kind transfer programs targeted to children. For example, in terms of nutrition, a problem may arise if parents, as main providers of most of the child's calorie intake, have considerable scope to alter intake when a nutritional program serves their child. Our results suggest that the program's nutritional in-kind transfer "stick" to the child as in Jacoby (2002) since parents did not alter their nutritional investments. Essentially, the results suggest that the positive impacts on cognition and nutrition are due to the intervention and not due to changes induced at home: parents do not change their investment patterns because of their children's participation in a high-quality early education program.

V. Discussion and Mechanisms

A. Health

We find strong positive effects on child health sustained throughout the study, with effect sizes between 0.08 SD in Y2 to 0.16 SD in Y5. These effects are larger for boys and for older children. We do not find changes on nutritional investments at home, which suggests that in-kind transfers targeted to children can "stick" to the child and do not get diluted as a result of within-household redistribution (Jacoby, 2002).

In contrast to our findings, Bernal and Fernández (2013) do not find nutritional gains of a home-based care program with a similar nutritional supplement (50% to 70% of the daily allowance) in a similar context. Andrew et al. (2016) assess the effects of a nutritional enhancement implemented in medium-size childcare centers in Colombia and find an effect of 0.10 SD on weight-for-age, but no statistically significant effects on BMI, weight-for-length, or height. Finally, Bernal et al. (2019) report a small effect of 0.05 SD on the nutritional status of children attending public center-based care in urban Colombia relative to children attending home-based care. The effect sizes in our study are comparable or larger and persist over time.

These effects are important in relation to the literature. A metanalysis by Nores and Barnett (2010) report average effects of 0.31 SD for interventions that include solely a nutritional component, and effect sizes of 0.23 SD for interventions that also include educational component. The findings in this study are therefore sizable even considering a larger body of international work. Particularly so for an intervention that developed its nutritional component based on regular meals, rather than based on supplementation add-ons.

The larger effect we find for older children (cohorts of children aged between 2 and 4 at baseline), could be driven by the fact that these children were significantly more vulnerable at baseline according to their anthropometric indicators and in terms of food fragility compared to younger children (Appendix Table G7). At baseline, stunting was 23.3% in the 2-4 cohort and 18.6% in the 0-2 cohort. Similarly, height-for-age was -1.25 SD for older children and -0.97 SD for younger children. As a result of this vulnerability, older children could have benefitted the most from the nutritional component of the intervention. A second potential explanation for the larger effects on older children might be related with the fact that younger children are less likely than older children to eat meals without an adult's help, as reported by Andrew et al. (2016).[35]

For gender, we find suggestive evidence that the effects on health are larger for boys. To understand this heterogeneity, we estimate differences in nutritional status between boys and girls at baseline. There were no meaningful differences by gender (Appendix Table G8). We also looked at differences on food fragility and nutritional investments by gender (results not reported). The differences are mostly non-statistically significant, but when they are not, they favor girls.[36] Overall, there is no evidence that gender differences were driven by baseline differences or by crowding out of investments. We speculate these results might be influenced by gender differences in children's eating behaviors that have been well documented in the literature. It has been shown that boys exhibit greater desire to drink, more emotional overating, and higher food responsiveness (the urge to eat when the child sees, smells, or tastes food) compared to girls. On the other hand, females show greater avoidance eating behaviors (i.e., slow eating and faster satiety; see Keller et al., 2019 for a review). Without data on eating behaviors, we can only speculate that girls could have eaten less of their meal portions, and that this could explain the larger effects we observe for boys.

B. Cognitive development

We find positive and significant impact of the program on child cognitive skills from Y2 to Y4, with effect sizes as large as 0.36 SD. 2SLS effects are larger and close to 0.40 SD. These effects

[35] Andrew et al. (2016) report that children younger under age 2 in center-based care in in Colombia consume 38% of their lunch portions without adult help, children between 2 and 3 consume 40% on their own, and children older than 3 consume 55%. Given that the centers in this study were comparably larger, it is reasonable to assume these types of differences in consumption would also be present, or at least not fully reduced with no additional staff during mealtimes.

[36] For example, in Y2, food fragility improved for girls by 4.4 percentage points (p-value 0.064) but it was not statistically significant for boys.

are sizable given the gap in cognitive development between high and low SES children in Colombia of close to one standard deviation by age 5 (Bernal, Martínez & Quintero, 2015).

These effect sizes are substantial relative to similar studies in the Colombian context. Bernal and Fernández (2013) find positive effects on cognitive skills after 15 months of exposure to a *Hogares Comunitarios* for children ages three to six, with effect sizes ranging between 0.15 and 0.30 SD. Bernal et al. (2019) report negative effects of -0.10 SD on cognitive development of public center-based care compared to home-based care. Andrew et al. (2019) report effects of 0.15 SD on cognitive development of an intensive teacher training and coaching intervention in small early education centers in urban Colombia. A meta-analysis by Nores and Barnett (2010) reports average effects of early education programs on cognitive development of about 0.25 SD. In the context of these evaluations, the impacts reported in this study are large and appear very early on (just 8 months after the start of the program).

Effects on cognitive skills are particularly strong for girls, with effect sizes close to 0.50 SD in Y2 and Y3 and positive troughout the study. Magnuson et al. (2016) report mixed evidence on gender differences from early education interventions, with effects varying by context, types and the quality of interventions, and developmental domains (also see, Garcia, Heckman & Ziff, 2018, and Muschkin, Ladd, Dodge & Bai, 2018). While we did not observe systematic differences in socioeconomic characteristics between boys and girls at baseline, there is evidence of higher parental-reported interactions with boys in play and reading (Nores, Bernal & Barnett, 2019). To the extent that girls experienced poorer home learning investments, they may have benefitted more from the enhanced learning environments in the program (Garcia, Heckman & Ziff, 2018).

Developmental differences by gender could also contribute to the results. Recent studies reported consistent advantages of girls during the first three years of life in early communicative gestures, early vocabulary growth, and vocabulary size and complexity (Barbu at al., 2015). Nores, Bernal and Barnett (2019) – the first study of this program assessing short term impacts of the program – reported that at baseline, girls outperformed boys in expressive vocabulary, language, and motor skills by as much as 0.16 SD. Such initial developmental differences might have allowed girls to realize higher returns from treatment.

All households in this study exhibit high socio-economic vulnerability at baseline and generalizations should therefore be limited to low-income households. Keeping this in mind, we find suggestive evidence that the effects on cognitive skills are concentrated on less vulnerable households, while the effects on health are stronger on children from more vulnerable households. Within our vulnerable sample, children in lower SES households experienced larger and more adverse risks such as food fragility, lower parental investments, and higher stunting rates. The fact that nutritional effects are concentrated on the most fragile households, while cognitive impacts are stronger for relatively less disadvantaged children suggests the existence of synergies between nutritional status, home environments, and high-quality early childhood experience in the production of cognitive skills.

Our results are consistent with the fact that at very high levels of nutritional and home fragility, children may require even higher investments and a reduction of their nutritional deficits, to be able to fully capture the benefits of early childhood programming. It may be that for some early childhood interventions to produce benefits on child cognitive skills, nutrition and home supports need to be in place, so that programs have multiplicative effects on children. On the other hand, nutritional effects depend critically on calorie intake and less so on parental investments. Since parents did not crowd out the nutritional resources that children received in centers, then it is reasonable that the most socioeconomically vulnerable children would have benefitted the most from the nutritional component of the intervention.

Finally, we find suggestive evidence that children with lower nutritional and cognitive developmental levels at baseline benefitted the most from the intervention. This strongly coincides with other studies in the early childhood literature (see Britto et al., 2017 for a review).

C. Socio-emotional development

We report no significant program impacts on child socio-emotional development. This aligns with previous research in Colombia. Bernal and Fernández (2013) report improved peer interactions of about 0.12 SD, but at the same time increased disruptive behaviors during play of about 0.17 SD for children attending home-based care. Bernal et al. (2019) find no effects of center-based early education compared to home-based care on socio-emotional development. Finally, Andrew et al. (2019) find no effects on socioemotional development of a training and coaching program for child care teachers.

Beyond the Colombian context, this null effect is not uncommon in the early intervention literature, and could be due to lack of sensitivity, biases, or inaccurate measurement of underlying constructs using parental-reported measures (Achenbach, McConaughy & Howell, 1987; Renk & Phares, 2004). A null effect could also be interpreted positively in relation to findings from some studies on center-based intervention in the U.S., which report negative effects on socio-emotional development. Center-based care offers children opportunities to develop social skills while also increasing the likelihood of disruptive behaviors occurring as children compete for limited resources (Belsky et al., 2007; Haskins, 1985; NICHD, 2003). The null effects in this study imply that such negative effects were not present in this setting.

D. Convergence in cognitive development over time

One puzzling finding is the convergence of the effects on child cognitive skills in year five for the children who remained eligible for treatment. To understand this result we explore several possible explanations.

First, we look at enrollment patterns in the formal education system over time by treatment group (Figure 6).[37,38] The figure shows that enrollment in the public education system is 0% in the first wave of the study, and this figure steadingly increases over time, reaching over 70% in 2016. Importantly, it also shows that enrollment rates in the formal system are lower for children assigned to the treatment group compared to children in the control group, and these differences are more pronounced for older children (Appendix Figure G4), suggesting that children in the control group progressed earlier into the formal education system.[39] To test the significance of the difference in Figure 6, we regress enrollment in the formal education system on the treatment dummy interacted with survey waves indicators. The results are reported in Table 8 and confirm the findings from Figure 6. There is only a small difference in enrollment between treatment and control children in Y2 (0.3 percentage points), but this increases over time. By the last study wave, enrollment rate in the formal education system for the treatment group is 9.1 percentage points (22%) lower than in the control group. This is virtually identical when using enrollment data from the government's data system

---

[37] This figure plots average enrollment rates in the formal education system by treatment group over time. The information comes from the combination of parental reports and administrative records from the Colombian Ministry of Education.

[38] As a reference, in this same period net enrollment rates for the first grade in the public education system (known as *Transición*) in the location of our study went from 53% in 2011 to 59% in 2014, which may have contributed to the trends we see in the data. See the data reported by the government, available here: https://www.datos.gov.co/Educaci-n/MEN_ESTADISTICAS_EN_EDUCACION_EN_PREESCOLAR-B-SICA/nudc-7mev/data

[39] Similar results for the whole sample (irrespectively of age-eligibility) are presented in Appendix Figure G5 and follow a similar pattern in terms of differences between treated and control children, despite of the fact that parents seem to over-report enrollment into the formal education system compared with the information collected from the administrative records in SIMAT.

(columns 3 and 4 of Table 8).[40] These trends occurred within a context of enrollments rates in the first grade of formal school at 53% in 2011 and increasing to 59% in 2014.

While the random assignment to the treatment group appears to relate to a later transition into the formal education system, this is not sufficient to explain the convergence in cognitive skills of treatment and control children in Y5. To explain this convergence, children in the formal education system would need to be learning more of the measured cognitive skills due, for example, to being exposed to more advanced learning materials or through their peers (Chen et al., 2020).

We test this hypothesis by comparing the effectiveness of the aeioTu *vis-à-vis* alternative care arrangements: being cared at home, attending alternative childcare (i.e. other pre-school centers different from aeioTu), and being enrolled in the formal education system.[41] Random assignment alone is insufficient to identify these effects, and we therefore use instrumental variables techniques to identify these effects. Following Garcia, Heckman & Ziff (2018), we consider alternative sets of instruments (including randomization status, household size, marital status of the child's mother, whether the families receive the transfers from the national CCT, and whether the father resided in the same household as the child), and alternative specifications for the instruments (in levels, and interacted with random assignment). Additional details of the instrumental variables approach are presented in Appendix F.

Table 9 shows that aeioTu is more productive than alternative childcare services, and than being cared at home. On the other hand, primary school enrollment has a larger effect than aeioTu on

[40] Similar effects for the full sample are reported in Apeendix Table G9.
[41] Information on alternative care arrangements comes from parental reports.

child cognitive development, with effect as large as 0.90 SD. These results are robust to the set of instruments used and across specifications.[42]

As an additional explanation, we consider whether the convergence is due to issues related to compliance and the counterfactual experienced by children in the control group. In Appendix Table G10, we present OLS comparisons between children in the treatment group and subgroups of children in the control group depending on their compliance to the initial random assignment and counterfactual. The first row reports the main ITT effects on cognitive development as a reference as reported in Table 5. The second row compares complier children in the treatment group (those who did enroll in aeioTu) with complier controls (those who did not attent aeioTu but could have enrolled in other centers, cared for at home or in school). The third and fourth rows compare complier children in the treatment group with control children cared for at home, and control children who attended other early education programs different from aeioTu, respectively. Finally, the last row compares complier children in the treatment group with control children not enrolled in primary school. Consistent with the evidence presented above, there is a positive correlation between enrollment in aeioTu and child cognitive skills, relative to all comparison groups and in all study waves. Moreover, the differences are positive, large, and statistically significant also in Y5 when comparing treatment children with control children cared for at home, or with control children not enrolled in primary school.

Finally, and as mentioned earlier, this study was initiated at the same time the national early childhood strategy DCAS was launched. DCAS increased the supply of center-based early education slots by 200% nationwide between 2011 and 2016 (Bernal et al., 2019). DCAS also included

[42] One disadvantage of the instrumental variable approach is that it does not allow us to look at how these effects vary over time (this would be too demanding on the data, as such models would include over 10 endogenous variables). For this reason, we also present correlations of child cognitive development and various educational alternatives over time in Appendix Table G10.

investments in quality of existing services, empashized the integration of health, nutrition, care and early education, and defined requirements for pedagogical components. A recent evaluation of this expansion of integrated early education services between 2011 and 2013 reported positive effects on children's vocabulary (0.35 SD) persisting five years after the implementation of the early childhood national strategy (Bernal & Ramírez, 2019). Without any information on access and quality of alternative services available in the study communities, we can only hypothesize that these trends had an impact on access to and on the quality of the counterfactual programs utilized by children in the control group. The fact that the program exhibited consistent positive impacts throughout the study when treated children are compared to controls cared for at home, provides evidence in favor of this hypothesis.

VI. Conclusions

This paper reports short and medium-term effects of a high-quality early education intervention, characterized by key elements of process quality such as project and play-based learning and rich adult-child interactions, on children younger than four years of age in two communities in northern Colombia. To this aim, we report on a five-year randomized control trial with a sizable sample of disadvantaged children under the age of four (0-4 years of age) from 2010 to 2014.

The results of this RCT, in alignment with Bernal and Ramirez (2018) seem to signal that the expansion of enhanced comprehensive programs with a focus on quality has important effects on early child development. Comprehensive (including nutrition and education) being the key concept. Our findings, much in line with the literature in the field, highlight the importance that attention be paid to both the nutritional and educational components of programs. In addition, the large group of studies that have emerged in Colombia support a necessary focus on quality of programming. These studies, including our own, also suggest there is more to understand in this context about strengthening (and maybe about measuring) children's socioemotional development. Lastly, these studies point to the critical importance of a cohesive evaluation agenda to understand and build quality in ECCE programming. This accumulation of national evidence (as argued more generally by Angrist & Pischke, 2010) is central to further ECCE expansion and further quality investments in the region. This evidence also supports the notion of creating and sustaining systems of continuous quality monitoring and improvement in the region to inform national strategies and priorities.

The differences in outcomes by gender, age and household vulnerability suggest a few things. Programs may have to pay attention at gender differences in ECCE; and for children in extreme vulnerability programs may need to include targeted programmatic components. Attention to measuring group differences in programming and support appear essential for program equity.

Our findings point towards the strength of the program. Understanding this evaluation in the context of the DCAS national strategy, as well as other contemporaneous evaluations of center and home-based care in Colombia is critical to inform public policy and interpret the results. The fact that the evaluation was embedded within a larger context of early childhood policy enhancements, and that it was long-term, exposed the randomization protocol to some of the challenges of RCTs already highlighted by research. That is, the difficulties and ethics of limiting crossover (Deaton, 2010) which we argue is more problematic in long-term experiments; and the plausibility of interactive cofounders (Leamer, 2010) given that the program and its evaluation informed DCAS, and DCAS impacted the counterfactual by increasing access and quality to center-based care nation-wide and over time. In line with the arguments by Deaton and Cartwright (2018), we explicitly analyze the impact of aeioTu *vis-à-vis* alternative care arrangements to understand and interpret treatment effects, as well contextualize the results in relation to larger changes occurring in the Colombian context at the time of this study. Assessing effects in relation to the different experiences within the control group further illuminates the centrality of understanding the counterfactual and the larger context in randomized trials, so as to better inform policy.

References

Aboud, F. E. (2006). “Evaluation of an early childhood preschool program in rural Bangladesh.” *Early Childhood Research Quarterly*, 21(1), 46-60.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). “Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity.” *Psychological bulletin*, 101(2), 213.

Almond, D., Currie, J., & Duque, V. (2018). “Childhood circumstances and adult outcomes: Act II.” *Journal of Economic Literature*, *56*(4), 1360-1446.

Andrew, A., Attanasio, O., Bernal, R., Cardona, L., Krutikova, S., Martínez, D., Medina, C., Peña, X., Rubio-Codina, M., & Vera-Hernández, M. (2016). “Evaluation of centers of infant development: an early years intervention in Colombia.” *Unpublished manuscript, Institute for Fiscal Studies*.

Andrew, A., Attanasio, O., Bernal, R., Cardona, L., Krutikova, S., & Rubio-Codina, M. (2019). “Preschool quality and child development.” *NBER working paper*, 26191.

Angrist, J. D., & Pischke, J. S. (2010). “The credibility revolution in empirical economics: How better research design is taking the con out of econometrics.” *Journal of economic perspectives*, 24(2), 3-30.

Araujo, M. C., Ardanaz, M., Armendáriz, E., Behrman, J. R., Berlinski, S., Cristia, J. P., ... & López Bóo, F. (2015). “The early years: child well-being and the role of public policy.” *IDB Publications* (Books).

Arbour, M., Yoshikawa, H., Willett, J., Weiland, C., Snow, C., Mendive, S., ... & Treviño, E. (2016). “Experimental impacts of a preschool intervention in Chile on children's language outcomes: Moderation by student absenteeism." *Journal of Research on Educational Effectiveness*, 9(1), 117-149.

Armecin, G., Behrman, J., Duazo, P., Ghuman, S., Gultiano, S., King, E., et al. (2006). "Early Childhood Development Through an Integrated Program: Evidence from the Philippines." World Bank Policy Research Working Paper 3922. Washington, D.C., *The World Bank*.

Arora, S., Bharti, S., & Sharma, S. (2007). "Comparative Study of Cognitive Development of ICDS and Non-ICDS Children (3-6 Years)." *Journal of Human Ecology*, 22(3), 201-204.

Attanasio, O., Baker-Henningham, H., Bernal, R., Meghir, C., Pineda, D., & Rubio-Codina, C. (2018). "Early Stimulation and Nutrition: The Impacts of a Scalable Intervention." *NBER Working Paper*, 25059.

Attanasio, O., Di Maro, V., & Vera-Hernández, M. (2013). "Community Nurseries and the Nutritional Status of Poor Children." Evidence from Colombia. *Econ J*, 123(571), 1025-58.

Bach, M., Koebe, J., & Peter, F. (2019). "Long Run Effects of Universal Childcare on Personality Traits." *DIW Berlin Discussion Paper*, 1815.

Bago, J. L., Ouédraogo, M., Akakpo, K., Lompo, M. L., & Ouédraogo, E. (2019). "Early Childhood Education and Child Development: Evidence from Ghana." *Children and Youth Services Review*, 104620.

Baker, M., Gruber, J., & Milligan, K. (2008). "Universal childcare, maternal labor supply, and family well-being." *Journal of political Economy*, 116(4), 709–745

Baker, M., Gruber, J., & Milligan, K. (2019). "The long-run impacts of a universal child care program." *American Economic Journal: Economic Policy*, 11(3), 1-26.

Banco de la República (2019). “Boletín de Indicadores Económicos, June 25/2019.” Retrieved from http://www.banrep.gov.co/sites/default/files/paginas/bie.pdf

Banerjee, A. V., Banerjee, A., & Duflo, E. (2011). “Poor economics: A radical rethinking of the way to fight global poverty.” *Public Affairs*.

Banerjee, A., Britto, P., Daelmans, B., Goh, E., & Peterson, S. (2019). “Reaching the dream of optimal development for every child, everywhere: what do we know about ‘how to’?.“ *Archives of Disease in Childhood*, 104(1).

Barbu, S., Nardy, A., Chevrot, J., Guellai, B., Glas, L., Juhel, J., & Lemasson, A. (2015). “Sex Differences in language Across Early Childhood: Family Socioeconomic Status does not Impact Boys and girls Equally.” *Frontiers in Psychology*, 6.

Barker, J. E., Semenov, A. D., Michaelson, L., Provan, L. S., Snyder, H. R., & Munakata, Y. (2014). “Less-structured time in children’s daily lives predicts self-directed executive functioning.” *Frontiers in Psychology*, 5(June),1–16.

Barnett, W. S. (2011). “Effectiveness of early educational intervention.” *Science*, 333(6045), 975-978.

Barnett, W. S., & Boocock, S. (Eds.) (1998). “Early Care and Education for Children in Poverty: Promises, Programs and Long-Term Results.” pp.11–44. Albany, NY: SUNY Press.

Barnett, W. S., & Lamy, C. E. (2013). Achievement Gaps Start Early. Preeschool Can Help. In P. L. Carter & K. G. Welner, “Closing the Opportunity Gap: What America Must Do to Give Early Child an Even Chance.” OUP USA: *Oxford Scholarship Online*

Barnett, W. S., & Masse, L. (2007). “Comparative Benefit–Cost Analysis of the Abecedarian Program and its Policy Implications.” *Econ of Ed Rev*, 26(1), 113-25.

Bartlett, M. S. (1937). “The statistical conception of mental factors.” *British journal of Psychology*, 28(1), 97–104.

Bassok, D., Gibbs, C. R., & Latham, S. (2019). “Preschool and children's outcomes in elementary school: Have patterns changed nationwide between 1998 and 2010?.” *Child Development*, 90(6), 1875-1897.

Bastos, P., Bottan, N. L., & Cristia, J. (2017). “Access to preprimary education and progression in primary school: Evidence from rural Guatemala.” *Economic Development and Cultural Change*, 65(3), 521-547.

Bayley, N. (2005). “Bayley Scales of Infant Development.” San Antonio, TX: The Psychological Corporation, Harcourt Brace & Company.

Beaton, G., & Ghassemi, H. (1982). “Supplementary feeding programs for young children in developing countries.” *American Journal of Clinical Nutrition*, 35, 1177-98.

Behrman, J., Cheng, Y., & Todd, P. (2004). “Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach.” *Review of Economics and Statistics,* 86(1), 108-132.

Belsky, J., Burchinal, M., McCartney, K., Vandell, D., Clarke-Stewart, K., & Owen, M., for the NICHD Early Child Care Research Network. (2007). “Are there long-term effects of early child care?.” *Child Development,* 78(2), 681-701.

Berlinski, S., & Schady, N. (Eds.) (2016). “The Early Years: Child Well-Being and the Role of Public Policy.” New York: Springer.

Berlinski, S., Galiani, S., & Gertler, P. (2009). “The effect of pre-primary education on primary school performance.” *Journal of Public Economics,* 93(1-2), 219-234.

Bernal, R. (2015). “The impact of a vocational education program for childcare providers on children’s well-being.” *Economics of Education Review*, 48, 165–183.

Bernal, R., & Camacho, A. (2014). “Early Childhood Policy in the Context of Equity and Social Mobility in Colombia.” In A. Montenegro and M. Meléndez, (Eds.) Equidad y Movilidad Social: Diagnósticos y Propuestas para la Transformación de la Sociedad Colombiana. Editorial Uniandes.

Bernal, R., & Fernández, C. (2013). “Subsidized Childcare and Child Development in Colombia: Effects of Hogares Comunitarios de Bienestar as a Function of Timing and Length of Exposure.” *Soc Sci Med*, 97(November), 241–49.

Bernal, R., & Ramírez, C. (2018). “Improving child care quality at scale: the effects of From Zero to Forever.” *CEDE Working Paper*, 40.

Bernal, R., Attanasio, O., Peña, X., & Vera-Hernández, M. (2019). “The Effects of the Transition from Home-Based Childcare to Center-Based Childcare in Colombia.” *Early Childhood Research Quarterly*, 47, 418-431.

Bernal, R., Martínez, M. A., & Quintero, C. (2015). “Situación de Niñas y Niños Colombianos Menores de Cinco Años entre 2010 y 2013.” Bogotá, Colombia: Editorial Kimpress.

Biroli, P., Del Boca, D., Heckman, J. J., Heckman, L. P., Koh, Y. K., Kuperman, S., ... & Ziff, A. L. (2018). “Evaluation of the Reggio approach to early education.” *Research in Economics*, 72(1), 1-32.

Black, M. M., & Dewey, K. G. (2014). “Promoting equity through integrated early child development and nutrition interventions.” *Annals of the New York Academy of Sciences*, 1308(1), 1-10.

Black, M. M., Walker, S., Fernald, L., Andersen, C., DiGirolamo, A., Lu, C., McCoy, D., Fink, G., Shawar, Y., Shiffman, J., Devercelli, A., Wodon, Q., Vargas-Barón, E., & Grantham-McGregor, S. (2016). "Early Childhood Development Coming of Age: Science through the Life Course." *The Lancet*, 389(10064), 77-90.

Bowman, B., Donovan, M., & Burns, M. (2001). "Eager to Learn: Educating our Preschoolers." Washington, DC: *National Academy Press*.

Bracken, B. A., Keith, L. K., & Walker, K. C. (1998). "Assessment of preschool behavior and social-emotional functioning: A review of thirteen third-party instruments." *Journal of Psychoeducational Assessment*, 16(2), 153-169.

Brinkman, S. A., Hasan, A., Jung, H., Kinnell, A., Nakajima, N., & Pradhan, M. (2016). "The role of preschool quality in promoting child development: evidence from rural Indonesia." Washington, D.C., *The World Bank*.

Britto, P. R., Lye, S., Proulx, K., Yousafzai, A., Matthews, S., Vaivada, T., … & the Early Childhood Development Interventions Review Group (2017). "Nurturing care: promoting early childhood development." *The Lancet*, 389, 91-102.

Brutti, Z., & Montolio, D. (2019). "Preventing criminal minds: early education access and adult offending behavior." *IEB Working Paper*, 2019/02.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). "Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development." *Teach Coll Rec*, 112(3), 579-620.

Carta, F., & Rizzica, L. (2018). "Early kindergarten, maternal labor supply and children's outcomes: evidence from Italy." *Journal of Public Economics*, 158, 79-102.

Cavanaugh, D. M., Clemence, K. J., Teale, M. M., Rule, A. C., & Montgomery, S. E. (2017). “Kindergarten scores, storytelling, executive function, and motivation improved through literacy rich guided play.” *Early Childhood Education Journal*, 45, 1–13.

Chen, J., Justice, L. M., Tambyraja, S. R., & Sawyer, B. (2020). Exploring the mechanism through which peer effects operate in preschool classrooms to influence language growth. *Early Childhood Research Quarterly*, 53, 1-10.

Comisión Inter-Sectorial para la Primera Infancia CIPI. (2013). “De Cero a Siempre: Atención Integral a la primera infancia. Estrategia de Atención Integral a la Primera Infancia. Fundamentos políticos, técnicos y de gestión.” Colombia: Presidencia de la República.

Cornelissen, T., Dustmann, C., Raute, A., & Schönberg, U. (2018). “Who benefits from universal child care? Estimating marginal returns to early child care attendance.” *Journal of Political Economy*, 126(6), 2356-2409.

Daelmans, B., Darmstadt, G. L., Lombardi, J., Black, M. M., Britto, P. R., Lye, S., ... & Richter, L. M. (2017). “Early childhood development: the foundation of sustainable development.” *The Lancet*, 389(10064), 9-11.

Deaton, A. (2010). “Instruments, randomization, and learning about development.” *Journal of economic literature*, 48(2), 424-55.

Deaton, A., & Cartwright, N. (2018). “Understanding and misunderstanding randomized controlled trials.” *Social Science & Medicine*, 210, 2-21.

DeBruin-Parecki, A. (2005). “Early Literacy Skills Assessment user guide.” *Ypsilanti*, MI: *High/Scope Press*.

Del Boca, D., Martino, E. M., Meroni, E. C., & Piazzalunga, D. (2019). "Early education and gender differences." *IZA Discussion Paper,* 12484.

Dietrichson, J., Kristiansen, I. L., & Nielsen, B. C. (2018). "Universal preschool programs and long-term child outcomes: A systematic review." (No. 2018: 19). *Working Paper*.

Doyle, A., Ostrander, R., Skare, S., Crosby, R. D., & August, G. J. (1997). "Convergent and criterion-related validity of the behavior assessment system for children-parent rating scale." *Journal of Clinical Child Psychology*, 26(3), 276-284.

Drange, N., & Havnes, T. (2019). "Early Childcare and Cognitive Development: Evidence from an Assignment Lottery." *Journal of Labor Economics*, 37(2), 581-620.

Drange, N., & Telle, K. (2017). "Preschool and school performance of children from immigrant families." *Empirical Economics*, 52(2), 825-867

Duncan, G. J., & Magnuson, K. (2013). "Investing in preschool programs." *Journal of Economic Perspectives*, 27(2), 109-32.

Duncan, G. J., Jenkins, J. M., Watts, T. W., Magnuson, K., Clements, D., Sarama, J., ... & Spitler, M. E. (2015). "Preventing Preschool Fadeout through Instructional Intervention in Kindergarten and First Grade." *Society for Research on Educational Effectiveness*.

Duncan, G. J., Leak, J. A., Li, W., Magnuson, K., Schindler, H., & Yoshikawa, H. (2011). "Timing Issues with Early Childhood Education Programs: How Effect Sizes Vary by Starting Age, Program Duration and Persistence of Effects." *Society for Research on Educational Effectiveness*.

Edwards, C. P., Gandini, L., & Forman, G. E. (Eds.) (1998). "The hundred languages of children: The Reggio Emilia approach--advanced reflections." Greenwood Publishing Group.

Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2015). "Early childhood education. In Economics of Means-Tested Transfer Programs in the United States." Volume 2 (pp. 235-297). University of Chicago Press.

Encuesta Longitudinal Colombiana. (2013). "Encuesta Longitudinal Colombiana (ELCA)." Bogotá, Colombia: Universidad de los Andes. Available at https://encuestalongitudinal.uniandes.edu.co/en/.

Englund, M. M., White, B., Reynolds, A. J., Schweinhart, L., & Campbell, F. A. (2014). "Health Outcomes of the Abecedarian, Child-Parent Center, and HighScope Perry Preschool Programs." *Health and education in early childhood: Predictors, interventions and policies*, 257, 285.

Feinstein, L. (2003). "Inequality in the Early Cognitive Development of British Children in the 1970 Cohort." *Economica*, 70(277),73-97.

Felfe, C., & Lalive, R. (2018). "Does early child care affect children's development?." *Journal of Public Economics*, 159, 33-53.

Felfe, C., Nollenberger, N., & Rodríguez-Planas, N. (2015). "Can't buy mommy's love? Universal childcare and children's long-term cognitive development." *Journal of population economics*, 28(2), 393-422.

Fernald, L., Kariger, P., Hidrobo, M., & Gertler, P., (2012). “Socioeconomic gradients in child development in very young children: evidence from India, Indonedia, Peru, and Senegal.” *Proceedings of the National Academy of Science*, 109(2), 17273-80.

Fernald, L., Prado, E., Kariger, P., & Raikes, A. (2017). “A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries.” Washington, D.C., *The World Bank*.

Fort, M., Ichino, A., & Zanella, G. (2017). “The cognitive cost of daycare 0-2 for children in advantaged families.” In Technical Report. *Working Paper*.

Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G. & Gardiner, B. A. (2020). “The State of Preschool 2019: State Preschool Yearbook.” New Brunswick, NJ: *National Institute for Early Education Research*.

García, J. L., Heckman, J. J., & Ziff, A. L. (2018). “Gender differences in the benefits of an influential early childhood program.” *European economic review*, 109, 9-22.

García, J., Heckman, J. J., Leaf, D., & Prados, M. J. (2016). “The Life-cycle Benefits of an Influential Early Childhood Program.” *NBER Working Paper*, 22993.

Gorsuch, R. (1983). “Factor Analysis.” *Lawrence Erlbaum Associates, Hillsdale*, NJ.

Gorsuch, R. (2003). “Factor analysis.” In *Handbook of psychology: Research methods in psychology*. Volume 2, 143–164. John Wiely & Sons, Inc., Hoboken, NJ.

Gupta, N. D., & Simonsen, M. (2010). “Non-cognitive child outcomes and universal high quality child care.” *Journal of Public Economics*, 94(1-2), 30-43.

Haskins, R. (1985). “Public school aggression among children with varying day-care experience.” *Child Development*, 56, 689-703.

Havnes, T., & Mogstad, M. (2015). "Is universal child care leveling the playing field?." *Journal of public economics*, 127, 100-114.

Heckman J. J., Pinto R. & Savelyev P. (2013). "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review*, 103(6), 2052-2086.

Heckman, J. J. (2008). "Schools, skills and synapses." *Economic Inquiry*, 46(3), 289-324.

Hojman, A., & López Bóo, F. (2019). "The Impacts of Public Daycare in a Developing Country." *Unpublished manuscript, LACEA-LAMES.*

Imbens, Guido W., and Joshua D. Angrist. "Identification and Estimation of Local Average Treatment Effects." Econometrica 62.2 (1994): 467-475.

Jacobs, G., & Crowley, K. (2007). "Play, projects, and preschool standards: Nurturing children's sense of wonder and joy in learning." Corwin Press.

Jacoby, H. (2002). "Is there an intrahousehold 'flypaper effect'? evidence from a school feeding programme." *The Ecnomic Journal*, 112, 196-221.

Jenkins, J. M., Watts, T. W., Magnuson, K., Gershoff, E. T., Clements, D. H., Sarama, J., & Duncan, G. J. (2018). "Do High-Quality Kindergarten and First-Grade Classrooms Mitigate Preschool Fadeout?." *Journal of research on educational effectiveness*, 11(3), 339-374.

Jensen, H., Pyle, A., Zosh, J. M., Ebrahim, H. B., Zaragoza Scherman, A., Reunamo, J., & Hamre, B. K. (2019). "Play facilitation: the science behind the art of engaging young children" White paper, Denmark: *The LEGO Foundation*.

https://www.legofoundation.com/media/1681/play-facilitation_the-science-behind-the-art-ofengaging-young-children.pdf

Keller, K., Kling, S., Fuchs, B., Pearce, A., Reigh, N., Masterson, T. & Hickok, K. (2019). "A biopsychosocial model of sex differences in children's eating behaviors." *Nutrients*, 11, 1-21.

Kottelenberg, M. J., & Lehrer, S. F. (2017). "Targeted or Universal Coverage? Assessing Heterogeneity in the Effects of Universal Child Care." *Journal of Labor Economics*, 35(3), 609-653.

Leamer, E. E. (2010). "Tantalus on the Road to Asymptopia." *Journal of Economic Perspectives*, 24(2), 31-46.

Little, T. D. (2013). "Longitudinal structural equation modeling." New York, NY: *Guilford*.

López Bóo, F., Dormal, M., & Weber, A. (2019). "Validity of four measures of childcare quality in a national sample of centers in Ecuador." *PloS one*, 14(2).

Magnuson, K., Kelchen, R., Duncan, G., Schindler, H., Shager, H., & Yoshikawa, H. (2016). "Do the Effects of Early Childhood Education Programs Differ by Gender? A Meta-Analysis." *Early Childhood Research Quarterly*, 36(3), 521-536.

Malmberg, L. E., Mwaura, P., & Sylva, K. (2011). "Effects of a preschool intervention on cognitive development among East-African preschool children: A flexibly time-coded growth model." *Early Childhood Research Quarterly*, 26(1), 124-133.

Martinez, S., Naudeau, S., & Pereira, V. (2017). "Preschool and child development under extreme poverty: evidence from a randomized experiment in rural mozambique." Washington, D.C., *The World Bank*.

McCoy, D. C., Gonzalez, K., & Jones, S. (2019). "Preschool Self-Regulation and Preacademic Skills as Mediators of the Long-Term Impacts of an Early Intervention." *Child development*, 90(5), 1544-1558.

Melhuish, E., Quinn, L., Hanna, K., Sylva, K., Sammons, P., Siraj-Blatchford, I., et al. (2004). "Effective Pre-School Provision in Northern Ireland [EPPNI] Project: Summary Report 1998-2004." *Department of Education, Northern Ireland Statistics and Research Agency*.

Muennig, P., Robertson, D., Johnson, G., Campbell, F., Pungello, E., & Neidell, M. (2011). "The Effect of an Early Education Program on Adult Health: the Carolina Abecedarian Project Randomized Controlled Trial." *Am J Public Health*, 101(3), 512-16.

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). "Batería III Pruebas de aprovechamiento." Itasca, IL: Riverside Publishing.

Muschkin, C., Ladd, H., Dodge, K., & Bai, Y. (2018). "Gender Differences in the Impact of North Carolina's Early Care and Education Iniatives on Student Outcomes in Elementary School." *Educational Policy*, https://doi.org/10.1177/0895904818773901

Nair, D., Alonge, O., Derakhshani Hamadani, J., Sharmin Salam, S., Islam, I., & Hyder, A. (2017). "Developmental assessments during injury research: is enrollment of very young children in crèches associated with better scores?." *International journal of environmental research and public health*, 14(10), 1130.

NICHD Early Child Care Research Network (2003). "Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten?." *Child Development,* 74, 976-1005.

Noboa-Hidalgo, G. E. & Urzua, S. (2012). “The effects of participation in public child care centers: Evidence from chile.” *Journal of Human Capital*, 6(1), 1–34

Nores, M., & Barnett. W.S. (2010). “Benefits of Early Childhood Interventions Across the World: (Under) Investing in the Very Young.” *Economics of Education Review*, 29(2), 271-82.

Nores, M., Bernal, R., & Barnett, W. S. (2019). “Center-based care for infants and toddlers: The aeioTU randomized trial.” *Economics of Education Review*, 72, 30-43.

Nores, M., Figueras-Daniel, A., López, M. A., & Bernal, R. (2018). “Implementing aeioTU: Quality improvement alongside an efficacy study. Learning while growing.” *Annals of the New York Academy of Sciences*, 1419(1), 201-217.

Nores, M & Prayag, R. (Forthcoming) Early Childhood Education Effects into Adulthood. In Berends, M., Schneider, B., & Lamb, S. (eds). *SAGE Handbook on the Sociology of Education*. Sage.

Özler, B., Fernald, L. C., Kariger, P., McConnell, C., Neuman, M., & Fraga, E. (2016). “Combining preschool teacher training with parenting education: A cluster-randomized controlled trial.” Washington, D.C., *The World Bank*.

Padilla, E., Lugo, D., & Dunn, L. (1986). “Test de Vocabulario en Imágenes Peabody (TVIP).” Circle Pines, MN: American Guidance Service (AGS), Inc.

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). “A structured observation of behavioral self-regulation and int contribution to kindergarten outcomes.” *Developmental Phychology*, 45(3), 605-619.

Ponitz, C. E., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). "Touch your toes! Developing a direct measure of behavioral regulation in early childhood." *Early Childhood Research Quarterly*, 23(2), 141-15.

Powell, D. R., & Diamond, K. E. (2012). Promoting Early Literacy and Language Development. In R. C. Pianta, W. S. Barnett, L. M. Justice, & S. M. Sheridan (Eds.), Handbook of Early Childhood Education (pp. 194-216). New York: The Guilford Press.

Pyle, A., Poliszczuk, D., & Danniels, E. (2018). "The Challenges of Promoting Literacy Integration within a Play-Based Learning Kindergarten Program: Teacher Perspectives and Implementation." *Journal of Research in Childhood Education*, 00(00), 1–15.

Ramey, C. T., & Campbell, F. A. (1991). Poverty, early childhood education, and academic competence: The Abecesarian experiment. In A. Huston (Ed.), Children reared in poverty. 190-221. New York, NY: *Cambridge University Press*.

Ramey, C. T., Campbell, F. A., Burchinal, M., Skinner, M. L., Gardner, D. M., & Ramey, S. L. (2000). "Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers." *Applied Developmental Science*, 4(1), 2-14.

Renk, K., & Phares, V. (2004). "Cross-informant ratings of social competence in children and adolescents." *Clinical Psychology Review*, 24(2), 239-254.

Reynolds, A. J., Ou, S. R., Mondi, C. F., & Hayakawa, M. (2017). "Processes of early childhood interventions to adult well-being." *Child development*, 88(2), 378-387.

Romano, J., & Wolf, M. (2005). "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73(4), 1237-82.

Rosero, J., & Oosterbeek, H. (2011). "Trade-offs between different early childhood interventions: Evidence from Ecuador." *Tinberg Institute Discussion Paper*, TI 2011-102/3.

Rubio-Codina, M., Attanasio, O., Meghir, C., Varela, N., & Grantham-McGregor, S. (2015). "The socio-economic gradient of child development: cross-sectional evidence from children 6-42 months in Bogota." *Journal of Human Resources*, Spring, 50(2), 464-483.

Schweinhart, L. J. (2019). "11 Lessons on Sustaining Early Gains from the Life-Course Study of Perry Preschool." *Sustaining Early Childhood Learning Gains: Program, School, and Family Influences*, 254.

Singer, D., Golinkoff, R. M., & Hirsh-Pasek, K. (2006). "Play= Learning: How play motivates and enhances children's cognitive and social-emotional growth." Oxford University Press.

Sparrow, S. S., Balla, D. A., Cicchetti, D. V., & Harrison, P. L. (1985). "Vineland adaptative behavior scales: classroom edition." *American Guidance Service*.

Squires, J., Bricker, D., and Twombly, E. (2009). "Technical Report on ASQ:SE." Baltimore, Co: Paul H. Brookes Publishing.

Tanner, J., Candland, T. & Odden, W. (2015). "Later impacts of early childhood interventions: a systematic review." *IEG Working Paper*, 2015/3. *World Bank Group*.

United Nations. (2016). "Sustainable Development Goals. Goal 4: Ensure Inclusive and Quality Education for All and Promote Lifelong Learning." Washington, DC: *United Nations*. Retrieved from http://www.un.org/sustainabledevelopment/education/

Weiland, C. (2016). “Launching Preschool 2.0: A road map to high-quality public programs at scale.” *Behavioral Science & Policy*, 2(1), 37-46.

Westat, Inc. (2011). “The Fragile Families and Child Wellbeing Study (Survey of Parents), Father’s Nine-Year Follow-up Survey Public Use Version.” K10 page 15. Retrieved from https://fragilefamilies.princeton.edu/sites/fragilefamilies/files/ff_dad_q9.pdf

Woldehanna, T. (2016). “Inequality, preschool education and cognitive development in Ethiopia: Implication for public investment in pre-primary education.” *International Journal of Behavioral Development*, 40(6), 509-516

Wolf, S. (2019). “Year 3 Follow-up of the ‘Quality Preschool for Ghana’ Interventions on Child Development.” *Accepted in Developmental Psychology*.

Word Health Organization. (2007). “WHO Child Growth Standards: Head Circumference-for-age, Arm Circumference-for-age, Triceps Skinfold-for-age and Subscapular Skinfold-for-age: Methods and Development.” WHO Multicentre Growth Reference Study Group. Geneva: *World Health Organization*.

World Bank. (2013). “Colombia. Early Childhood Development.” SABER Country Report, Washington, D.C., *The World Bank*. Retrieved from http://hdl.handle.net/10986/16280.

World Health Organization. (2006). “WHO Child Growth Standards: Length/height-for-age, Weight-for-age, Weight-for-length, Weight-for-height and Body Mass Index-for-age: Methods and Development.” WHO Multicentre Growth Reference Study Group. Geneva: *World Health Organization*.

Yoshikawa, H., Leyva, D., Snow, C., Treviño, E., Barata, M., Weiland, C., … & Arbour, M. C. (2015). "Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes." *Dev Psychol*, 51(3), 309–322.

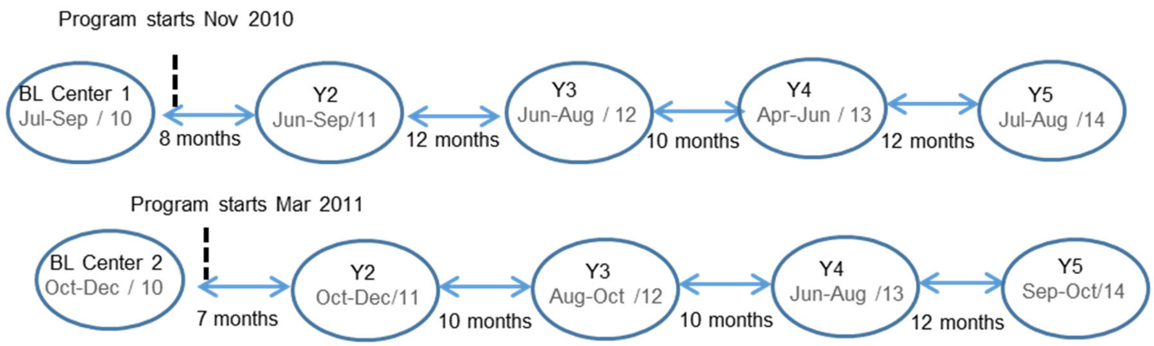
Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). "When does preschool matter?." *The Future of Children, 26*(2), 21-36.

Yoshikawa, H., Wuermli, A. J., Raikes, A., Kim, S., & Kabay, S. B. (2018). "Toward High-Quality Early Childhood Development Programs and Policies at National Scale: Directions for Research in Global Contexts." *Social Policy Report*, 31(1), 1-36.

Yousafzai, A. K., & Aboud, F. (2014). "Review of implementation processes for integrated nutrition and psychosocial stimulation interventions." *Annals of the New York Academy of Sciences*, 1308(1), 33-45.

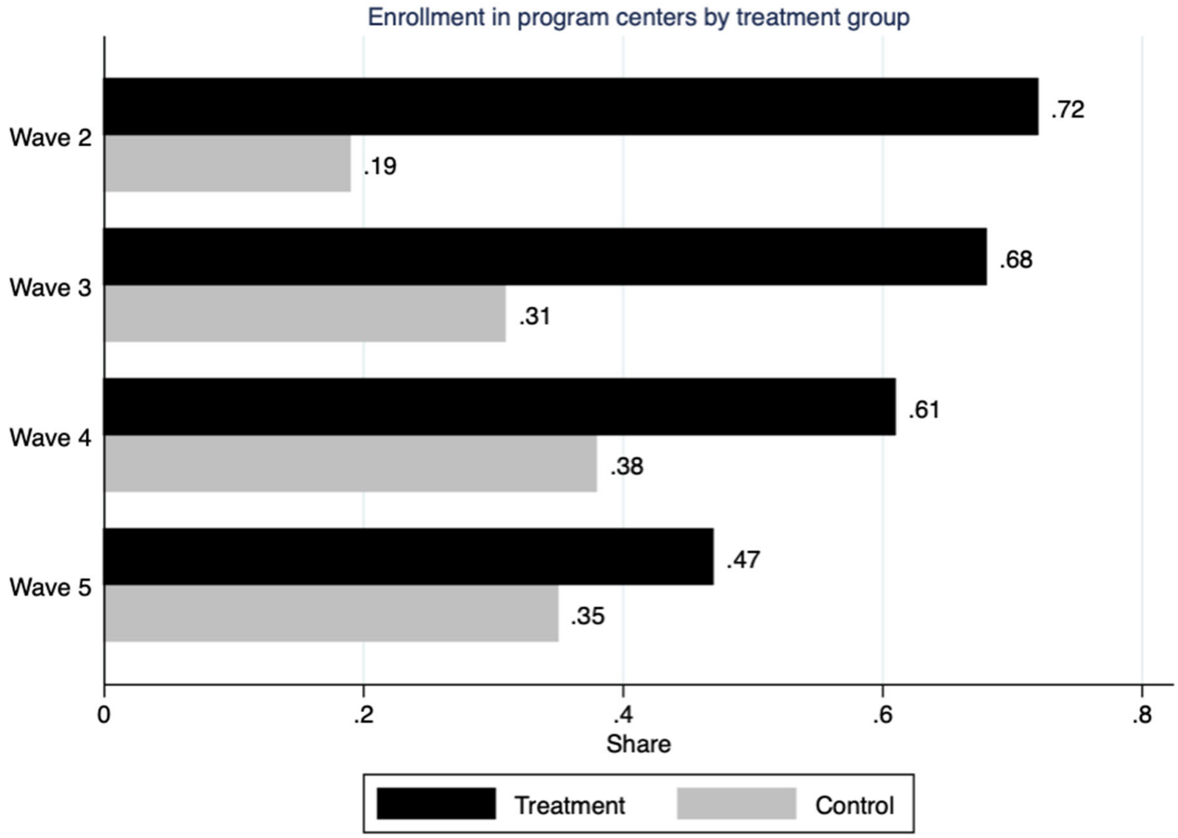
Tables & Figures

Figure 1. Study timeline



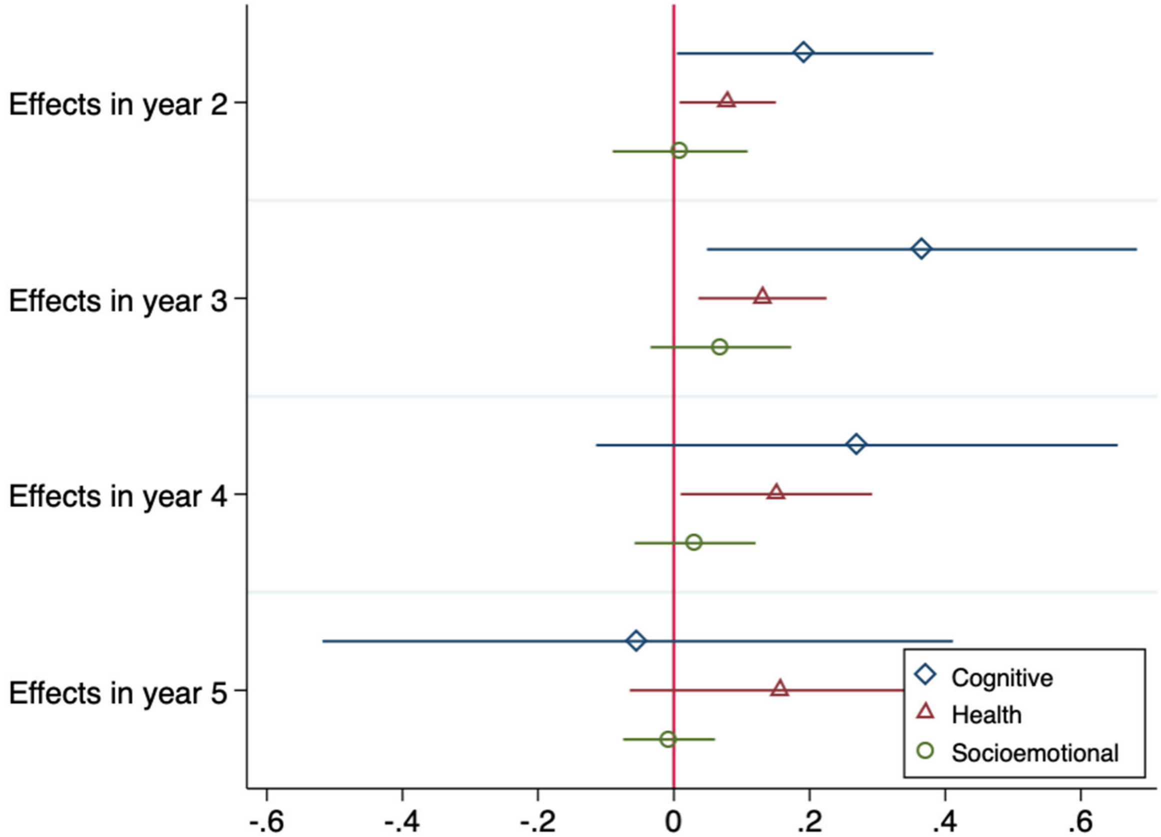
Notes: This figure plots the study timeline. Each yearly assessment is shown and we refer to these as waves throughout the paper.

Figure 2: Enrollment by study wave and treatment group



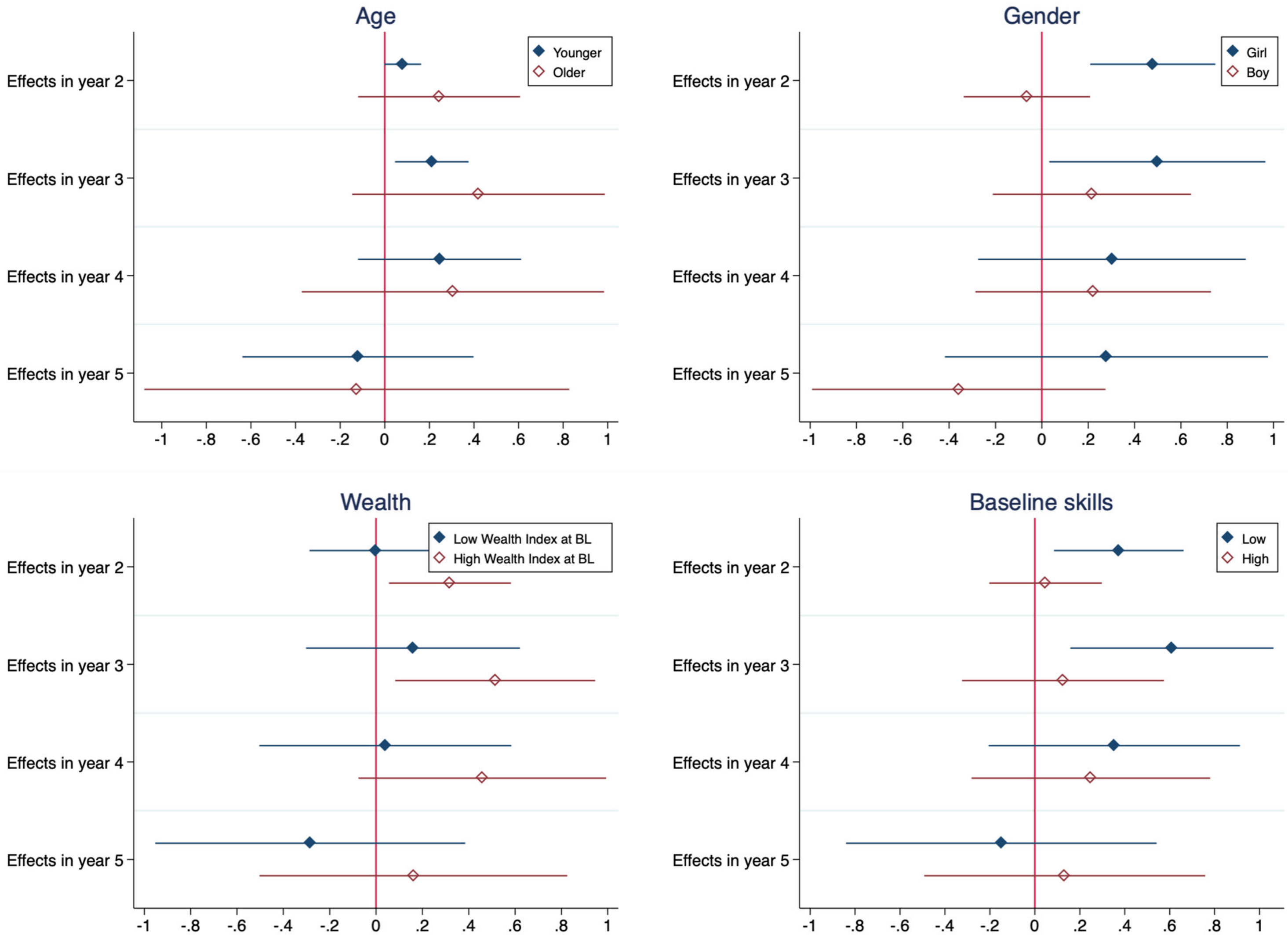
Notes: This figure plots for each study wave the share of children who were initially assigned to the treatment group and effectively enrolled in aeioTu, and the share of children who were initially assigned to the control group and were actually enrolled in aeioTu. That is, it depicts the share of the children in the treatment group that attended aeioTu in a given year and the share of the children in the control group that attended aeioTu in a given year (which we define as 'non-compliers'). The sample is restricted to age eligible children as defined in the main text.

Figure 3: ITT effects



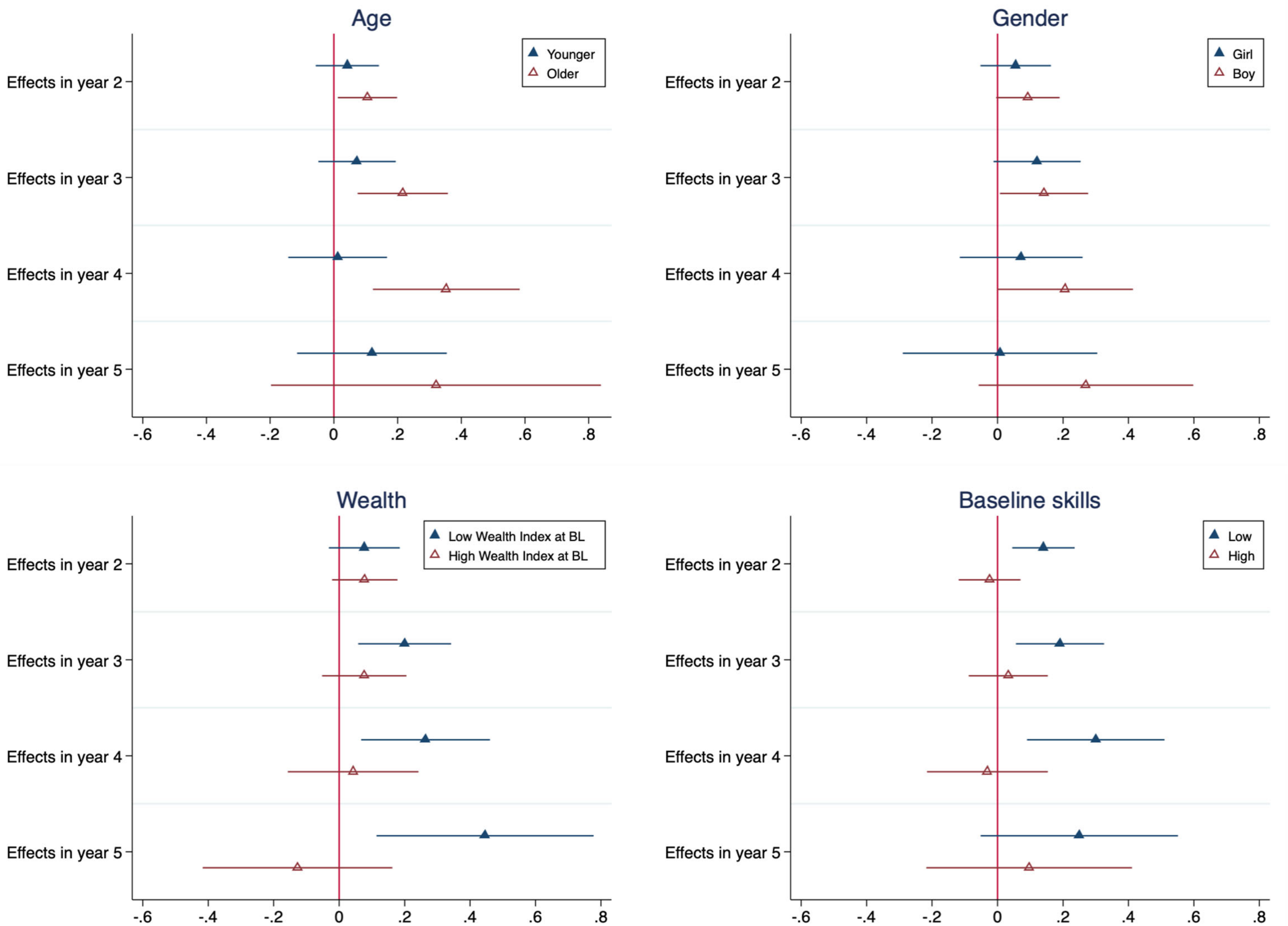
Notes: This figure reports the estimates of equation (1) and the corresponding 95% two-sided confidence intervals.

Figure 4: Heterogeneous effects on child cognitive skills



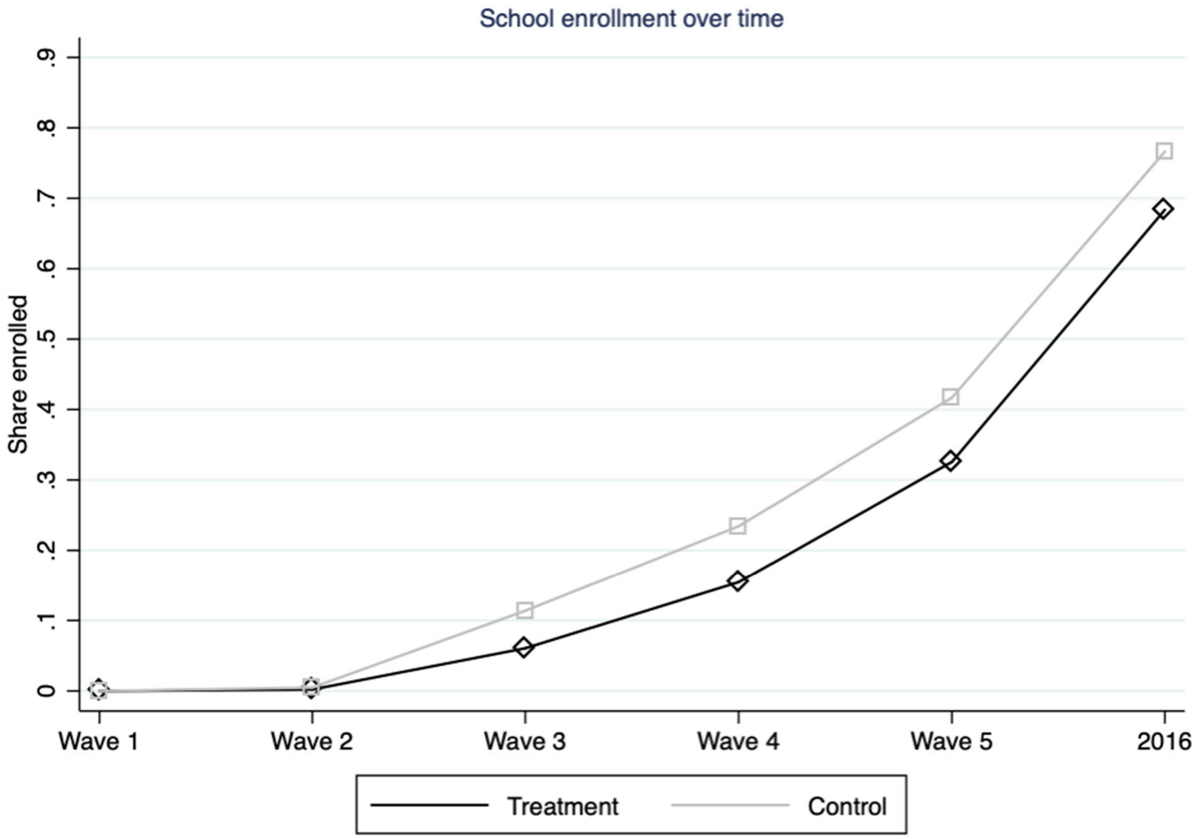
Notes: This figure plots treatment effects by subgroup. The top left panel compares younger (0-2) and older children (2-4). The top right panel compares boys and girls. The bottom left panel compares higher and lower SES children (based on whether household wealth is above or below the median in the sample). The bottom right panel compares children with lower or higher development at baseline (based on whether the outcome variable at baseline is above or below the median in the sample). The sample is restricted to age eligible children as defined in the main text.

Figure 5: Heterogeneous effects on child health



Notes: This figure plots treatment effects by subgroup. The top left panel compares younger (0-2) and older children (2-4). The top right panel compares boys and girls. The bottom left panel compares higher and lower SES children (based on whether household wealth is above or below the median in the sample). The bottom right panel compares children with lower or higher development at baseline (based on whether the outcome variable at baseline is above or below the median in the sample). The sample is restricted to age eligible children as defined in the main text.

Figure 6: School enrollment over time by treatment group



Notes: This figure plots the share of children enrolled in the formal education system in each study wave. The data for 2016 come from the administrative school records of the Integrated Enrollment System (Sistema Integrado de Matricula, SIMAT) of the Ministry of Education. The sample is restricted to age eligible children as defined in the main text.

Table 1: Study sample

	Baseline sample			Eligible in Y2 (%)			Eligible in Y3 (%)			Eligible in Y4 (%)			Eligible in Y5 (%)		
	Total	C	T	C	T	P-val	C	T	P-val	C	T	P-val	C	T	P-val
All	1,073	602	471	100%	100%	1.000	99%	98%	0.122	92%	84%	0.000	69%	55%	0.000
By cohort:															
0-1	190	109	81	100%	100%	1.000	100%	100%	1.000	100%	100%	1.000	100%	100%	1.000
1-2	321	205	116	100%	100%	1.000	100%	100%	1.000	100%	100%	1.000	100%	100%	1.000
2-3	308	168	140	100%	100%	1.000	100%	100%	1.000	100%	100%	1.000	60%	44%	0.004
3-4	254	120	134	100%	100%	1.000	96%	93%	0.374	61%	43%	0.005	0%	0%	1.000

Notes: The table reports the fraction of children that are age-eligible at each follow-up wave. C: Control; T: Treatment; P-val: P-value of the difference between treatment and control.

Table 2. Measures of early development by developmental domain and wave

	Baseline	Y2	Y3	Y4	Y5
Health					
Height					
Weight					
Arm circumference	0+	0+	0+	0+	0+
Cognitive Development					
Bayley 3rd edition	0-42	0-42	0-42	0-42	-
Peabody Picture Vocabulary Test (Spanish)	30+	30+	30+	30+	30+
ELSA Early Literacy Skills Assessment	36+	36+	36+	36+	-
Woodcock-Muñoz III subscale 10 [a]	36+	36+	36+	36+	36+
Woodcock-Muñoz III (subscales 1 [b] and 9 [c])	-	-	-	36+	36+
Head Toes Knees and Shoulders (HTKS)	48+	48+	48+	48+	48+
Socio-emotional development					
Ages & Stages: socio-emotional domain	6-60	6-60	6-60	6-60	6-60
Behavior assessment system for children	-	-	36+	36+	36+
Vineland Adaptive Behavior Scales-II	-	-	-	36+	36+

Notes: Each cell reports the ages for which the measure is available (in months) [a] Applied problems; [b] Word identification
[c] Text comprehension. The Table excludes four instruments, which were used only in Y5 and for children above four years of age: the Copy Design, Dimensional Change Card Sort and Peg Tapping for non-verbal cognitive development and the Strengtts and Difficulties Questionnaire for socio-emotional development.

Table 3: Baseline socio-demographic characteristics of children and families

		All		Control		Treatment		
	Observations	Mean	SD	Mean	SD	Mean	SD	P-value
Age in months	1,073	25.332	12.470	24.091	11.833	26.919	13.082	0.000
Gender (male)	1,073	0.523	0.500	0.523	0.500	0.522	0.500	0.975
Race (black)	1,073	0.615	0.487	0.603	0.490	0.631	0.483	0.357
Childcare use at baseline	1,073	0.171	0.377	0.145	0.352	0.206	0.405	0.008
Child has health insurance	1,073	0.778	0.416	0.779	0.415	0.777	0.417	0.938
Mother secondary complete +	1,073	0.364	0.481	0.365	0.482	0.363	0.481	0.936
Mother is single	1,073	0.276	0.447	0.282	0.451	0.268	0.443	0.588
Mother works	1,073	0.236	0.425	0.244	0.430	0.225	0.418	0.464
Father secondary complete +	1,007	0.399	0.490	0.390	0.488	0.411	0.493	0.504
Father lives at home	1,069	0.688	0.463	0.691	0.462	0.685	0.465	0.832
Wealth index	1,073	0.110	4.875	0.269	4.952	-0.094	4.772	0.226
Household size	1,073	5.355	2.007	5.409	2.089	5.287	1.898	0.323
No. of children <= 5 yrs	1,073	2.692	0.806	2.626	0.775	2.775	0.838	0.003
Children books at home	1,072	1.433	2.575	1.399	2.237	1.477	2.954	0.623
Monthly household income	914							0.272
$0- $200,000 (%)		0.158	0.365	0.148		0.170		
$200,000-$400,000 (%)		0.279	0.449	0.263		0.300		
$400,000-$700,000 (%)		0.395	0.489	0.403		0.385		
$700,000-$1,000,000 (%)		0.109	0.312	0.126		0.087		
>$1,000,000 (%)		0.059	0.236	0.060		0.058		
CCT beneficiary	1,073	0.341	0.474	0.342	0.475	0.340	0.474	0.932
Interviewed for SISBEN	1,073	0.851	0.356	0.834	0.372	0.873	0.334	0.077
Community	1,073	0.563	0.496	0.608	0.489	0.505	0.501	0.001
Neighborhood	1,073							0.010
Neighborhood 1		0.563	0.496	0.608		0.505		
Neighborhood 2		0.057	0.232	0.050		0.066		
Neighborhood 3		0.191	0.393	0.174		0.212		
Neighborhood 4		0.189	0.392	0.168		0.217		

Notes: This table reports baseline child and household characteristics by treatment group. For categorical variables the p-value reported is that of a Chi2.

Table 4: Baseline child development

		All		Control		Treatment		
	Observations	Mean	SD	Mean	SD	Mean	SD	P-value
Cognition								
Bayley Raw Scores								
Cognitive	798	48.455	14.997	47.949	14.592	49.176	15.550	0.255
Receptive	790	19.284	7.838	18.738	7.621	20.065	8.087	0.019
Expressive	794	19.618	9.326	19.201	8.819	20.214	9.988	0.132
Fine Motor	793	32.439	9.624	32.101	9.352	32.923	9.994	0.236
Gross Motor	797	46.890	13.400	46.608	12.946	47.288	14.029	0.481
Total	767	167.031	52.573	164.497	51.101	170.649	54.484	0.111
Peabody Picture Vocabulary Test								
Raw Score	384	6.995	6.024	6.235	5.165	7.659	6.626	0.021
Woodcock-Munoz Test								
Applied Problems Raw Score	244	89.148	9.750	90.170	10.395	88.280	9.117	0.132
ELSA test								
Reading Comprehension Score	245	2.469	3.087	2.263	3.043	2.649	3.125	0.330
Latent cognitive skills†	1,050	0.000	0.165	-0.010	0.148	0.013	0.183	0.021
Nutrition								
Weight-for-age Z-Score	1,046	-0.365	1.031	-0.317	1.061	-0.427	0.990	0.088
Height-for-age Z-Score	1,048	-1.123	1.080	-1.134	1.143	-1.110	0.995	0.725
Weight-for-height Z-Score	1,038	0.339	0.976	0.395	0.989	0.268	0.955	0.038
BMI Z-score	1,034	0.466	0.974	0.528	0.983	0.388	0.958	0.021
Latent health†	1,065	-0.000	1.582	0.018	1.601	-0.022	1.559	0.684
Socioemotional								
ASQ Total Score	1,060	46.239	29.369	44.101	28.263	48.953	30.530	0.008
Latent socioemotional skills†	1,068	0.000	0.656	-0.012	0.652	0.015	0.660	0.514

Notes: This table reports child developmental outcomes at baseline by treatment group. *†* We describe the construction of the latent skills in Appendix B.

Table 5: ITT effects on child development

Developmental Domain	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value	Observations
Cognitive	0.193	0.022	0.366	0.012	0.270	0.084	-0.053	0.589	3,418
	(0.096)	(0.068)	(0.162)	(0.051)	(0.196)	(0.155)	(0.237)	(0.602)	
Health	0.079	0.014	0.131	0.003	0.151	0.018	0.157	0.083	3,484
	(0.036)	(0.038)	(0.048)	(0.010)	(0.072)	(0.038)	(0.113)	(0.081)	
Socioemotional	0.009	0.571	0.069	0.904	0.031	0.753	-0.007	0.420	3,490
	(0.051)	(0.920)	(0.053)	(0.937)	(0.045)	(0.937)	(0.035)	(0.886)	

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline. The sample is restricted to age eligible children as defined in the main text.

Table 6: Effects of program enrollment on child development

Developmental Domain	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value	Observations (F-stat)
Cognitive	0.359	0.022	0.389	0.012	0.227	0.084	-0.039	0.588	3,418
	(0.177)	(0.062)	(0.171)	(0.046)	(0.164)	(0.154)	(0.176)	(0.600)	(29.851)
Health	0.146	0.015	0.136	0.004	0.124	0.018	0.116	0.084	3,484
	(0.067)	(0.085)	(0.050)	(0.019)	(0.059)	(0.020)	(0.084)	(0.074)	(30.976)
Socioemotional	0.016	0.570	0.073	0.905	0.026	0.756	-0.005	0.420	3,490
	(0.093)	(0.630)	(0.056)	(0.989)	(0.038)	(0.989)	(0.025)	(0.528)	(31.423)

Notes: 2SLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline. The sample is restricted to age eligible children as defined in the main text.

Table 7: ITT effects on intermediate outcomes

Intermediate outcome	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value	Observations
Parental time	-0.004	0.912	0.021	0.571	-0.025	0.620	-0.037	0.517	3,510
	(0.040)	(0.946)	(0.038)	(0.946)	(0.050)	(0.946)	(0.057)	(0.946)	
Nutrition	0.046	0.201	-0.002	0.971	-0.001	0.978	-0.017	0.708	3,482
	(0.036)	(0.585)	(0.041)	(0.971)	(0.034)	(0.977)	(0.045)	(0.971)	
Discipline	0.010	0.186	0.006	0.467	0.012	0.147	0.005	0.765	3,385
	(0.008)	(0.465)	(0.009)	(0.704)	(0.008)	(0.465)	(0.015)	(0.771)	

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. Two-tailed p-values and Romano-Wolf two-tailed p-values (in parentheses) are reported in the same column. All regressions include randomization strata and interviewer fixed effects. Covariates include a second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline. The sample is restricted to age eligible children as defined in the main text.

Table 8: Enrollment in formal education system over time

	Formal Education Enrollment			
	Parental reports		Administrative records (2016)	
	(1)	(2)	(3)	(4)
Treatment X Wave 2	-0.003	-0.021		
	(0.424)	(0.011)		
Treatment X Wave 3	-0.053	-0.071		
	(0.002)	(0.000)		
Treatment X Wave 4	-0.080	-0.091		
	(0.002)	(0.000)		
Treatment X Wave 5	-0.091	-0.088		
	(0.016)	(0.013)		
Treatment			-0.079	-0.088
			(0.004)	(0.002)
Control mean wave 2	0.005			
Control mean wave 3	0.114			
Control mean wave 4	0.234			
Control mean wave 5	0.417			
Control mean			0.734	
Controls	N	Y	N	Y
Observations	3755	3678	1073	1050

Notes: This table presents the results for a model of enrollment in the formal education system. The outcome variable in columns (1) and (2) is school enrollment computed from parental reports. The outcome variable in columns (3) and (4) is school enrollment in 2016 computed from the administrative school records of the Integrated Enrollment System (Sistema Integrado de Matricula, SIMAT) of the Ministry of Education. Controls include randomization strata, child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended child-care prior to baseline. Standard errors are adjusted for clustering at the child level. P-values are reported below the point estimate. The sample is restricted to age eligible children as defined in the main text.

Table 9: Comparing aeioTu to alternative education services

Child cognitive skills	IV1		IV2		IV3	
	Levels	Interacted	Levels	Interacted	Levels	Interacted
Formal education system	0.885	0.931	0.854	0.906	0.833	0.859
	(0.002)	(0.001)	(0.003)	(0.001)	(0.003)	(0.002)
aeioTu	0.289	0.276	0.303	0.291	0.327	0.302
	(0.040)	(0.047)	(0.033)	(0.039)	(0.020)	(0.032)
Alternative childcare	-0.272	-0.277	-0.273	-0.285	0.106	0.074
	(0.438)	(0.439)	(0.438)	(0.430)	(0.800)	(0.860)
Observations	3218	3218	3218	3218	3418	3418

Notes: This table presents the results of the control function approach described in the main text. IV1, IV2 and IV3 refer to different sets of instruments. For each set of instruments two models are presented. The first model labeled Levels uses the instruments in levels, while the second labeled Interacted uses the instruments in levels and interactions of randomization status with the instruments. “School” is a variable for whether the child is enrolled in school, aeioTu is a variable for whether the child is enrolled in aeioTu, and “Alternative childcare” is a dummy variable for whether the child attends other childcare services different from aeioTu. Robust standard errors clustered at the child level in parenthesis. P-values are reported below the point estimate. The sample is restricted to age eligible children as defined in the main text.

Appendices

Appendix A. Description of child assessments

Table 2 presents a summary of the child assessments included in this study, by wave, and the age-range for which each is available.

Health: As is standard practice in early intervention studies in developing countries (Fernald, Gertler, and Neufeld 2008), we measured height, weight, and arm circumference to assess the child's nutritional following World Health Organization (WHO) standards (WHO 2006; WHO 2007) for all children and all waves.

Cognitive Development: We used the Cognitive, Motor, and Language scales from the Bayley Scales of Infant Development III (BSID), the most used assessment of infant development (Bayley, 2005). The BSID is a good predictor of later measures of cognitive ability (Blaga et al. 2009; Feinstein 2003). This was administered to all children younger than 36 months of age, following guidelines for conducting this assessment. In particular, we used a translation provided under a license by the publisher (Pearson), that had been issued for another study on a similar population in Colombia (Attanasio et al., 2014), reporting a test-retest reliability of this translation of 0.95–0.98 .

As children outgrew the BSID, we administered the Peabody Picture Vocabulary test in Spanish (Test de Vocabulario en Imágenes Peabody, TVIP) (Padilla, Lugo, and Dunn 1986). The TVIP is a measure of receptive language and has been used extensively in preschool studies (Early, Maxwell, Burchinal, Alva, Bender, Bryant, et al., 2007), and shown sensitivity to early interventions (Leroy, Garcia-Guerra, Garcia, Dominguez, Rivera and Neufeld, 2008). Receptive language has been shown to be highly predictive of later development (Powell & Diamond, 2012).

We also measured child development using the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1985). The Vineland is a parent-completed that assesses children's personal and social skills in communication, daily living skills, socialization, and motor skills. This instrument was used for all children older than three years of age starting from Y4.

We measured emerging math and literacy skills using the Woodcock-Muñoz III Tests of Achievement (WM-III), which is a comprehensive set of individually administered tests of children's early literacy and mathematical skills and knowledge (Muñoz-Sandoval, Woodcock, McGrew and Mather, 2005). We used subtests #1, #9 and #10, letter-word identification, text comprehension and applied problems, respectively. Subtest # 10 was used every year for children older than 3, while subtests # 1 and #9 were included only from Y4. The scales have been translated into Spanish and adapted for Latin American contexts, and used to evaluate effects of early childhood interventions on cognitive development in infants and older children (Fernald, Kariger, Engle and Raikes, 2009).

Finally, we used the Early Literacy Skills Assessment (ELSA) measures of early literacy development (DeBruin-Parecki, 2005). This has 23 items and appears to the child to be a children's storybook. The Spanish ELSA has acceptable reliability and discriminates change (Cheadle, 2007). In our case, however, only the reading comprehension subscale exhibited reliabilities by Cronbach's alpha higher than 0.7 at all ages and all waves, so we excluded all other subscales from the statistical analyses. The ELSA was collected for all children older than three from Y1 to Y4.

Executive Function: We measured executive function using the Head-Toes-Knees and Shoulders (HTKS), which examines behavioral regulation in children (Ponitz, McClelland, Jewkes, Conner, Farris and Morrison, 2008; Ponitz, McClelland, Matthews and Morrison, 2009). HTKS requires children to remember and respond to behavioral commands. It has predictive validity with achievement and teacher-ratings of self-regulation. We measured HTKS for all children older than 4 years of age in all waves.

Socio-emotional Development: The Ages and Stages Questionnaire for the Socio-Emotional domain (ASQ:SE) is a parent-completed assessment system for children 6–60 months old. The ASQ:SE measures self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interactions with others (Squires, Bricker, and Twombly 2009a). The ASQ:SE has high levels of reliability and validity (Squires, Bricker, and Twombly 2009b). It was collected for children up until 66 months of age and all waves. Higher scores represent higher levels of socio-emotional risk or negative behaviors. To reduce the impact of illiteracy, ASQ is done as an interview.

As children grew older, in Y3 we included to the Behavior Assessment System for Children, Second Edition (BASC-II). BASC-II measures adaptive and problem behaviors through 134-160 items. The BASC-II has high levels of consistency, reliability and validity (Bracken, Keith, & Walker, 1994; Doyle, Ostrander, Skare, Crosby & August, 1997). We collected the BASC for all children older than 36 months of age.

Appendix B. Construction of the latent factors

As described in Section III.E., each developmental domain (cognitive development, socio-emotional development, and health) is measured using a variety of instruments in each wave. We used height, weight, and arm circumference to measure *health* for children at all ages in all the waves. In the case of *cognitive* and *socio-emotional* development, we used a variety of measures that changed over time, as children grew older. In Appendix Tables B1 and B2, we show the set of measures available at each study wave for each cohort, for cognitive and socio-emotional development, respectively.

There are different ways in which the information contained in these measures can be summarized in latent skills in order to estimate program impacts. For example, one could take simple averages of measures relating to the same domain, as it is commonly done in the psychology literature. As discussed in Heckman et al. (2013), this method makes somewhat arbitrary assumptions on the weights used to form averages and only controls for measurement error in these measurements through averaging.

Following Heckman et al. (2013), we implement a factor analytic approach to summarize the information contained in the different measures in a single factor. As it is standard in the psychometric literature, we specify a dedicated measurement system where each measure is associated with at most one factor (Gorsuch 1983 and 2003),

Formally, we define $N_{k,t}$ as the number of measures available to proxy for the child's skill of type k at age t and denote $m^j_{i,k,t}$ the j-th measure of skill of type k for child i at time t. We specify a linear relationship between the individual measures and the factors and write:

$$m^j_{i,k,t} = \alpha^j_{k,t} + \lambda^j_{k,t}\theta_{i,k,t} + \varepsilon^j_{i,k,t} \qquad j = 1, \dots, N_{k,t} \qquad (1)$$

where the terms $\alpha^j_{k,t}$ are intercepts, $\lambda^j_{k,t}$ are factor loadings, $\theta_{i,k,t}$ is the latent factor of skill type k, and the terms $\varepsilon^j_{i,k,t}$ are mean-zero error terms assumed to be independent of the latent factors and from each other. The above specification makes the implicit assumption that the measurement system is invariant to the treatment status. This means that any observed treatment effect operates only through the latent factors and not through changes in the measurement equations.

Because the latent factors have no natural location or scale, some normalizing restrictions are necessary for the identification of the factor model (Anderson and Rubin, 1956). By assuming mean zero for all the factors i.e. $E(\theta_{k,t}) = 0 \ \forall\, k, t$, we identify the location of our factors. The mean zero assumption is an innocuous one when one is not interested in modelling the dynamic growth of the factors over time (as would be the case in a human capital production function framework) but only in comparing the levels across two groups, as in our case.[43]

Turning to the question of scaling, valid comparison of program impacts over time requires that the scale of the factors is comparable across periods. This point is similar to the one made by Agostinelli and Wiswall, (2016) for the identification of a child human capital production function. If one had at least one developmental measure spanning all periods, then the scale of the factor could be identified by fixing the loading on this measure to be equal to one in all periods. In this way, program impacts could be expressed always in the same metric and could be compared over time. In our data, we have such measure for health, but not for socioemotional or cognitive development. This is because different age-appropriate measures were used at different points in time.[44]

To overcome this issue, we exploit the fact that we have at least one time period where a subset of measures for younger and older children, were jointly administered in order to link measures across time and express the scale of the latent factor on a common metric. Formally, consider skill k and define $M_{k,t}$ as the set of available proxy measures for skill k at time t.[45] Suppose that the set of available measures available for skill k can be partitioned in two subsets: M^A_k whose generic element is a, and M^B_k whose generic element is b. These two subsets are such that $M^A_k \cap M^B_k = \emptyset$. If there exists at least one time period t such that $a, b \in M_{k,t}$, then one can express the scale of the latent factor on a common metric.

To fix ideas consider the following simple example for skill k. Suppose we have $T = 3$ time periods and that $M^A_k = \{a_1, a_2, a_3\}$, $M^B_k = \{b_1, b_2, b_3\}$. Assume further that $M_{k,1} = M^A_k$, $M_{k,2} = M^A_k \cup M^B_k$ and $M_{k,3} = M^B_k$. Omitting the subscript k to avoid notational clutter, we can write:

[43] See Agostinelli and Wiswall (2016), for a discussion on the issues related with the location of the factors in a production function framework.

[44] Agostinelli and Wiswall (2016) further notice that age-standardizing the different measures to have a mean of zero and a standard deviation of one in the whole sample does not solve the issue related to the scaling of the factors.

[45] In panels A, B and C of Table A3 in this appendix, we describe the different measures available for each developmental domain (or intermediate outcomes such as parental investments) by wave and cohort.

$$
\text{t =1}\begin{cases} m_{i,1}^{a_1} = \alpha_1^{a_1} + \lambda_1^{a_1}\theta_{i,1} + \varepsilon_{i,1}^{a_1} \\ m_{i,1}^{a_2} = \alpha_1^{a_2} + \lambda_1^{a_2}\theta_{i,1} + \varepsilon_{i,1}^{a_2} \\ m_{i,1}^{a_3} = \alpha_1^{a_3} + \lambda_1^{a_3}\theta_{i,1} + \varepsilon_{i,1}^{a_3} \end{cases}
$$

$$
\text{t =2}\begin{cases} m_{i,2}^{a_1} = \alpha_2^{a_1} + \lambda_2^{a_1}\theta_{i,2} + \varepsilon_{i,2}^{a_1} \\ m_{i,2}^{a_2} = \alpha_2^{a_2} + \lambda_2^{a_2}\theta_{i,2} + \varepsilon_{i,2}^{a_2} \\ m_{i,2}^{a_3} = \alpha_2^{a_3} + \lambda_2^{a_3}\theta_{i,2} + \varepsilon_{i,2}^{a_3} \\ m_{i,2}^{b_1} = \alpha_2^{b_1} + \lambda_2^{b_1}\theta_{i,2} + \varepsilon_{i,2}^{b_1} \\ m_{i,2}^{b_2} = \alpha_2^{b_2} + \lambda_2^{b_2}\theta_{i,2} + \varepsilon_{i,2}^{b_2} \\ m_{i,2}^{b_3} = \alpha_2^{b_3} + \lambda_2^{b_3}\theta_{i,2} + \varepsilon_{i,2}^{b_3} \end{cases}
$$

$$
\text{t =3}\begin{cases} m_{i,3}^{b_1} = \alpha_3^{b_1} + \lambda_3^{b_1}\theta_{i,3} + \varepsilon_{i,3}^{b_1} \\ m_{i,3}^{b_2} = \alpha_3^{b_2} + \lambda_3^{b_2}\theta_{i,3} + \varepsilon_{i,3}^{b_2} \\ m_{i,3}^{b_3} = \alpha_3^{b_3} + \lambda_3^{b_3}\theta_{i,3} + \varepsilon_{i,3}^{b_3} \end{cases}
$$

The intuition is the following: one can exploit the fact that at $t = 2$ we observe both sets of measures to express the scale of the latent factor on a common metric across the three time periods. In particular, we can fix the scale of θ_1 to be comparable to that of θ_2 by setting $\lambda_1^{a_1} = \lambda_2^{a_1} = 1$.[46] Furthermore, by imposing $\lambda_2^{b_1} = \lambda_3^{b_1}$ we make sure that θ_3 is expressed in the same metric as θ_2 (and therefore in the same metric as θ_1).[47]

In our empirical application, we set the scale of the cognitive factor by normalizing the loading on the Peabody Picture Vocabulary Test in Spanish (TVIP) to one, and that of the socio-emotional factor by setting the loading on the self-regulation subscale of the Ages & Stages Questionnaire (ASQ:SE) to one.[48] For health, we observe the same measures in all periods, thus the choice of which one to normalize is arbitrary. We set the scale of the health factor by setting the loading on the weight of the child to be one in every period.

Given these assumptions and normalizations, Carneiro et al. (2003) show that the parameters in (1), the distribution of the latent factors and the distribution of the measurement errors are non-parametrically identified.[49] Once we have recovered the parameters of the measurement system, we use them to estimate a factor score for each developmental domain *k* for each observation at each time period *t* using the Bartlett scoring method (Bartlett, 1937).[50] These factor scores, which summarize the information contained in the different measures are used in the estimation of treatment impacts.

[46] The normalization $\lambda_1^{a_1} = 1$ implicitly sets the scale of the latent factor in terms of measure a_1.

[47] Notice that, in our example, we have assumed that whenever we observe a measure belonging to a subset, we also observe all other measures in that same subset. This does not need to be the case for the results to hold.

[48] We use the TVIP because it is the only cognitive measure that was administered in at least one occasion in which the Bayley scales were also administered. We use the ASQ:SE because it is the only measure that was administered in at least one occasion in which the Vineland was also administered.

[49] See Heckman et al. (2013) and Cunha and Heckman (2008). Cunha et al. (2010) discuss identification in the case where there is an unknown and potentially non-linear mapping between the measures and the factors.

[50] In practice, we estimate separate measurement systems for children of different cohorts.

Appendix Table B1: Measures used for the construction of the cognitive factor by cohort and wave

Cohort	Baseline	Y2	Y3	Y4	Y5
0-1	BSID [†]	BSID	BSID TVIP	BSID TVIP WM10 [§] ELSA RC [‡] HTKS	TVIP WM10 HTKS
1-2	BSID	BSID TVIP	BSID TVIP WM10 ELSA RC	TVIP WM10 ELSA RC HTKS	TVIP WM10 HTKS
2-3	BSID TVIP	BSID TVIP WM10 ELSA RC	TVIP WM10 ELSA RC HTKS	TVIP WM10 ELSA RC HTKS	TVIP WM10 HTKS
3-4	TVIP WM10 ELSA RC	TVIP WM10 ELSA RC HTKS	TVIP WM10 ELSA RC HTKS	TVIP WM10 ELSA RC HTKS	TVIP WM10 HTKS

Notes:[†] Bayley scales for Infant Development 3rd edition. Subscales: cognitive, receptive language, expressive language, gross and fine motor.[§] WM1 and WM9 were excluded from the construction of the latent cognitive skill because these subscales were collected starting on Y4 only, while all the other instruments were available for the span of the project.[‡] ELSA Reading Comprehension.

Appendix Table B2: Measures used for the construction of the socio-emotional factor by cohort and wave

Cohort	Baseline	Y2	Y3	Y4	Y5
0-1	ASQ:SE	ASQ:SE	ASQ:SE	ASQ:SE Vineland	Vineland
1-2	ASQ:SE	ASQ:SE	ASQ:SE	ASQ:SE Vineland	Vineland
2-3	ASQ:SE	ASQ:SE	ASQ:SE BASC	ASQ:SE BASC Vineland	Vineland
3-4	ASQ:SE	ASQ:SE	ASQ:SE BASC	BASC Vineland	Vineland

Notes:[†] Ages and Stages Questionnaire: Socio-Emotional Subscales: self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interaction

Appendix Table B3: Measures used for the construction of parental investment factors by cohorts

Factor	Measures
Parental time	• Number of hours spent by mother and father with child during a weekday • Number of hours spent by mother and father with child during weekend • Dummies for mother read, fed, walked/went out and played with child last week • Dummies for father read, fed, walked/went out and played with child last week
Discipline	• Frequency of use of different discipline strategies at home (higher if positive discipline used more frequently than physical/verbal punishment or negligent methods)
Food consumption	• Child skipped a meal last week • Nutritional content of each meal during the weekend (by nutritional group)

Appendix C: Attrition

Appendix Table C1: Attrition by wave

	Y2					Y3					Y4					Y5				
	Control		Treatment		P-val	Control		Treatment		P-val	Control		Treatment		P-val	Control		Treatment		P-val
All	40	(7%)	32	(7%)	0.923	29	(5%)	31	(7%)	0.196	32	(6%)	39	(10%)	0.018	21	(5%)	17	(7%)	0.403
By cohort:																				
0-1	9	(8%)	7	(9%)	0.925	7	(6%)	3	(4%)	0.407	10	(9%)	6	(7%)	0.664	8	(7%)	5	(6%)	0.753
1-2	10	(5%)	6	(5%)	0.907	8	(4%)	8	(7%)	0.236	10	(5%)	14	(12%)	0.019	9	(4%)	6	(5%)	0.75
2-3	15	(9%)	9	(6%)	0.415	10	(6%)	11	(8%)	0.509	10	(6%)	13	(9%)	0.268	4	(4%)	6	(10%)	0.132
3-4	6	(5%)	10	(7%)	0.420	4	(3%)	9	(7%)	0.203	2	(3%)	6	(10%)	0.071	-		-		

Notes: This table reports the number of children not re-interviewed in a given wave by treatment group and the corresponding percentages of children with respect to the age-eligible sample. C: Control; T: Treatment; P-val: P-value of the difference between treatment and control.

Appendix Table C2: Determinants of attrition

	Interviewed in all waves		
	(1)	(2)	(3)
Treatment	-0.017	-0.019	-0.028
	(0.426)	(0.396)	(0.209)
Child's gender (male)	0.002	0.004	0.012
	(0.920)	(0.860)	(0.581)
Neighborhood 1	-0.093	-0.098	-0.114
	(0.045)	(0.038)	(0.022)
Neighborhood 2	-0.066	-0.059	-0.043
	(0.018)	(0.039)	(0.138)
Neighborhood 3	-0.019	-0.014	0.000
	(0.499)	(0.629)	(0.999)
Cohort 1-2	0.017	0.003	-0.020
	(0.584)	(0.963)	(0.738)
Cohort 2-3	0.017	-0.009	-0.014
	(0.605)	(0.916)	(0.875)
Cohort 3-4	0.027	0.001	0.008
	(0.427)	(0.991)	(0.943)
Age		0.002	0.001
		(0.788)	(0.896)
Child's race (black)		0.012	0.012
		(0.603)	(0.609)
Mother secondary complete and above		0.021	-0.012
		(0.347)	(0.601)
Wealth Index		-0.005	-0.005
		(0.036)	(0.046)
No. of children <=5 yrs		-0.012	-0.015
		(0.368)	(0.318)
Childcare by baseline		0.029	0.016
		(0.334)	(0.598)
Maternal marital status (single)			-0.010
			(0.817)
Mother works			0.023
			(0.380)
Father secondary complete and above			0.039
			(0.087)
Father present			0.024
			(0.541)
Household size			-0.003
			(0.616)
Children books at home			-0.004
			(0.324)
Health insurance for child			0.050
			(0.057)
Interviewed for SISBEN			0.047
			(0.122)
CCT*			0.068
			(0.006)
Observations	1073	1073	1005

Notes: The dependent variable is a dummy variable equal to one if the child is observed in all four follow-up survey waves. P-values are reported below the point estimate. *CCT= Familias en Accion.

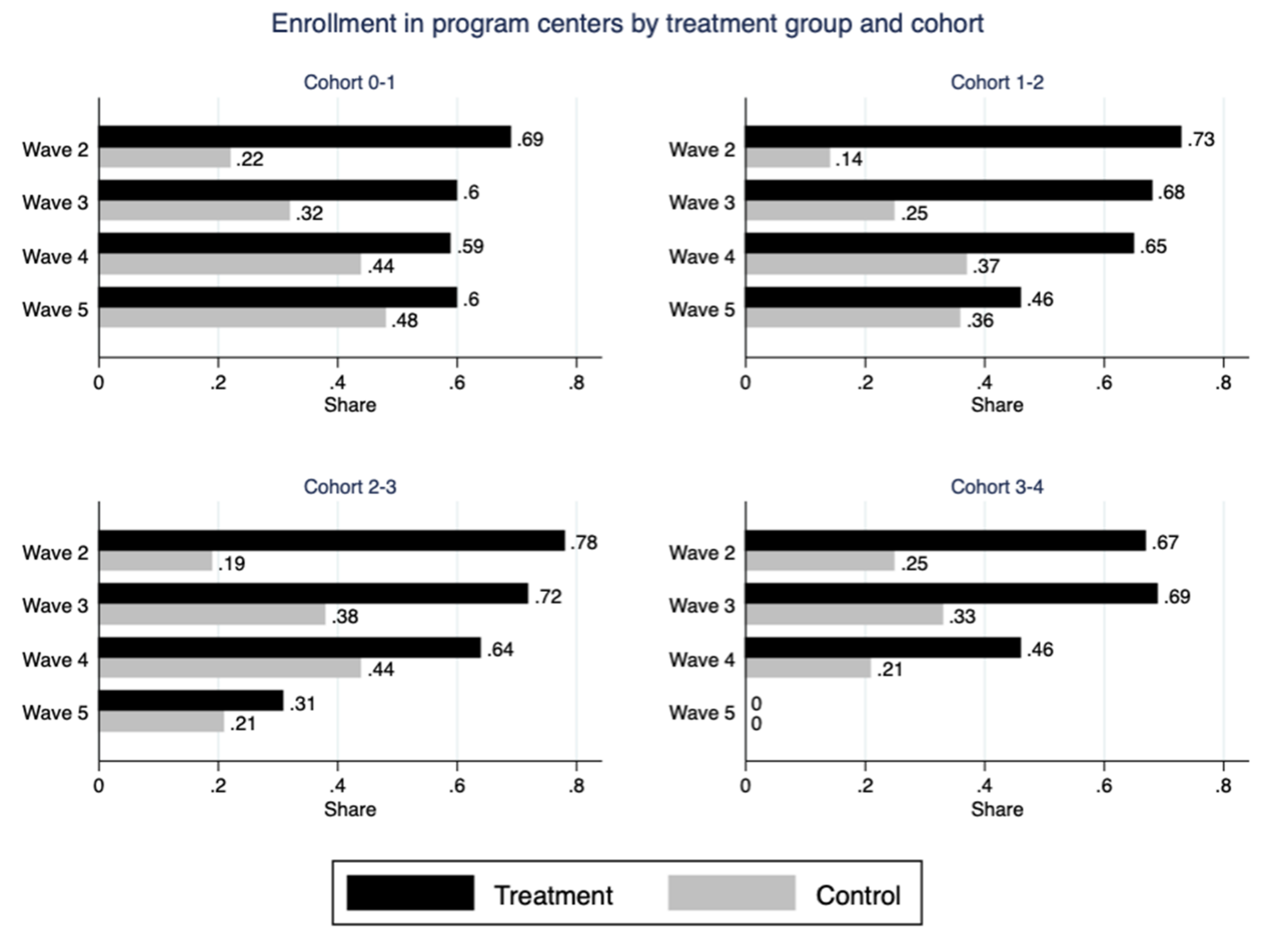
Appendix Table C3: Difference in baseline outcome variables by attrition

	Cognitive	Health	Socio-emotional
Non-attrited	0.073	0.055	-0.067
	(0.465)	(0.533)	(0.452)
Constant	0.013	-0.168	-0.078
	(0.923)	(0.141)	(0.502)
Observations	1050	1065	1068

Notes: Difference in terms of baseline outcomes by attrition. P-values are reported below the point estimate.

Appendix D: Compliance

Appendix Figure D1: Enrollment by study wave and treatment group, by cohort



Notes: This figure plots for each study wave and cohort the share of children whore were initially assigned to the treatment group and were actually enrolled in aeioTu, and the share of children who were initially assigned to the control group and were actually enrolled in aeioTu. This is "how many out of all children in the treatment group actually attended aeioTu in a given year?" and "how many out of all children in the control group actually attended aeioTu in a given year?". The sample is restricted to age eligible children as defined in the main text. There are no age-eligible children in Wave 5 in the cohort 3-4.

Appendix Table D1: Determinants of compliance

		Compliance	
	All	Treated	Control
	(1)	(2)	(3)
Treatment	-0.008	-	-
	(0.749)		
Child's gender (male)	-0.015	0.011	-0.034
	(0.552)	(0.786)	(0.280)
Neighborhood 1	0.009	0.044	0.001
	(0.833)	(0.499)	(0.984)
Neighborhood 2	0.006	0.015	-0.015
	(0.863)	(0.780)	(0.752)
Neighborhood 3	0.056	-0.013	0.122
	(0.095)	(0.803)	(0.005)
Cohort 1-2	0.071	0.006	0.135
	(0.164)	(0.946)	(0.032)
Cohort 2-3	0.030	0.034	0.028
	(0.717)	(0.810)	(0.785)
Cohort 3-4	0.020	0.009	0.035
	(0.869)	(0.965)	(0.815)
Age in months	-0.001	0.015	-0.013
	(0.830)	(0.036)	(0.033)
Child's race (black)	0.009	0.026	0.005
	(0.711)	(0.528)	(0.879)
Mother secondary complete and above	0.049	0.065	0.051
	(0.075)	(0.114)	(0.158)
Wealth Index	0.000	0.003	-0.002
	(0.915)	(0.411)	(0.653)
No. of children <=5 yrs	0.009	0.066	-0.034
	(0.619)	(0.008)	(0.138)
Childcare by baseline	0.029	-0.016	0.067
	(0.433)	(0.766)	(0.180)
Maternal marital status (single)	0.027	0.044	-0.002
	(0.601)	(0.598)	(0.977)
Mother works	-0.004	0.114	-0.081
	(0.906)	(0.017)	(0.044)
Father secondary complete and above	-0.035	0.026	-0.070
	(0.182)	(0.524)	(0.040)
Father present	0.056	0.145	-0.010
	(0.265)	(0.070)	(0.862)
Household size	0.007	-0.020	0.022
	(0.327)	(0.096)	(0.012)
Children books at home	-0.005	-0.007	-0.007
	(0.286)	(0.351)	(0.348)
Health insurance for child	-0.029	0.013	-0.044
	(0.357)	(0.789)	(0.268)
Household was interviewed for SISBEN	0.028	0.043	0.016
	(0.463)	(0.515)	(0.725)
CCT*	0.011	0.097	-0.048
	(0.691)	(0.029)	(0.201)
Observations	3298	1381	1917

The dependent variable is a dummy for whether the child complies to initial random assignment. Standard errors are adjusted for clustering at the child level. P-values are reported below the point estimate. * CCT is Familias en Accion.

Appendix Table D2: Determinants of enrollment

	Enrollment		
	All	Treated	Control
	(1)	(2)	(3)
Treatment	0.346	-	-
	(0.000)		
Child's gender (male)	0.030	0.011	0.034
	(0.219)	(0.786)	(0.280)
Neighborhood 1	0.011	0.044	-0.001
	(0.823)	(0.499)	(0.984)
Neighborhood 2	0.021	0.015	0.015
	(0.536)	(0.780)	(0.752)
Neighborhood 3	-0.068	-0.013	-0.122
	(0.046)	(0.803)	(0.005)
Cohort 1-2	-0.083	0.006	-0.135
	(0.102)	(0.946)	(0.032)
Cohort 2-3	-0.004	0.034	-0.028
	(0.964)	(0.810)	(0.785)
Cohort 3-4	-0.020	0.009	-0.035
	(0.870)	(0.965)	(0.815)
Age in months	0.015	0.015	0.013
	(0.001)	(0.036)	(0.033)
Child's race (black)	0.010	0.026	-0.005
	(0.694)	(0.528)	(0.879)
Mother secondary complete and above	0.001	0.065	-0.051
	(0.958)	(0.114)	(0.158)
Wealth Index	0.002	0.003	0.002
	(0.547)	(0.411)	(0.653)
No. of children <=5 yrs	0.047	0.066	0.034
	(0.007)	(0.008)	(0.138)
Childcare by baseline	-0.041	-0.016	-0.067
	(0.260)	(0.766)	(0.180)
Maternal marital status (single)	0.022	0.044	0.002
	(0.647)	(0.598)	(0.977)
Mother works	0.086	0.114	0.081
	(0.005)	(0.017)	(0.044)
Father secondary complete and above	0.049	0.026	0.070
	(0.061)	(0.524)	(0.040)
Father present	0.070	0.145	0.010
	(0.138)	(0.070)	(0.862)
Household size	-0.019	-0.020	-0.022
	(0.006)	(0.096)	(0.012)
Children books at home	-0.000	-0.007	0.007
	(0.950)	(0.351)	(0.348)
Health insurance for child	0.036	0.013	0.044
	(0.235)	(0.789)	(0.268)
Household was interviewed for SISBEN	0.005	0.043	-0.016
	(0.894)	(0.515)	(0.725)
CCT*	0.058	0.097	0.048
	(0.039)	(0.029)	(0.201)
Observations	3298	1381	1917

Notes: The dependent variable is a dummy for whether the child in enrolled in aeioTu. Standard errors are adjusted for clustering at the child level. P-values are reported below the point estimate. * CCT is Familias en Accion.

Appendix E: Robustness

Appendix Table E1: ITT effects on child development without tester effects

Developmental Domain	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Observations
Cognitive	0.189	0.026	0.364	0.013	0.253	0.099	-0.044	0.573	3,418
	(0.097)	(0.078)	(0.162)	(0.055)	(0.196)	(0.180)	(0.238)	(0.582)	
Health	0.084	0.010	0.131	0.003	0.148	0.020	0.142	0.105	3,484
	(0.036)	(0.028)	(0.048)	(0.008)	(0.072)	(0.033)	(0.113)	(0.104)	
Socioemotional	0.034	0.742	0.048	0.809	0.014	0.606	-0.013	0.354	3,490
	(0.052)	(0.936)	(0.055)	(0.936)	(0.053)	(0.936)	(0.035)	(0.823)	

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf p-values (in parentheses) are reported in the same column. All regressions include randomization strata. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline. The sample is restricted to age eligible children as defined in the main text.

Appendix Table E2: ITT effects on child development without controls

Developmental Domain	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Observations
Cognitive	0.162 (0.092)	0.039 (0.108)	0.388 (0.166)	0.010 (0.042)	0.273 (0.198)	0.085 (0.152)	-0.106 (0.242)	0.669 (0.678)	3,418
Health	0.068 (0.035)	0.026 (0.070)	0.119 (0.047)	0.006 (0.024)	0.139 (0.071)	0.025 (0.070)	0.123 (0.113)	0.139 (0.140)	3,484
Socioemotional	0.045 (0.051)	0.809 (0.964)	0.052 (0.055)	0.827 (0.964)	0.018 (0.053)	0.633 (0.953)	-0.007 (0.035)	0.417 (0.882)	3,490

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf p-values (in parentheses) are reported in the same column. All regressions include randomization strata. Covariates include child baseline score. The sample is restricted to age eligible children as defined in the main text.

Appendix Table E3: ITT effects on child development for the sample of children included in all waves

Developmental Domain	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Observations
Cognitive	0.179	0.002	0.208	0.054	0.324	0.052	-0.071	0.618	2,396
	(0.061)	(0.020)	(0.129)	(0.147)	(0.199)	(0.147)	(0.236)	(0.621)	
Health	0.050	0.128	0.092	0.060	0.089	0.142	0.155	0.087	2,442
	(0.044)	(0.235)	(0.059)	(0.208)	(0.083)	(0.235)	(0.114)	(0.227)	
Socioemotional	0.034	0.687	0.071	0.803	0.035	0.766	-0.005	0.443	2,437
	(0.069)	(0.967)	(0.083)	(0.967)	(0.048)	(0.967)	(0.035)	(0.895)	

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline.

Appendix Table E4: ITT effects on child development (50% eligible sample)

Developmental Domain	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Observations
Cognitive	0.179	0.030	0.359	0.014	0.264	0.095	0.022	0.464	3,075
	(0.095)	(0.115)	(0.162)	(0.070)	(0.201)	(0.187)	(0.245)	(0.430)	
Health	0.075	0.018	0.135	0.003	0.151	0.022	0.215	0.039	3,137
	(0.036)	(0.048)	(0.049)	(0.010)	(0.075)	(0.048)	(0.122)	(0.048)	
Socioemotional	0.009	0.570	0.057	0.842	0.045	0.832	-0.019	0.329	3,141
	(0.051)	(0.922)	(0.057)	(0.973)	(0.047)	(0.973)	(0.042)	(0.801)	

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline.

Appendix Table E5: ITT effects on child development (full sample)

Developmental Domain	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Observations
Cognitive	0.175	0.035	0.366	0.012	0.400	0.016	-0.007	0.513	3,918
	(0.097)	(0.079)	(0.161)	(0.059)	(0.185)	(0.059)	(0.206)	(0.473)	
Health	0.083	0.011	0.144	0.002	0.172	0.005	0.231	0.009	3,992
	(0.036)	(0.017)	(0.048)	(0.004)	(0.067)	(0.014)	(0.098)	(0.017)	
Socioemotional	0.008	0.563	0.065	0.894	0.027	0.730	0.008	0.627	4,001
	(0.051)	(0.959)	(0.052)	(0.959)	(0.044)	(0.959)	(0.024)	(0.959)	

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline.

Appendix Table E6: ITT effects on individual tests

Test	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value	Observations
Bayley cognitive	0.054 (0.028)	0.028	0.060 (0.035)	0.043	0.001 (0.053)	0.496	-	-	735
Bayley receptive	0.096 (0.042)	0.012	0.152 (0.055)	0.003	0.100 (0.096)	0.148	-	-	726
Bayley expressive	0.096 (0.044)	0.015	-0.082 (0.072)	0.873	0.134 (0.158)	0.199	-	-	728
Bayley fine motor	0.040 (0.027)	0.072	0.061 (0.045)	0.087	-0.031 (0.091)	0.635	-	-	732
Bayley gross motor	0.037 (0.026)	0.078	-0.024 (0.038)	0.735	-0.008 (0.066)	0.546	-	-	726
TVIP	0.333 (0.141)	0.009	0.370 (0.133)	0.003	0.196 (0.149)	0.095	0.065 (0.205)	0.376	3,001
WCM	-0.141 (0.103)	0.914	0.061 (0.087)	0.243	0.068 (0.090)	0.227	-0.076 (0.108)	0.759	2,725
ELSA	0.286 (0.087)	0.001	0.195 (0.064)	0.001	0.061 (0.049)	0.108	-	-	2,087
HTKS	-0.004 (0.206)	0.507	-0.022 (0.154)	0.557	-0.038 (0.096)	0.654	-0.106 (0.122)	0.808	2,121

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed unadjusted p-values are reported. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline.

Appendix F: Instrumental variables approach

To identify the model using an instrumental variable strategy we need to use in addition to the initial randomization status also other instruments. We follow Garcia, Heckman & Ziff (2018) and consider combinations of different instruments and alternative specifications. In particular, we use the three following sets of instruments:

1. IV1: the first set of instruments includes household size, an indicator for whether the mother works, maternal marital status, an indicator for whether the child's father is a member of the household, father's education, mother's marital status, and an indicator for whether the family receives Familias en Accion.

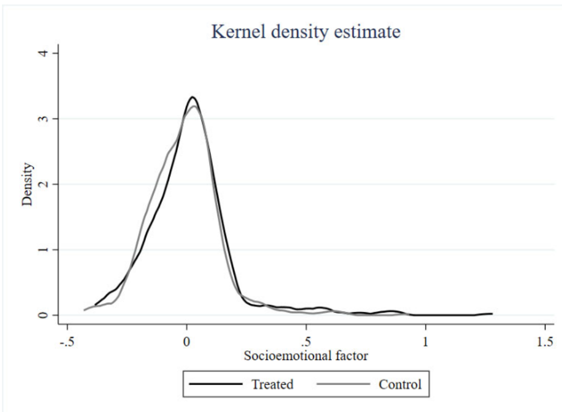
2. IV2: the second set of instruments includes, maternal marital status, an indicator for whether the child's father is a member of the household, father's education, mother's marital status, and an indicator for whether the family receives Familias en Accion.

3. IV3: the third set of instruments includes household size, maternal marital status, an indicator for whether the child's father is a member of the household, and indicator for whether the family receives Familias en Accion.

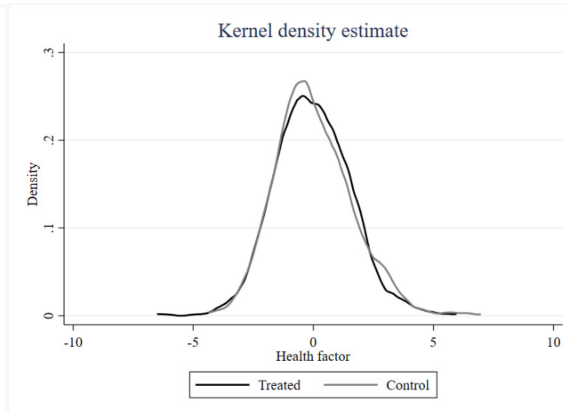
For each of these three sets of instruments we test two distinct specifications. In the first specification we use the instruments measured in levels, while in the second specification we use the instruments measured in levels and interactions of the initial randomization status with the instruments.

Appendix G: Additional Tables and Figures

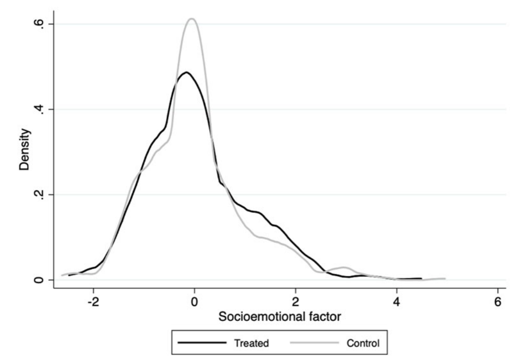
Appendix Figure G3: Distributions of latent factors



Cognitive skills

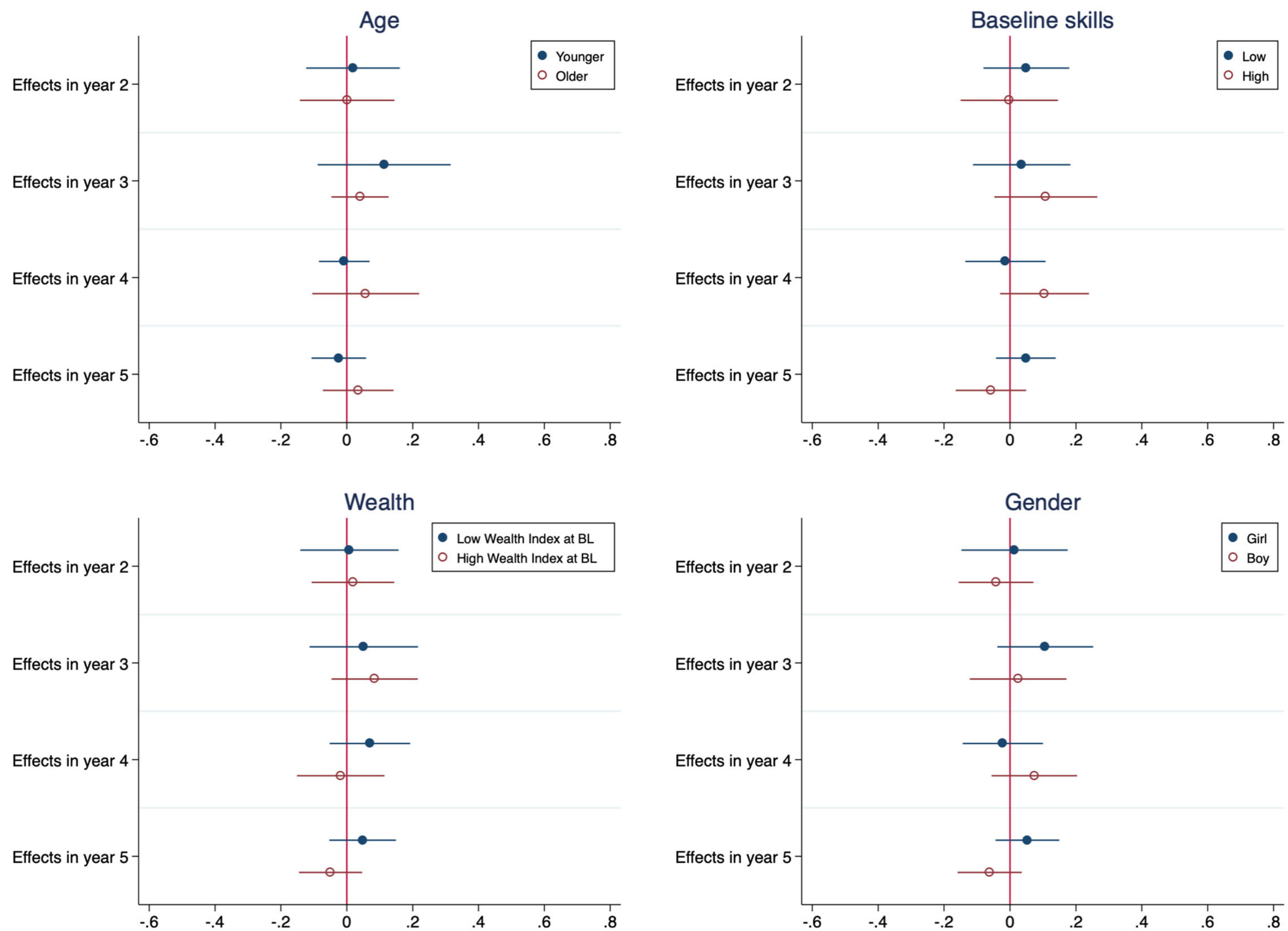


Health



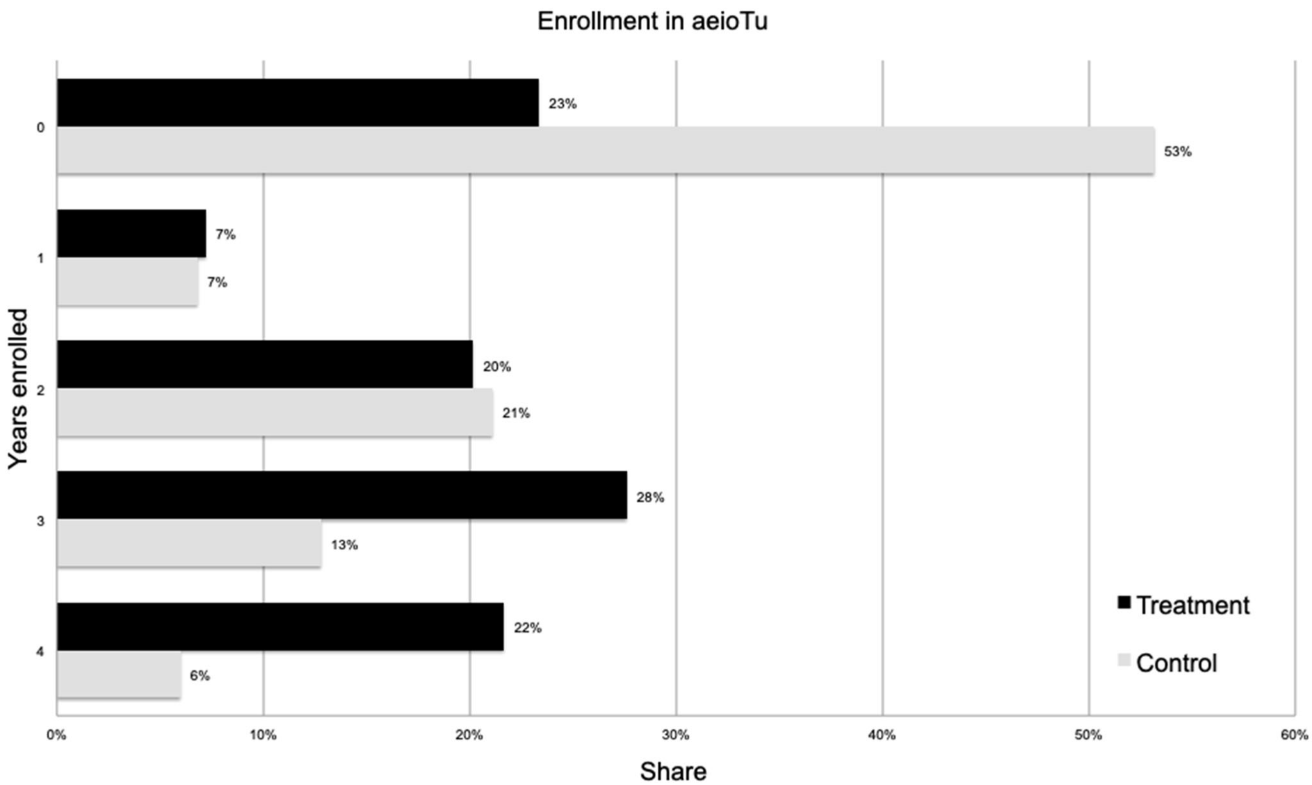
Socio-emotional skills

Appendix Figure G2: Heterogeneous effects on child socio-emotional skills



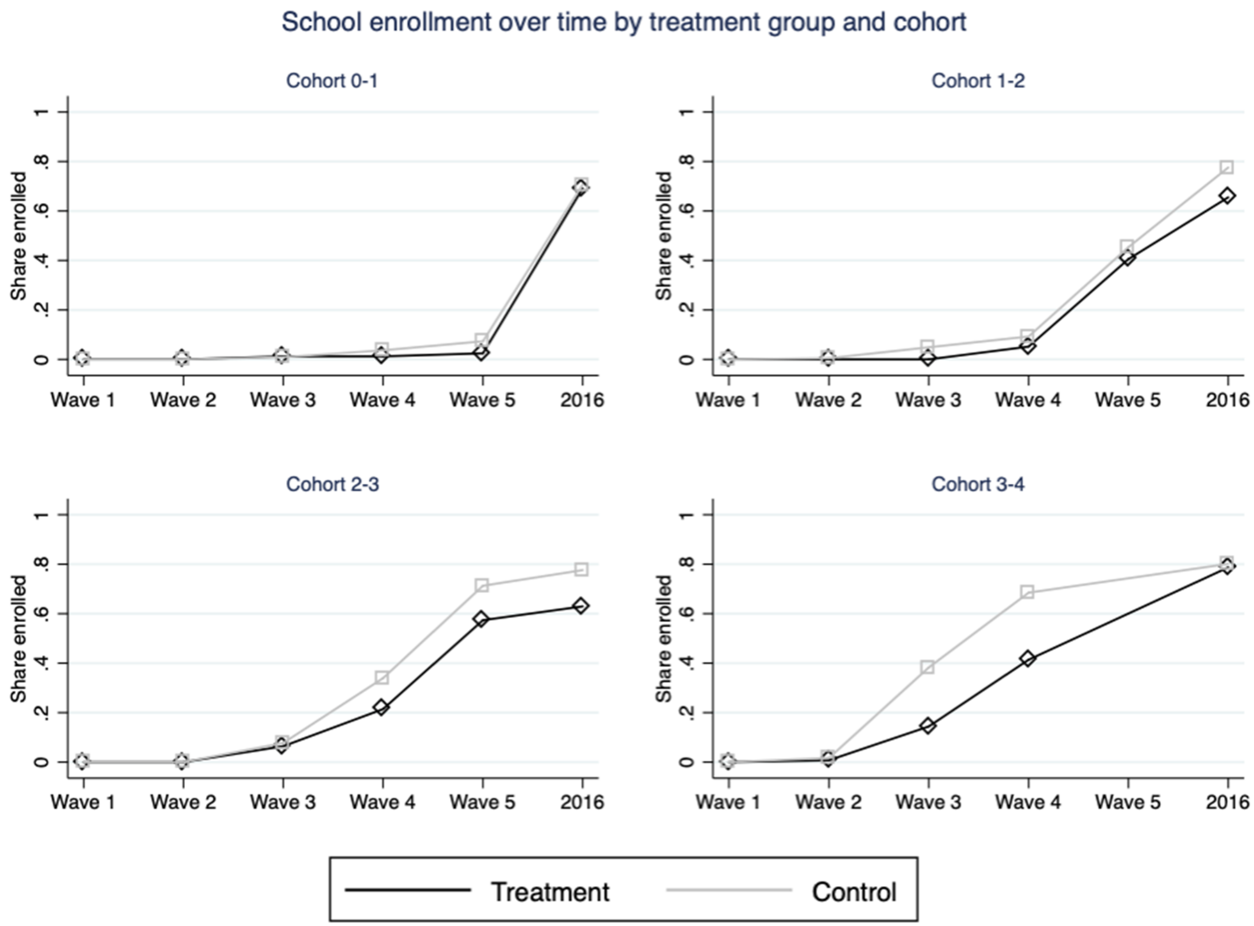
Notes: This figure plots treatment effects by subgroup. The top left panel compares younger (0-2) and older children (2-4). The top right panel compares The sample is restricted to age eligible children as defined in the main text.

Appendix Figure G3: Years of enrollment in aeioTu by treatment group



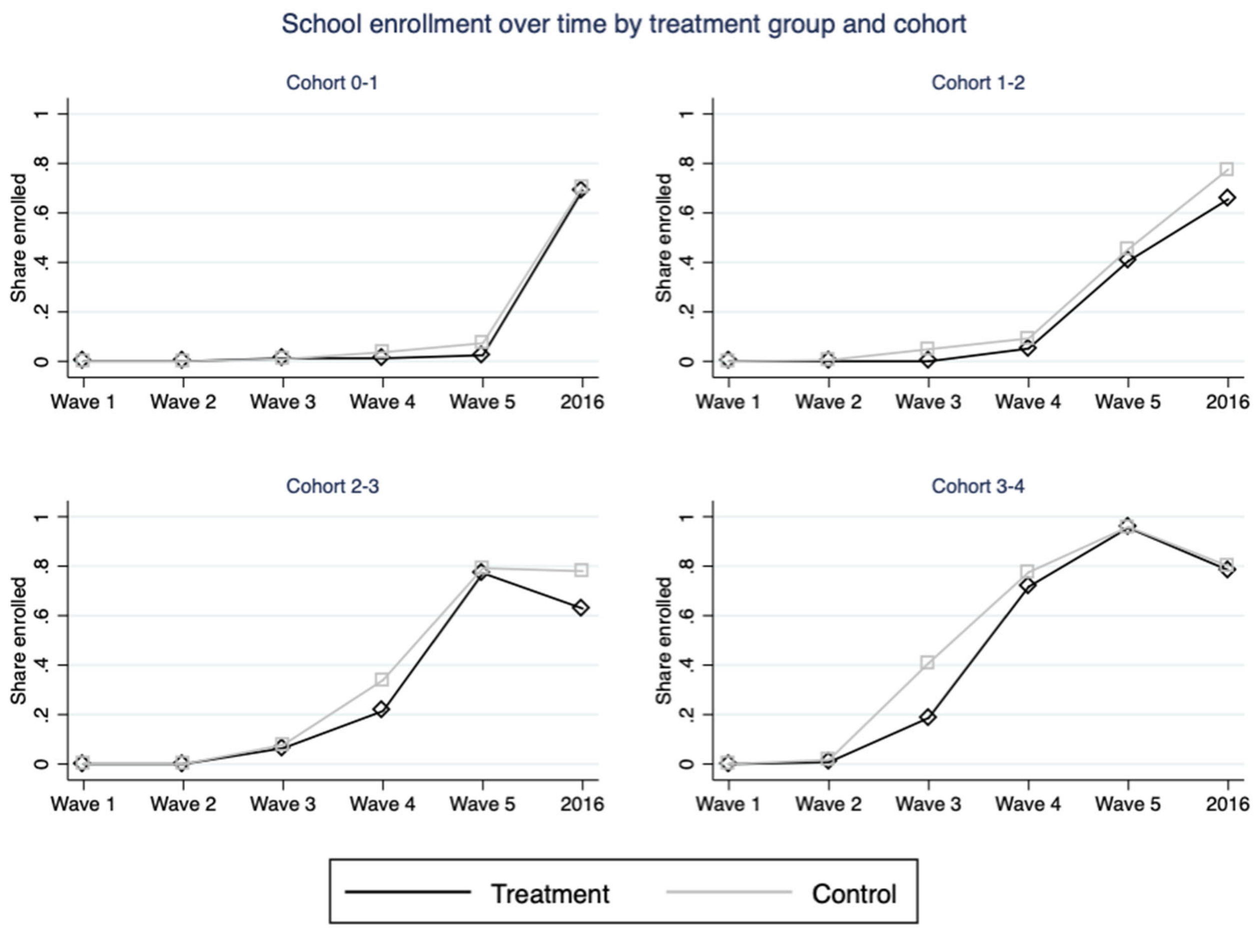
Notes: This figure plots the number of years of enrollment in an aeioTu center by treatment group. Each bar represents the share of children who was enrolled in a center for a given number of years out of the total number of children in the treatment or control group. The sample is restricted to age eligible children as defined in the main text.

Appendix Figure G4: School enrollment over time by treatment group and cohort



Notes: This figure plots the share of children enrolled in school in each study wave by cohort. The data for 2016 come from the administrative school records of the Integrated Enrollment System (Sistema Integrado de Matricula, SIMAT) of the Ministry of Education. The sample is restricted to age eligible children as defined in the main text. There are no age eligible children from cohort 3-4 in Wave5.

Appendix Figure G5: School enrollment over time by treatment group and cohort (full sample)



Notes: This figure plots the share of children enrolled in school in each study wave by cohort. The data for 2016 come from the administrative school records of the Integrated Enrollment System (Sistema Integrado de Matricula, SIMAT) of the Ministry of Education. The sample include all children irrespectively of their age-eligibility.

Appendix Table G1: Correlations of skills with sociodemographic characteristics at baseline

Sociodemographic characteristics	Cognitive	Health	Socio-emotional
BL: Maternal Education	0.059*	0.074**	-0.028
BL: Wealth Index	-0.029	-0.010	-0.041
BL: Children Books at Home	0.095***	0.011	-0.079***
Y2: Maternal Education	0.122****	0.069**	0.031
Y2: Wealth Index	-0.037	-0.006	-0.007
Y2: Children Books at Home	0.046	0.014	0.031
Y3: Maternal Education	0.176***	0.107***	-0.006
Y3: Wealth Index	0.052*	-0.010	-0.034
Y3: Children Books at Home	0.101***	0.030	-0.073**
Y4: Maternal Education	0.192***	0.098***	0.018
Y4: Wealth Index	0.047	0.004	0.026
Y4: Children Books at Home	0.066**	0.007	-0.068**
Y5: Maternal Education	0.183***	0.114***	-0.062*
Y5: Wealth Index	0.060*	0.011	-0.070**
Y5: Children Books at Home	0.061*	0.041	-0.080**

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

Appendix Table G2: Baseline child development by treatment group controlling for age

	Observations	Control Mean	Control SD	Treatment Mean	Treatment SD	P-value
Cognition						
Bayley Raw Scores						
Cognitive	798	8.369	1.020	9.015	1.098	0.131
Receptive	790	5.140	0.713	5.502	0.658	0.020
Expressive	794	1.826	0.906	2.464	1.049	0.560
Fine Motor	793	5.434	0.668	6.736	0.722	0.039
Gross Motor	797	5.160	0.926	5.825	0.966	0.191
Total	767	25.458	2.973	29.516	3.266	0.043
Peabody Picture Vocabulary Test						
Raw Score	384	6.976	9.403	17.992	7.098	0.084
Woodcock-Mu√±oz Test						
Applied Problems Raw Score	244	125.198	151.353	77.874	12.558	0.072
ELSA test						
Reading Comprehension Score	245	3.413	6.575	21.018	26.750	0.737
Latent cognitive skills †	1,050	-0.072	0.024	-0.015	0.032	0.077
Nutrition						
Weight-for-age Z-Score	1,046	0.238	0.174	0.279	0.176	0.308
Height-for-age Z-Score	1,048	-0.720	0.188	-0.429	0.180	0.539
Weight-for-height Z-Score	1,038	0.888	0.166	1.047	0.169	0.111
BMI Z-score	1,034	0.650	0.166	0.815	0.172	0.098
Latent health †	1,065	-0.841	0.263	-0.508	0.279	0.497
Socioemotional						
ASQ Total Score	1,060	20.317	4.271	13.477	4.962	0.219
Latent socioemotional skills †	1,068	0.017	0.108	-0.106	0.119	0.600

Notes: This table reports child developmental outcomes at baseline by treatment group controlling for child age. † We describe the construction of the latent skills in Appendix 3.

Appendix Table G3: Heterogeneity by age

Developmental Domain	Obs	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value
Cognitive: younger	1,870	0.087	0.304	0.212	0.104	0.253	0.070	-0.110	0.744
		(0.169)		(0.169)		(0.171)		(0.168)	
Cognitive: older	1,548	0.300	0.124	0.480	0.032	0.370	0.101	-0.148	0.617
		(0.260)		(0.259)		(0.290)		(0.497)	
Difference		0.213	0.249	0.268	0.372	0.117	0.763	-0.038	0.944
Health: younger	1,902	0.039	0.311	0.071	0.186	0.006	0.469	0.114	0.075
		(0.080)		(0.079)		(0.081)		(0.079)	
Health: older	1,582	0.103	0.096	0.213	0.004	0.350	0.000	0.320	0.017
		(0.079)		(0.079)		(0.089)		(0.150)	
Difference		0.064	0.354	0.143	0.130	0.343	0.014	0.206	0.472
Socioemotional: younger	1,901	0.020	0.614	0.114	0.954	-0.006	0.463	-0.023	0.367
		(0.068)		(0.068)		(0.069)		(0.068)	
Socioemotional: older	1,589	-0.001	0.492	0.037	0.723	0.054	0.780	0.035	0.615
		(0.061)		(0.062)		(0.070)		(0.118)	
Difference		-0.021	0.835	-0.077	0.483	0.060	0.503	0.058	0.397

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values are reported. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline.

Appendix Table G4: Heterogeneity by gender

Developmental Domain	Obs	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value
Cognitive: Girl	1,602	0.479	0.019	0.499	0.017	0.303	0.115	0.279	0.174
		(0.232)		(0.234)		(0.251)		(0.297)	
Cognitive: Boy	1,816	-0.065	0.620	0.216	0.152	0.221	0.158	-0.359	0.913
		(0.211)		(0.210)		(0.220)		(0.264)	
Difference		-0.543	0.005	-0.283	0.375	-0.081	0.834	-0.637	0.179
Health: Girl	1,629	0.056	0.243	0.121	0.067	0.072	0.204	0.008	0.470
		(0.080)		(0.081)		(0.087)		(0.103)	
Health: Boy	1,855	0.093	0.129	0.142	0.041	0.207	0.008	0.270	0.004
		(0.082)		(0.081)		(0.086)		(0.103)	
Difference		0.037	0.612	0.021	0.824	0.134	0.339	0.262	0.238
Socioemotional: Girl	1,640	0.014	0.581	0.107	0.942	-0.022	0.385	0.053	0.727
		(0.067)		(0.068)		(0.073)		(0.087)	
Socioemotional: Boy	1,850	-0.042	0.250	0.025	0.654	0.074	0.869	-0.062	0.216
		(0.063)		(0.062)		(0.066)		(0.078)	
Difference		-0.056	0.572	-0.082	0.428	0.095	0.287	-0.114	0.097

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values are reported. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline.

Appendix Table G5: Heterogeneity by household wealth

Developmental Domain	Obs	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value
Cognitive: Low Wealth	1,677	-0.003	0.505	0.159	0.242	0.040	0.435	-0.284	0.835
		(0.228)		(0.228)		(0.245)		(0.291)	
Cognitive: High Wealth	1,741	0.319	0.069	0.514	0.008	0.458	0.022	0.161	0.275
		(0.215)		(0.214)		(0.227)		(0.270)	
Difference		0.322	0.099	0.355	0.264	0.418	0.275	0.446	0.348
Health: Low Wealth	1,707	0.077	0.174	0.200	0.007	0.264	0.002	0.446	0.000
		(0.082)		(0.082)		(0.089)		(0.105)	
Health: High Wealth	1,777	0.078	0.165	0.077	0.168	0.043	0.308	-0.127	0.895
		(0.080)		(0.079)		(0.085)		(0.101)	
Difference		0.001	0.985	-0.124	0.200	-0.221	0.117	-0.573	0.010
Socioemotional: Low Wealth	1,710	0.008	0.548	0.051	0.773	0.070	0.829	0.048	0.709
		(0.068)		(0.069)		(0.074)		(0.088)	
Socioemotional: High Wealth	1,780	0.019	0.616	0.085	0.912	-0.018	0.393	-0.049	0.269
		(0.063)		(0.063)		(0.067)		(0.080)	
Difference		0.010	0.916	0.033	0.754	-0.088	0.331	-0.097	0.165

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values are reported. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline.

Appendix Table G6: Heterogeneity by baseline outcome

Developmental Domain	Obs	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value
Cognitive: low	1,713	0.320	0.072	0.542	0.006	0.323	0.083	-0.164	0.726
		(0.219)		(0.218)		(0.233)		(0.273)	
Cognitive: high	1,705	0.078	0.361	0.227	0.151	0.231	0.165	0.104	0.358
		(0.219)		(0.220)		(0.237)		(0.287)	
Difference		-0.242	0.205	-0.315	0.329	-0.092	0.812	0.268	0.563
Health: low	1,767	0.120	0.046	0.212	0.001	0.210	0.003	0.102	0.122
		(0.071)		(0.070)		(0.075)		(0.088)	
Health: high	1,717	-0.014	0.564	0.020	0.407	0.069	0.231	0.197	0.040
		(0.086)		(0.087)		(0.094)		(0.113)	
Difference		-0.134	0.053	-0.191	0.037	-0.141	0.311	0.095	0.665
Socioemotional: low	1,736	-0.114	0.035	0.090	0.925	-0.013	0.422	0.007	0.536
		(0.063)		(0.063)		(0.068)		(0.081)	
Socioemotional: high	1,754	0.116	0.958	0.045	0.748	0.078	0.860	-0.031	0.355
		(0.067)		(0.068)		(0.072)		(0.084)	
Difference		0.231	0.021	-0.045	0.667	0.091	0.309	-0.039	0.573

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values are reported. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline.

Appendix Table G7: Nutritional differences at baseline by cohort

	N	Cohort 0-2		Cohort 2-		P-value
		Mean	SD	Mean	SD	
Food fragility	1,062	0.206	0.405	0.246	0.431	0.128
Weigh-for-age	1,046	-0.141	1.060	-0.568	0.962	0.000
BMI-for-age	1,034	0.612	1.011	0.337	0.923	0.000
Weigh-for-length	1,038	0.503	1.015	0.193	0.916	0.000
Height-for-age	1,048	-0.978	1.168	-1.252	0.978	0.000
Arm circumference	1,061	14.930	1.062	15.652	1.017	0.000
Stunting	1,048	0.186	0.390	0.233	0.423	0.065
Risk of stunting	1,048	0.306	0.461	0.357	0.480	0.076

Appendix Table G8: Nutritional differences at baseline by gender

	N	Girls		Boys		P-value
		Mean	SD	Mean	SD	
Food fragility	1,062	0.243	0.429	0.213	0.410	0.244
Weigh-for-age	1,046	-0.369	1.021	-0.362	1.042	0.917
BMI-for-age	1,034	0.425	0.934	0.504	1.009	0.193
Weigh-for-length	1,038	0.308	0.932	0.367	1.013	0.336
Height-for-age	1,048	-1.083	1.096	-1.159	1.065	0.253
Arm circumference	1,061	15.259	1.110	15.353	1.088	0.167
Stunting	1,048	0.203	0.403	0.218	0.413	0.564
Risk of stunting	1,048	0.326	0.469	0.339	0.474	0.646

Appendix Table G9: School enrollment over time by treatment group (full sample)

	School enrollment			
	Parental reports		Administrative records (2016)	
	(1)	(2)	(3)	(4)
Treatment X Wave 2	-0.003	-0.047		
	(0.424)	(0.000)		
Treatment X Wave 3	-0.047	-0.092		
	(0.009)	(0.000)		
Treatment X Wave 4	-0.005	-0.058		
	(0.857)	(0.011)		
Treatment X Wave 5	0.025	-0.029		
	(0.402)	(0.248)		
Treatment			-0.079	-0.084
			(0.004)	(0.004)
Control mean wave 2	0.005			
Control mean wave 3	0.121			
Control mean wave 4	0.287			
Control mean wave 5	0.580			
Control mean			0.718	
Controls	N	Y	N	Y
Observations	4292	4200	1073	1050

Notes: This table presents the results for a model of school enrollment. The outcome variable in columns (1) and (2) is school enrollment computed from parental reports. The outcome variable in columns (3) and (4) is school enrollment in 2016 computed from from the administrative school records of the Integrated Enrollment System (Sistema Integrado de Matricula, SIMAT) of the Ministry of Education. Controls include randomization strata, child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline. Standard errors are adjusted for clustering at the child level. P-values are reported below the point estimate.

Appendix Table G10: Comparing aeioTu to counterfactual alternatives over time

Child cognitive skills

	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value	Observations
Main sample	0.193	0.022	0.366	0.012	0.270	0.084	-0.053	0.589	3,418
	(0.096)		(0.162)		(0.196)		(0.237)		
Compliers control	0.237	0.014	0.474	0.006	0.298	0.111	-0.115	0.644	2,304
	(0.108)		(0.189)		(0.244)		(0.311)		
Compliers treatment vs control in home care	0.258	0.012	0.501	0.007	0.335	0.153	0.907	0.036	1,556
	(0.114)		(0.204)		(0.327)		(0.504)		
Compliers treatment vs control in alternative childcare	0.095	0.272	0.668	0.001	0.544	0.032	0.394	0.174	1,325
	(0.157)		(0.215)		(0.292)		(0.419)		
Compliers treatment vs control not enrolled in school	0.241	0.011	0.631	0.000	0.514	0.023	0.599	0.051	1,936
	(0.105)		(0.179)		(0.257)		(0.365)		

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the sample reported in the row header. Effects interpreted in terms of SD in the control group at baseline. Robust standard errors clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values are reported. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, second order polynomial in age, race, maternal education, household wealth, number of children younger than five in household, whether the child had attended childcare prior to baseline. The sample is restricted to age eligible children as defined in the main text.