CSEF
Centre for Studies in Economics and Finance

WORKING PAPER NO. 665

The Effect of Center-Based Early Education on Disadvantaged Children's Developmental Trajectories: Experimental Evidence from Colombia

Raquel Bernal, Michele Giannola, and Milagros Nores

February 2023

This version January 2024

A previous version circulated under the title "*The Effects of a Project and Play-Based Early Education Program on Medium Term Developmental Trajectories of Young Children in a Low-Income Setting*".

University of Naples Federico II | University of Salerno | Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS AND STATISTICS – UNIVERSITY OF NAPLES FEDERICO II
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unina.it
ISSN: 2240-9696

WORKING PAPER NO. 665

The Effect of Center-Based Early Education on Disadvantaged Children's Developmental Trajectories: Experimental Evidence from Colombia

Raquel Bernal[*], Michele Giannola[†], and Milagros Nores[‡]

Abstract

Early childhood development is recognized as a global priority, but questions remain regarding how to best serve the needs of disadvantaged children in low- and middle-income countries. We study the effects of a high-quality center-based early childhood intervention on the development of disadvantaged children in Colombia. We conduct a randomized controlled trial with over 1000 children followed longitudinally over a five-year period. The intervention leads to positive and sustained effects on children's health, and positive effects on cognitive development that dissipate in the last year of the program. We find no effect on socio-emotional development or on the home environment.

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

JEL classification: J13, I10, I20, H43.

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 opening the doors to our evaluation team, and iQuartil for their excellent work managing data collection on site. 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 manuscript preparation: RB, MG & MN. All authors reviewed the results and approved the final version of the manuscript. The study obtained ethical approvals from Universidad de Los Andes, Comité de Etica (IRB Approval Number 032-2009) and from Rutgers IRB (IRB Approval Number E09-568). Trail Registry # AEARCTR- 0001903, https://www.socialscienceregistry.org/trials/1903.

[*] Economics Department and Centro de Estudios sobre Desarrollo Económico, Universidad de Los Andes

[†] Università di Napoli Federico II, CSEF, and the Institute for Fiscal Studies

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

Interest in early childhood care and education (ECCE) investments as a mean to improve the development of disadvantaged children has risen globally (Berlinski & Schady (2016); Britto et al. (2017); Nores & Barnett (2010)). Poverty hampers the development of millions of children at a great cost to individuals and their countries (Black et al. (2016)). Socio-economic disparities in cognitive and language development emerge as early as age three and widen over time (Barnett & Lamy (2013); Duncan & Magnuson (2013); Fernald et al. (2012); Heckman (2008); Rubio-Codina et al. (2015)). However, early interventions have demonstrated the potential to alter children's developmental trajectories and reduce opportunity gaps (Almond, Curie & Duque (2018); Black & Dewey (2014); Daelmans et al. (2017)). The importance of high-quality and comprehensive programming for early childhood development is now recognized as a target under the United Nations' Sustainable Development Goals for 2030 (United Nations (2016)) and has been highlighted as a global priority by the G20 (2018 communiqué, point 14). Despite this interest, questions remain regarding the most effective approaches to achieve these goals in low- and middle-income countries.

This paper presents findings from a five-year Randomized Controlled Trial (RCT) studying the impacts of a high-quality, center-based early education intervention on the development of disadvantaged children in Colombia. In 2010, we randomly assigned over 1000 children between zero and four years old from two disadvantaged communities in the North of the country to a treatment group and a control group. We then followed these children longitudinally by conducting four yearly follow-ups throughout the end of 2014. We study the program's effects on children's developmental trajectories while they are still age-eligible for pre-primary education and, thus, focus on the years preceding their enrolment in primary school. We leave the study of longer-run effects for future research.

The program we study is an integrated intervention including an educational and health component, as well as family participation. The program is named aeioTU, after the NGO that runs the ECCE centers, and its philosophy is inspired by the Reggio Emilia approach. Children enrolled in aeioTU centers receive 9 hours of care per day for 11 months of the year and are provided with 70 percent of their daily nutritional requirements through breakfast, two snacks and lunch. In line with the Reggio Emilia philosophy, the program features key elements of process quality, such as project and play-based learning, rich adult-child interactions, and intentional integration of multiple domains of child development including cognitive, socio-emotional, and health development. Centers are further characterized by low child-to-teacher ratios, high qualification for teachers, and extensive pre- and in-service training opportunities. aeioTU centers also operate an open-door policy with families, and encourage their active participation through workshops, regular communication, and meetings. This study does not seek to disentangle the distinct effect of each component of this integrated intervention.

We take a comprehensive perspective on child development and evaluate the effects of the program along several domains, including children's cognitive and socio-emotional skills, and their health status. To assess children cognitive and socio-emotional development over time, we use a large battery of well-established age-appropriate instruments that are recommended for ECCE evaluations in developing countries and have strong psychometric properties (Fernald et al. (2017)). To measure the effects of the program on children's health, we collect anthropometric information on their height, weight and arm circumference following World Health Organization guidelines (WHO (2006, 2007)). We address issues related with multiple inference, by constructing index variables and by conducting multiple hypothesis-testing corrections to control for family-wise error rate (FWER). To gain a deeper understanding of the mechanisms at play and explore potential parental behavioral responses to the intervention (Das et al. (2013); Jacoby (2002)), we further measure children's home environments through detailed interviews with their primary caregivers.

We find strong positive effects of the program on children's cognitive development, with effect sizes as large as 0.36 standard deviations (SD) relative to children in the control group. We also find positive impacts on children's health, with effect sizes between 0.08 and 0.16 SD on a summary index of child health, and a reduction in the incidence of stunting – a key marker of chronic nutritional deprivation that bears long run consequences for human capital development (Sudfeld et al. (2015)) – of 5 percentage points. We do not find impacts on children's socioemotional development or on their home environment. We further show that these treatment effects are heterogeneous along several dimensions. First, similarly to Garcia, Heckman & Ziff (2018), we document that the effects on cognitive skills are stronger for girls than for boys. Simultaneously, the effects on health are more pronounced among boys. Second, in line with recent evidence (Britto et al. (2017)), we show that children with a lower baseline level of development benefit the most from the program both in terms of cognitive development and in the health domain. Third, we find that treatment impacts are stronger for children who were older at program enrollment.

Following children over time, we then document the longitudinal effects of the intervention in the years preceding their transition to primary education. We find that health impacts are remarkably stable over time. In contrast, the effects on children's cognitive development are mitigated in the fifth year of the program, even among children who are still enrolled in program centers at that point in time. We investigate several reasons for this. First, we document that, compared to treated children, children from the control group transition earlier to primary education. We also show that primary school enrollment is associated with improved cognitive outcomes. These two findings combined could explain the convergence in cognitive outcomes that we observe and highlight the importance of considering counterfactual alternatives to understand the returns to ECCE programs (Kline & Walters (2016)). Second, we show that the probability of children from the control group enrolling in program centers increases over time, which might contribute to the mitigation in the treatment impacts if enrolled children from the control group catch-up, or if the inflow of control children disrupts learning in the centers (Carrell et al. (2018)). Third, as we explain below, aeioTU was part of a wider national early childhood policy aimed at increasing access and improving the quality of childcare services in the country. This national strategy plausibly contributed to improving the counterfactual childcare alternatives available to children in the control group. We argue that the contextual interpretation of our findings is key to understand and interpret our results (Banerjee & Duflo (2011); Deaton (2010); Deaton & Cartwright (2018)).

This study makes several contributions to the existing literature. First, we examine the dynamic impacts of aeioTU over a five-year period. Longitudinal evaluations of early childhood programs are rare in developing countries (Tanner, Candland & Odden (2015)), but necessary to fully understand the effects of such programs on disadvantage children's developmental trajectories. Second, the key features of program quality embedded in the aeioTU model, and its growth to serve over 13 thousand children nation-wide make of it a noteworthy case study, particularly so given the financial and human capital constraints present in Colombia that also preponderate in other low- and middle-income contexts. Third, our results underscore the critical role that program quality plays to produce impacts on children's development. Fourth, our findings illustrate the importance of considering the potential for impacting disadvantaged children's health in center-based programs. This is relevant for several reasons. First, integrated programs – combining a nutritional and education component – are currently the focus of the new 2023 WHO and UNICEF guidelines and goals for children under 5, so understanding their potential is fundamental to inform policy. Moreover, studies assessing the benefits of center-based programs are typically limited to children's cognitive and socio-emotional development, although these programs normally also include a health component. Finally, health benefits are typically stronger in nutrition-specific interventions.

Related literature: This paper speaks to a growing literature studying the effects of ECCE interventions in low- and middle-income countries as a strategy to improve the outcomes of vulnerable children. Two broad classes of ECCE interventions exist: home-based programs and center-based programs. This study fits within the latter.

Center-based early childhood programs have shown positive impacts on children's skills.[1] Benefits appear to be present both in developed and developing countries (Nores & Barnett (2010); Joo et al. (2020)), with the most convincing evidence coming from long-term experimental studies that followed children longitudinally in the U.S. (Barnett (2011); Barnett, and Jung (2021); Bailey et al. (2022); Englund et al. (2014); Gray-Lobe et al. (2023)). These studies document that even when initial convergence in academic outcomes between treated and untreated children occurs, long-term effects persist across other domains such as health and schooling (Nores & Prayag (forthcoming); Reynolds et al. (2017)). However, it is unclear how results from the U.S. would generalize to low- and middle-income settings for several reasons. First, the quality of care widely varies across contexts and is typically lower in developing contexts (Yoshikawa et al. (2018)). Second, the counterfactual experiences of children in the control groups – which are key to understand program impacts (Kline & Walters (2016)) – are also likely to be very different in the U.S. compared to low- and middle-income countries.

At the same time, some programs have shown neutral or negative effects of center-based care on children's development, as is the case of studies in Quebec (Canada), Chile, Denmark, Italy and Norway (Baker, Gruber & Milligan (2008, 2019); Carta & Rizzica (2018); Ichino, Fort & Zanella (2020); Gupta & Simonsen (2010); Havnes & Mogstad (2015); Kottelenberg & Lehrer (2017); Noboa-Hidalgo & Urzua (2012)). Factors related to the higher socio-economic status of attending children, the low quality of the services or the comparable quality of alternative childcare arrangements contribute to such results (Cornelissen et al. (2018)). This body of work highlights the critical role that program quality plays in producing long-term gains. Consequently, the field has shifted from a focus solely on access to ECCE services, to an equal emphasis on quality (Araujo et al. (2015); López Bóo, Dormal & Weber (2019); Nores et al. (2018)). Therefore, quality of ECCE, transitions to, and the subsequent quality of primary education have garnered increased attention to understand when convergence does and does not occur (Dietrichson, Kristiansen & Viinholt (2020); Duncan et al. (2015); Jenkins et al. (2018); McCoy, Gonzalez & Jones (2019); Bassok, Gibbs & Latham (2019)).

[1] See Bastos, Bottan, & Cristia (2017), Behrman, Cheng & Todd (2004), Berlinski, Galiani & Gertler (2009), Hojman & Lopez Boo (2022), and Nores, Bernal & Barnett (2019) for research on Latina America. See Aboud (2006), Brinkman et al. (2016), Noboa-Hidalgo & Urzua (2012) for research on Asia. See Bago et al. (2019), Martinez, Naudeau & Pereira (2017), Woldehanna (2016), Wolf (2019) for studies in Africa. See Camilli, et. al. (2010); Duncan, et. al. (2011), and Elango et al. (2015) for North America. See Biroli et al. (2018), Drange & Havnes (2019), and Felfe & Lalive (2018) for Europe.

Only a few studies outside the U.S. track a randomized set of children for more than one year to investigate the longitudinal effects of center based ECCE. Bernal et al. (2019) follow children transitioned from home-based care to center-based care in Colombia over an 18-month period and report neutral or negative effects on their development. Özler et al. (2018) study a preschool teacher and parent training program in Malawi finding short run positive effects that dissipate over time. Dean and Jayachandran (2020) consider the effects of pre-primary education in India, finding positive effects on children's cognitive development that persist in first grade. We build on this literature by experimentally studying the impact of a high-quality center-based intervention on children's developmental trajectories over a five-year period, and evaluating its effects on children's cognitive, socio-emotional and health development in the years preceding their enrolment in primary school.

Roadmap: This paper is structured as follows. We provide an overview of the early childhood policy landscape in Colombia and describe in detail the aeioTU program in Section I. Section II describes the study design, sample, measures, and empirical strategy. Section III presents the results. Section IV discusses the results and mechanisms, and Section V concludes.

I. Background and intervention

A. Early childhood education in Colombia

Colombia's growth rate declined from 7 percent in 2010 to 1.7 percent in 2018. At the same time, inequality decreased from a Gini coefficient of 0.56 in 2008 to 0.50 in 2018 (Banco de la República (2020)). About 65 percent of the 4.3 million children under the age of five in the country are born to socioeconomically disadvantaged families.[2] This vulnerable population is the focus of our study.

In Colombia, children are *eligible* for public ECCE from six months to five years of age, with the expectation that they would transition to primary school thereafter, although enrollment in first grade (*Transición*) is not universal. Language and cognitive development gaps between low and high-income children emerge as early as 12 months of age and increase to about one standard deviation by age five (Bernal, Martínez & Quintero (2015)). The Colombian Institute of Family Welfare (*Instituto Colombiano de Bienestar Familiar*, ICBF) has led efforts to mitigate these inequalities through investments in ECCE.

[2] Socioeconomic disadvantage is measured in Colombia using SISBEN scores (a proxy means-indicator based on a household socio-demographic survey).

However, enrollment in public centers for socioeconomically vulnerable children remains at 38 percent (Bernal & Camacho (2014); World Bank, (2013)). About 20 percent of children nationwide attend home-based care known as *Hogares Comunitarios de Bienestar* (HCB). Despite the original intended purpose of promoting female labor supply and the low quality of the service, recent evaluations have shown positive impacts of HCB on children's health and cognitive development (Attanasio, Di Maro & Vera-Hernández (2013); Bernal & Fernández, (2013)).

In this context, in 2011 the Colombian government launched a new national early childhood strategy called "*De Cero a Siempre*" (DCAS, "From Zero to Forever"). This strategy aims to increase enrollment in comprehensive childcare centers by expanding access, while simultaneously improving the quality of existing ECCE services (Bernal & Ramírez (2019); Comisión Inter-Sectorial para la Primera Infancia (2013)).[3] As a result, the number of children enrolled in center-based care grew from 125,000 in 2011 to about 380,000 children by 2016. The increased access to integrated center-based care facilitated by the implementation of DCAS resulted in improvements in children's language and cognitive development (Andrew et al. (forthcoming); Bernal & Ramírez (2019)).

The aeioTU program makes part of this national strategy and should therefore be interpreted in the context of these large-scale investments in ECCE that have been taking place in the country since the early 2010s.

B. The aeioTU program

The program we study is an integrated center-based early education intervention including an educational and health component, as well as family participation. The program is named aeioTU, after the NGO that runs the ECCE centers.

At the time this evaluation was planned in 2010, the program offered full-day (9 hours per day) center-based care for children under the age of five for 11 months of the year and was delivered through a public-private partnership between aeioTU and DCAS. At the start of the program, the Colombian government provided a stipend of USD 1,500 per child per year to the centers, which was supplemented by aeioTU with additional own resources.[4] This changed in 2014, when aeioTU ceased supplementing the government's stipend after careful planning and cost-structure monitoring.

[3] 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 concurrently offer nutrition, health, care, and early education.

[4] Using the average COP/USD exchange rate in 2010.

During the first years, aeioTU centers were characterized by low child-to-teacher ratios (4:1 for infants and 6:1 for toddlers), highly qualified teacher (32 percent had a BA and the rest a vocational degree in ECCE) and extensive pre- and in-service training opportunities for teachers (120 hours pre-service and 130 hours in-service). These structural aspects align with recommendations for program quality (Friedman-Krauss et al. (2023)), and contrast with comparable public center-based programs in Colombia, which had inferior teacher qualifications, higher child-to-teacher ratios (25:1 for toddlers) and lacked teacher training and coaching strategies (Bernal et al. (2019)). Over time, teacher training, curricular support, classroom materials and quality monitoring in aeioTU centers improved. Nores et al. (2018) report increasing levels of process quality comparable to other high-quality programs in the region. However, teacher-child ratios and teacher qualifications deteriorated due to public requirements and funding constraints (e.g., the government required providers to hire personnel from the HCB as some of these programs were phased off).

Another key aspect of the program is its health and nutritional component: aeioTU provides 70 percent of children's daily nutritional requirements through breakfast, two snacks and lunch. This daily nutritional requirement became mandatory in all public child-care centers in 2012, but this was not the case at the beginning of this study.[5] Moreover, an on-site nutritionist periodically monitors children's nutritional status. aeioTU also provides families with nutritional supplementation during holidays in the form of micronutrients, which stands out relative to the national policy. Research on the integration of school feeding and supplementation suggests this might be an important pathway to support disadvantaged children's development (Ogunlade et al. (2011); Simeon et al. (1998)).

Inspired by the Reggio Emilia approach (Malaguzzi (1993)), the program emphasizes project- and play-based learning, rich adult-child interactions, intentional integration of learning across multiple domains, and a balance of teacher-directed and child-initiated activities.[6] The program is further characterized by projects and themes of study, which emerge from children's play and investigations. Daily activities are guided and structured through specific pedagogical guidelines and group planning

[5] This component was underfunded by the government's stipend in about 20-25 percent, which was covered with aeioTU's own resources. We did not collect any data to monitor nutritional intakes in the centers.
[6] 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 enrvironment through explorations. These explorations belong spontaneously to children's everyday experiences, their play, their speaking, thinking, and negotiating (Malaguzzi1(993)). 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.

sessions, and incorporate play and art. To fulfill this educational plan, centers are staffed with a team of professionals including the atelier (on-site artist) and a pedagogical coordinator.[7] In line with the Reggio Emilia philosophy, aeioTU centers also operate an open-door policy with families, and encourages their active participation through workshops, regular communication, and meetings, including for example encouraging parents to come into the center to breastfeed their infants.

Finally, aeioTU utilizes data on children and classrooms for quality improvements, a feature that is critical for ECCE quality (Bernal (2015); Bowman, Donovan & Burns (2001); Yoshikawa, Weiland & Brooks-Gunn, (2016)). These characteristics align with critical aspects of high-quality ECCE highlighted in the literature (Jensen, et. al (2019); Singer, Golinkoff & Hirsh-Pasek (2006); Yoshikawa et al. (2018)). In contrast, other programs under the DCAS strategy have the freedom to define their own curricula following only broad national standards. Consequently, comparable public center-based programs are not mandated to use research-based curricula or include pedagogical guidelines for teachers' activities (Bernal et al. (2019)). We do not have comparative assessments to confirm that aeioTU had better quality than comparable programs targeting the same communities. We can only qualify that the inputs often linked to center-based quality (e.g., children to teacher ratios, training, qualifications) were higher on average.

In sum, much like the Abecedarian program in the U.S. (Ramey and Ramey (2023)), aeioTU is an integrated ECCE program including an educational and nutritional component, and family participation. For ethical reasons, it was not possible to evaluate the contribution of these different components in isolation (e.g., by offering some children only the educational component). Therefore, we estimate the impacts of the integrated intervention on children's development. Integrated programs are currently the focus of the new 2023 WHO and UNICEF guidelines and goals for children under 5 (WHO (2023); UNICEF (2023)), so understating their potential is fundamental from a policy perspective.

By 2016, aeioTU operated 20 centers and provided comprehensive ECCE to about 13,300 low-income children under the age of five around the country, which speaks to its scalability potential.

II. Methods

A. Study timeline

We evaluate the effects of aeioTU using a Randomized Controlled Trial with families of young children assigned to a treatment and control groups in two centers from two disadvantaged communities

[7] The hiring of a pedagogical coordinator was implemented across all government partners starting in 2012.

in northern Colombia. Figure 1 depicts the study's timeline, including yearly assessments. We conducted baseline assessments in late 2010 (Y1), prior to randomly assigning children and before the start of the intervention. Baseline data collection for the first center occurred between July and September 2010, and the program started in November 2010. In the second center, baseline data collection took place between October and December 2010, and the program began in March 2011. After the start of the intervention, we tracked children longitudinally over five years, throughout the end of 2014. Assessments took place mid-school year (which spans from the third week of January to November) and were conducted approximately 8, 20, 32, and 41 months after the start of the intervention. In what follows, we refer to these time points as *waves* or *years (Y)*.

B. Sampling, randomization, and masking

The evaluation was designed as a two-site randomized controlled trial using an oversubscription model. The evaluation sites were selected from centers scheduled to open around the time the study was planned. To be included, centers had to meet two criteria: (i) sufficient size to ensure adequate statistical power for detecting program impacts, and (ii) a large enough demand (i.e., oversubscription), allowing for the implementation of an assignment lottery among applicants. Two centers, which were under construction and set to open in 2010 were identified in two disadvantaged communities in northern Colombia. These centers were deemed suitable for the evaluation given the scarcity of alternative ECCE services in these communities. Location choices for the centers also included the political will of local mayors who supported the initiative by funding the infrastructure, the ICBF's approval, which prioritized underserved areas, and community support which was garnered through meetings with the community.

The two study communities are part of the Caribbean coastal region of Colombia, which is the second most populated region in the country, known for its racial and ethnic diversity. In 2010, the poverty rate in these communities was above the national average (at about 40%) and similar to that of other mid-size cities in the country. The high poverty rate partly depends on the large influx of displaced populations that moved to the region during the preceding decade and on their dependency on informal casual labor (Meisel-Roca & Ricciuli-Marín (2018)). Much like in the rest of the country, poverty rates in these communities decreased in the years 2010s. In Section III.A below, we describe how families and children in our sample compare to families in the rest of the country.

In partnership with aeioTU, we conducted a comprehensive door-to-door census with the assistance of community leaders to identify *all* children under age five residing in these communities. A total of 1,288 eligible children was identified. All families met the income-eligibility criteria based on SISBEN scores and expressed interest in enrolling their children in aeioTU if offered a slot. From the initial sample of children identified in the census, 70 were excluded from randomization: 66 children were offered a slot for reasons including being related to center staff, 2 children moved out of the communities before randomization, and 2 children exceeded the program's age eligibility by the start of services. The final sample includes a total of 1,218 children under the age of five. All families provided active consent to participate in the study.

Each of the two study centers had a capacity of 320 children, with slightly over half of that capacity allocated to children under the age of three. Random assignment to treatment and control groups was stratified by age groups (referred to as cohorts), gender and location (four neighborhoods). We used computer generated random ordered lists to assign children to the treatment and control groups. The randomization took place in a public in-person community event.

Subsequently, the centers contacted parents for enrollment. As some lottery winners opted out of the program, centers filled the available slots with children from the control group using the random-ordered lists of children. As we detail below, we consider the initial random assignment to treatment in an Intention-To-Treat (ITT) framework for the first set of results, and then use an instrumental variable approach exploiting the initial random assignment to study the effects of actual program enrolment on children's development. Therefore, our estimates do not experience issues related to imperfect compliance in randomized waitlist designs (Chaisemartin & Behagel (2020)).

Over the years, enforcement and monitoring of the study's assignment protocol were relaxed due to ethical and programmatic considerations, allowing eligible children in the control group to enroll and fill available slots. This crossover from the control to the treatment increased over time. We discuss details on compliance and crossover in section III.C, and implications for the interpretation of our results in section IV.

Child assessors and parent interviewers were blind to the treatment status of participants. However, it is plausible that parents may have communicated their status at post-testing, potentially leading to assessors learning this information over time.

C. Study sample

From a policy perspective, it is crucial to distinguish the effects of aeioTU on children's development while they are still enrolled in program centers, from the longer-run effects once children transition to primary school. This paper focuses only on the former question, estimating the impact of aeioTU on pre-school age children, and leaves the study of program impacts on longer-run outcomes for future research.

Consequently, our analysis is restricted to children who are *age-eligible* for the program in a given year. A child is age-eligible in a specific year if they are younger than five by March 31st of that year, which is when enrollment in primary school is mandated. Since data collection took place during the second half of each school year, we determined a child's eligibility in a particular year based on their eligibility for at least 30 percent of the time between a wave *t*'s data collection and the third week of January of the previous year (we show below that our findings are robust to alternative eligibility definitions). The analytical sample therefore consists of 1,073 children aged between zero and four at baseline, ensuring that we observe at least one follow-up while the child is still eligible for the program. Given the longitudinal nature of our data, older cohorts are progressively excluded as children in these cohorts transition to primary education. If we were to pool *eligible* and *non-eligible* children, the results regarding the program's impact would be confounded with the effects of enrollment in primary school. However, as a robustness check, we also report results pooling eligible and non-eligible children together, even though these likely cofound program effects with subsequent schooling experiences.

Table 1 describes the sample by treatment group and shows the changes in sample size at each wave. At baseline 1,073 children under the age of four were randomized into a treatment group (471 children were offered a slot in the centers) and a control group (602 children). By Y3, 4-7 percent of children in the oldest cohort (three to four years old at baseline) had lost eligibility, and in Y4 39 percent of the control group and 57 percent of the treatment group in this cohort was no longer eligible. By Y5, all children in the oldest cohort had lost eligibility, and 40 percent of the control group and 56 percent of the treatment group in the two-to-three-year-old cohort was also expected to transition to primary school based on their age. Children in the two youngest cohorts remained eligible throughout the five-year study period.

With a power of 80 percent, at a significance level $\alpha=0.05$ our sample size of 1,073 children allows a Minimum Detectable Effect (MDE) of 0.17 SD on children's development, without accounting for efficiency gains resulting from controlling for baseline covariates. By the last year of the study, the sample of 675 age-eligible children allows a MDE of 0.21 SD (without efficiency gains from controlling for baseline covariates). The magnitude of these MDEs is smaller than the mean effect size on child development found in a recent meta-analysis of early childhood interventions (Nores & Barnett, 2010).

D. Theory of Change

We hypothesized that the program would positively impact children's development, particularly considering the limited supply of ECCE alternatives in the two low-income communities served by the centers in the study (as we describe below, at baseline only 17.1 percent of children had used childcare services in the previous year). Specifically, we expected the program's quality components, infrastructure, and operational supports to enhance children's learning opportunities, improving their language, cognitive and socio-emotional development. Moreover, we anticipated that the nutritional component of the program would lead to improvements in children's health.

Furthermore, we predicted that the program could influence the home environment and parental engagement, but the direction of these changes was ambiguous a priori. On one hand, the program's emphasis on active family participation could enhance parental knowledge and positively influence parenting practices. In addition, if the marginal return of parental investments increases with improvements in the quality of early education (because of complementarities between these two), parents could invest additional resources in their child. On the other hand, parents might redistribute resources, including time, to non-treated children or themselves, potentially mitigating program impacts (Barrera-Osorio et al. (2011); Giannola (forthcoming)).

E. Measures and outcomes

Because of the integrated nature of the intervention, we take a comprehensive perspective on child development and evaluate the effects of the program along several domains, including children's cognitive and socio-emotional skills, and their health status. To study the effects of the program on children's development, we used well-established instruments with strong psychometric properties, commonly used in ECCE evaluations, including those in developing countries (Fernald et al. (2017)). Child assessments were administered by trained psychology graduates and seniors who achieved high reliability standards (100 percent agreement with the trainer) through a two-week training which included live reliability with children. Assessors involved in child assessments were offered a refresher training every year, and new assessors (if any) were fully trained in similar conditions. The data collection was conducted in rented and adapted spaces, ensuring identical conditions for treatment and control children. Parent survey data were collected through direct interviews conducted in separate rooms alongside the child's assessment. Families and children received small incentives for participation, and an on-site snack.

Table 2 provides an overview of the child assessments used in the study (Appendix A reports a detailed description of each measure). It is important to note that the assessments changed over time to ensure that age-appropriate measures were used at each study wave. We describe below the instruments used to measure children's health, cognitive and socio-emotional skills.

Health: We measured children's health through measurements of height, weight, and arm circumference, following contemporaneous standards and procedures set by the World Health Organization (WHO (2006, 2007)).

Cognitive skills: We measured children's cognitive development for children younger than 42 months, using the cognitive, motor, and language scales from the Bayley Scales of Infant Development III (BSID-III; Bayley (2005)). We administered the Peabody Picture Vocabulary Test in Spanish (TVIP) to assess receptive vocabulary in children over 30 months of age (Padilla, Lugo & Dunn (1986)). In all waves except the last, we measured early literacy skills in children older than 36 months using the Early Literacy Skills Assessment (ELSA; DeBruin-Parecki (2005)). In addition, for children older than three we used three subtests from the Woodcock-Muñoz III Tests of Achievement (WM-III) measuring letter-word identification, text comprehension, and applied problems resolution (Muñoz-Sandoval et al. (2005)). The applied problems subtest was used every year, while the literacy scales were added in the third follow-up. Executive functions were assessed using the Head-Toes-Knees and Shoulders (HTKS) which measures short term memory and inhibitory control in children older than four (Ponitz et al. (2008)).

Socio-emotional skills: We measured socio-emotional development using The Ages and Stages Questionnaire: Socio-Emotional (ASQ:SE), a parent-completed questionnaire measuring self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interactions with others (Squires, Bricker & Twombly (2009)). We also used the Behavior Assessment System for Children, 2nd Edition (BASC-II) for children older than 36 months to measure adaptive functioning and behavioral problems (Bracken, Keith, & Walker (1998); Doyle et al. (1997)). Starting in Y4, we included the Vineland Adaptive Behavior Scales, a parent questionnaire on personal and social skills, daily living skills, socialization, and motor skills, for children older than three (Sparrow et al. (1985)).

Home environment: We assessed children's home environments through a parental survey measuring: (1) discipline strategies, (2) nutritional and feeding habits, and (3) parental engagement with children (see Appendix A and Appendix Table B3).

Outcome variables: To keep the number of outcome variables contained, thus allowing greater statistical power, we employ a factor analytical approach following Heckman, Pinto & Savelyev (2013). This approach offers several advantages. First, it summarizes the extensive information from various developmental measures into a lower dimensional and interpretable construct. Second, it corrects for measurement error in the individual measures.

Given the longitudinal nature of the study, a critical question arises regarding how to compare treatment effects across different time points. Little (2013) emphasizes the importance of making measures comparable when they change throughout a study to capture the evolution of the same underlying developmental domain over time. Unfortunately, longitudinal studies often compare scores from different measures over time, even if it is not clear how score points translate between different assessments.[8] When age-appropriate measures for younger children are replaced with measures that are appropriate at later ages, there is limited ability to compare program effects over time. In this study, we exploit the availability of a common subset of measures available for both younger and older children in the same wave to "link" measures and compare treatment effects over time. We estimate a dedicated measurement system in which each measure is associated with at most one developmental domain (Gorsuch (1983)). We then estimate factor scores for each child at each wave using the Bartlett scoring method (Bartlett (1937)). Appendix B provides additional details on the estimation of the factor model. Our main results report treatment effects on these latent factors, but we also report results using individual measures in our analysis.[9]

F. Empirical Strategy

We estimate intention-to-treat 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 denotes survey wave indicators (*t=2, 3, 4, 5*). T_i is an indicator that takes value 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 account for baseline imbalances between the treatment and control group

[8] 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.

[9] Appendix Figure G1 plots the distributions of the factors scores, and Appendix Table G1 reports the correlations between the factors scores and baseline sociodemographic characteristics. The distributions are well behaved, and correlations show expected signs. Cognitive development is positively correlated with maternal education, household wealth and the number of books at home. The correlation between cognitive development and maternal education increases with child's age. Health is only correlated 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.

(further discussed below). We control for tester effects in the analysis but conduct robustness checks excluding these. As in similar studies (e.g., Attanasio et al. 2014), 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 level of delay as indicative of a potential disability. We cluster standard errors at the individual level, the unit of randomization.

β_t in equation (1) identifies the intention-to-treat program impact at each survey wave, that is, the effect of being offered a slot in an aeioTU center in the original lottery. To benchmark the magnitude of the effects, we report the program impacts in terms of standard deviation units of the outcome variable of the control group at baseline. Variations of this model inquire into heterogeneous effects by age (pre-specified), gender, baseline development, and household socio-economic status (SES).

Due to the comprehensive nature of the program and the longitudinal design of the evaluation, we report program impacts on numerous outcomes. To address issues of multiple inference, we compute Romano & Wolf (2005) step-down p-values, where we test within each developmental domain four hypothesis, one for each study wave. Our results below reports both unadjusted and adjusted p-values. These p-values correspond to one-tailed tests for the intervention's impact, reflecting the assumption that the intervention would not harm the children, as detailed by our "Theory of Change" (we also report two-sided p-values below).

Given crossover between the treatment and control groups (discussed in detail in section IV.C), we also report instrumental variable estimates for the impact of effective enrollment in aeioTU centers on child outcomes, by using the initial random assignment as an instrument for actual enrollment (treatment on the treated, ToT). Initial assignment to the program is a valid instrument as it was randomly allocated, and it significantly explains enrollment. A regression of program enrollment on random assignment yields a strong first stage (as we report in Table 5 below). Information on effective enrollment in aeioTU centers is obtained from administrative records. A child is recorded as enrolled if they were registered in center rosters for at least one month during the period between wave *t* and wave *t-1*. Using this variable, we construct *cumulative* enrollment at each wave and estimate the effect of cumulative participation in the program (that is the effect of being enrolled one *additional* year). This is a local average treatment effect for the "compliers", that is those children whose program participation was influenced by the lottery (Imbens & Angrist (1994)).

III. Results

A. Baseline characteristics

Table 3 describes the baseline characteristics of children and families in the sample, and reports balance by treatment assignment. Panel A reports demographic characteristics of children and families. 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 percent lived in single-headed households. Children's mothers had on average 8.5 years of education, and only 36 percent had a high school degree. There are some statistically significant differences between groups: Treatment families have a slightly higher number of children under the age of five (2.78 vs 2.63, p-value=0.037), and treated children are more likely to have attended childcare the year prior to baseline (20 vs. 14 percent, p-value=0.093). There were fewer babies than toddlers and preschoolers in the treatment group by design, as class sizes for younger children were smaller and corresponding slots were fewer. As a result, children in the treatment group are on average three months older than children in the control group (27 vs 24 months, p-value=0.005). When we replicate the analysis for the subsample of children re-interviewed in each wave, these main facts are sustained (Appendix Table G2). We control for these baseline imbalances between groups in the analysis (but also report robustness checks without controlling for these covariates).

Panel B of Table 3 describes children's development at baseline. To assess children's degree of vulnerability in terms of cognitive development, we compute standardized BSID III scores at baseline (these are not reported in Table 3, which reports raw scores).[10] The scores for cognitive, language and motor development were 90.4 (SD=13.3), 88.9 (SD=13.2), and 93.6 (SD=13.6), respectively. That is, children in the sample were about 0.7 SD below published norms (Feinstein, 2003), and slightly less developed than low-income children in rural areas of Colombia, whose scores were 92.0 for cognition and 91.6 for language (Attanasio et al. (2022)). In comparison, children between 18 and 36 months of age in Bogotá (Colombia's capital) from middle-income households score only 0.1 SD below the norming sample (Rubio-Codina et al. (2015)). Therefore, children in our sample scored significantly lower than their higher-income peers in Colombia.

Average socio-emotional (ASQ:SE) scores were instead slightly above the validation sample and comparable to children from low socio-economic urban households in the ELCA (Colombian Longitudinal Household Survey (2013)). At baseline, children were also highly nutritionally vulnerable,

[10] Bayley III composites computed based on published norms provided by test developers have mean 100 and standard deviation 15. These are reported here only for comparison but are not used in the analysis.

with a height-for-age one standard deviation below WHO standards (WHO (2006, 2007)). About 21.6 percent of children in the sample were stunted, compared to a national rate of 15.2 percent in rural areas and 12 percent in urban areas (Colombian Longitudinal Household Survey (2013)). In sum, children in our sample show levels of cognitive and health deprivation comparable to those of other low-income children in Colombia, and lower levels of development compared to wealthier children in the country and globally.

There are no significant differences between children in the treatment and control group at baseline in terms of cognitive development and health, but children in the treatment group have slightly higher ASQ:SE scores than children in the control group.

B. Attrition

A detailed analysis of attrition is reported in Appendix C. The study has low levels of attrition ranging from 5 to 10 percent, notable for a five-year longitudinal study (Appendix Table C1). Children missing in each wave varied slightly, as some children who were not found in one wave could be found in later years. We estimate the probability of not attriting as a function of treatment status and household background characteristics (Appendix Table C2). Results show that: (i) attrition rates do not differ between the treatment and control groups, (ii) wealthier families are less likely to be surveyed in all waves, and (iii) beneficiaries of Colombia's conditional cash transfer (CCT) program are less likely to leave the sample. We also estimate the difference in the main outcome variables at baseline between attriters and non-attriters, revealing non-significant differences between these two groups (Appendix Table C3).

C. Compliance

In Appendix Table D1, we regress the total number of years a child is enrolled in a program center on an indicator for being randomized to the treatment group in the original lottery. Winning the lottery has a strong positive impact on program participation. The results in column 1, where we only control for randomization strata fixed effects, show that children randomized to the control group attended the program for, on average, 1.12 years, while children randomized to the treatment group attended aeioTU for, on average, 2.56 years, a 100% increase (p-value for the difference < 0.001). In column 2 we further control for baseline covariates and find a very similar effect.

Compliance with treatment assignment decreased over time, as it is often the case in longitudinal evaluations of ECCE programs (e.g., Attanasio et al. (2022)). Figure 2 reports enrollment rates by randomization status at each study wave (cohort-specific rates are reported in Appendix Figure D1). In Y2, 72 percent of lottery winners enrolled in centers, and 81 percent of lottery losers did not. Enforcement and monitoring over the study's assignment protocol was relaxed over the years due to ethical and programmatic reasons, allowing eligible children in the control group to enroll. By the last year of the study, a significant fraction of children in the treatment group moved to other public ECCE options, as these expanded due to the national early childhood strategy. Concurrently, enrollment of control children in aeioTU centers increased from 19 percent in Y2 to 35 percent in Y5. These patterns are heterogeneous between cohorts, with higher enrollment rates for the youngest cohorts and lower for the oldest cohorts, consistent with older children gradually transitioning to primary education (Appendix Figure D1).

These results show cross-over from control to treatment during the study, particularly for children enrolled as babies or toddlers. This has important implications for the interpretation of our results. We address this in two ways. First, we use an intention-to-treat approach and consider initial random assignment to treatment. Second, we use initial random assignment as an instrument for effective program enrollment to estimate a Local Average Treatment Effect, that is the average treatment effect for children whose participation in aeioTU was affected by the lottery (Imbens & Angrist (1994)).

We analyze the determinants of compliance for both the entire sample, and separately for the treatment and control groups in Appendix Table D2. We find that: (i) compliance is unrelated to treatment assignment, and (ii) on average, compliers and non-compliers do not differ significantly in observable characteristics. We explore the determinants of enrollment in Appendix Table D3 and show that: (i) as discussed above, 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 in program centers.

D. Intention-to-treat effects on child development

Table 4 presents intention-to-treat effects of aeioTU by domain of child development. The estimates are based on equation (1), and each coefficient represents the impact of aeioTU in wave t for children who are still age eligible for ECCE in that year. The effects are reported as fractions of a standard deviation (SD) of the outcome in the control group at baseline. We include both unadjusted p-values, and p-values adjusted for multiple hypotheses testing. For each domain of child development, the adjusted Romano-Wolf p-values test the significance of the treatment effects across the four study waves.

We find a positive and significant impact of the program on child cognitive skills. At the first follow-up (Y1), just eight months into the program, treated children's cognitive skills are 0.193 SD higher than children in the control group (step-down p-value=0.068). These effects increase to 0.366 SD in Y3 (step-down p-value=0.051) and 0.27 SD in Y4 (step down p-value=0.155). The effect in Y5 is not statistically different from zero, although the coefficient has a negative sign (-0.053 SD, step-down p-value=0.602). We discuss potential reasons for this convergence in section IV.D. The effect in Y2, Y3 and Y4 are not statistically different from each other, while the effect in Y5 is statistically different from the effect in Y3 at the 10 percent level.

We find positive and significant effects on child health throughout the study. The effect size is 0.079 SD in Y2 (step-down p-value=0.038). The point estimates increase over time with an effect size of 0.131 SD in Y3 (step-down p-value=0.010), 0.151 SD in Y4 (step-down p-value=0.038), and 0.157 in Y5 (step-down p-value=0.081). These estimates are not statistically different across study waves. Finally, throughout the study we find that the effects on children's socio-emotional development are always small in magnitude (effect sized between -0.007 and 0.069 SD) and never statistically significant.

We run several robustness checks for these results in Appendix E. Appendix Table E1 demonstrates the robustness of the results when excluding tester effects. Appendix Table E2 examines the sensitivity of our estimates to the exclusion of household characteristics that were unbalanced at baseline. Again, we confirm our main findings. To investigate whether the changes in treatment effects over time for child cognitive skills are influenced by changes in the estimation sample (since we progressively exclude older children), Appendix Tables E3 reports the treatment effects only on the children who remain age-eligible until the final study year. The patterns are consistent with those in Table 4, but less precisely estimated due to the smaller sample size. Appendix Tables E4 reports the effects for children who are age eligible for ECCE for 50 percent of the time between two data waves (instead of 30 percent as in our main specification). These coefficients are very similar to the ones reported in Table 4, with a slight increase in the estimated impacts on child health when using the more conservative 50 percent eligibility threshold (Y5 effect is 0.21 SD compared to 0.16 SD in our main specification). In Appendix Tables E5-E7 we report similar results for the full sample of children that were originally randomized to treatment and control, regardless of their age-eligibility for ECCE in each study wave. Appendix Table E5 shows that, when using the full sample, the effects on children's cognitive skills are similar in Y2 and Y3 and larger in Y4 (0.40 compared to 0.27 in our main specification), while the effects on children's health are larger in Y5 (the effect is 0.231 compared to 0.157 in our main specification), suggesting that the effects of aeioTU are likely to last once children progress into primary school. Appendix Figure E1 reports treatment effects at each study wave together with 95 percent two-sided confidence intervals.

Finally, in Appendix Tables E8 and E9 report the effects of the program on each individual test included in the cognitive factor and on the individual health measures. Appendix Table E8 shows that the effects on the individual tests mirror those estimated on the cognitive latent factor. In line with the results in Table 4, in Appendix Table E9 we show that children in the treatment group have a higher weight-for-age and height-for-age compared to children in the control group. We also find that the probability of being stunted (length-for-age Z-score < -2, a primary manifestation of chronic nutritional deprivation) decreases by between 3 and 5 percentage points for children in the treatment group, corresponding to a 13-22 percent decrease relative to children in the control group at baseline.

E. Effects of program enrollment on child development

Appendix Figure G2 plots the distribution of cumulative enrollment by randomization status. As previously discussed, compliance decreased over time, with 47 percent of control children enrolled for at least one year, and 19 percent enrolled for three or more years throughout the study period. Nonetheless, as shown in Appendix Table D1, enrollment is significantly higher for children in the treatment group, allowing us to use random assignment as an exogenous shifter for program participation: a regression of enrollment on program random assignment yields a highly significant first stage (the F-statistic is close to 30, as reported in Table 5).

Table 5 reports the two stages least squares (2SLS) results instrumenting *cumulative* program enrollment with random assignment to treatment, thus, the coefficients in Table 5 answer the question "what is the effect of one *additional* year in aeioTU?". The 2SLS results mirror the ITT effects presented in Table 4, but the effects sizes are generally larger, as expected. Pooling all waves together, our estimates imply that one additional year in the program improves children's cognitive skills by 0.226 SD (p-value=0.038) and their health by 0.129 SD (p-value=0.009). The effect of one additional year of enrollment in aeioTU on child cognitive skills is 0.359 SD in Y2 (step-down p-value=0.062), 0.389 SD in Y3 (step-down p-value=0.046), 0.227 SD in Y4 (step-down p-value=0.154) and -0.039 SD in Y5 (step-down p-value=0.600). The effects on health are 0.146 SD in Y2 (step-down p-value=0.085), 0.136 SD in Y3 (step-down p-value=0.019), 0.124 SD in Y4 (step-down p-value=0.020), and 0.116 SD in Y5 (step-down p-value=0.074). We find no effects on children's socio-emotional skills (the point estimates are always small in magnitude and never statistically different from zero).

F. Effects on intermediate outcomes

We report the program's impacts on intermediate outcomes in Table 6. We find no effects on parental time, nutritional investments, or discipline strategies. The point estimates are consistently close to zero and lack statistical significance, indicating that the program does not have meaningful impacts on these intermediate outcomes. These findings imply that the treatment neither crowded-in nor crowded-out parental investments. Beaton & Ghassemi (1982) and Jacoby (2002) discuss concerns over in-kind transfers targeted to children, as parents could alter their child's nutritional intake in response to such programs. Similarly, Das et. al (2013), find that parents in India and Zambia respond to increases in public education spending by reducing private investments in children. Our results indicate that the program's nutritional in-kind transfer "sticks" to the child, as in Jacoby (2002). Similarly, parents do not change their time investment and discipline strategies at home. The fact that parents did not alter their investments suggests that the positive impacts on children's cognitive development and health are due to the intervention itself rather than due to changes induced at home.

G. Heterogeneous effects

Figures 3 and 4 and Appendix Tables G3-G6 present heterogeneous ITT effects on child development along four dimensions: child age (younger 0-2 vs. older children 2-4, corresponding to the groups we use to stratify children to treatment), gender, baseline development (splitting the sample at the median level of the outcome), and household wealth (splitting the sample at the median level of baseline household wealth). However, it should be noted that, with the exception of child age, these heterogeneity analyses were not pre-specified at the outset, and we may be underpowered to detect small differences across groups.

We find positive effects on child cognitive development for both younger and older children from Y2 to Y4 (although we do not reject the null hypothesis of equality of coefficients across the two groups, see Appendix Table G3). The effects on cognitive skills are particularly strong for girls. In Y2 and Y3 the effect sizes on girls are 0.479 (p-value=0.019) and 0.499 (p-value=0.017). For girls, treatment effects remain positive throughout the study although they are not statistically different from zero in later waves. Effects for boys are positive but never statistically different from zero in Y2, Y3 and Y4. In Y5, the effect on boys' cognitive skills is negative (coefficient = -0.359, p-value=0.913), while it stays positive for girls (coefficient = 0.279, p-value=0.174). We reject the null hypothesis of equality of coefficients across genders in Y2 (p-value = 0.005) but not in later years (Appendix Table G4).

In terms of socio-economic background, we find suggestive evidence of stronger effects on cognitive development for children from relatively less disadvantaged households compared to children from the most disadvantaged households (Figure 3 and Appendix Table G5), although we can only reject the equality of coefficients in Y2. The effect on children from less vulnerable households is 0.319 SD in Y2 (p-value=0.069) and increases to 0.514 SD in Y3 (p-value=0.008) and 0.458 SD in Y4 (p-value=0.022). On the other hand, the effects for more vulnerable children are smaller and never statistically different from zero. Finally, we find suggestive evidence of larger effects for children with lower cognitive development at baseline in all waves but the last (Figure 3 and Appendix Table G6), although we cannot reject the null hypothesis of equality of coefficients across the two groups.

Figure 4 and Appendix Tables G3-G6 show similar results for health. Older children benefitted the most from the intervention, with consistently larger point estimates. Treatment effects for older children increase over time, with effect size of 0.103 SD in Y2 (p-value=0.096), 0.213 SD in Y3 (p-value=0.004), 0.350 SD in Y4 (p-value<0.001), and 0.320 in Y5 (p-value=0.017). Treatment effects are consistently smaller for younger children (Appendix Table G3), although we only reject the equality of coefficients in Y4. In terms of gender, we observe stronger effects for boys, with effect sizes of 0.27 SD in Y5 (p-value=0.004). This contrasts with the results for cognitive skills. We also find suggestive evidence that the positive effects on health are driven by the most socioeconomically vulnerable children, with effects size of 0.264 SD in Y4 (p-value 0.002), and 0.446 SD in Y5 (p-value < 0.001). In contrast, the effects on the health of children from less vulnerable households are small and never 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 coefficients across the two groups in Y2 and Y3 but not in later waves, see Appendix Table G6).

There are no meaningful heterogeneous effects on socioemotional development. The results are reported in Appendix Figure G3 and Appendix Tables G3-G6. We discuss these results in detail in the next section.

IV. Discussion and Mechanisms

A. Health

We find sustained positive effects of aeioTU on children's health, with effect sizes ranging from 0.08 SD in Y2 to 0.16 SD in Y5. We do not find changes in nutritional investments at home, supporting the notion that these in-kind transfers "stick" to the child (Jacoby (2002)). These findings contrast with those in Bernal & Fernández (2013) who find no nutritional gains in a home-based care program with a comparable nutritional component in Colombia. Similarly, Bernal et al. (2019) report small effects (0.05 SD) on children's nutritional status in public center-based care. Most relevantly, Andrew et al. (forthcoming) finds that improvements in center-based care brought about by DCAS lead to no significant gains on children's health. In contrast, we find that aeioTU has large effects on children's health development that persist over time. The magnitudes of these effects are also important in the context of the existing international literature: Nores & Barnett's (2010) meta-analysis report average effects of 0.31 SD for nutrition-specific interventions, and effect sizes of 0.23 SD for interventions that combined a nutritional and educational component.

The larger effect on older children, may be due to their higher baseline vulnerability on anthropometric indicators and in terms of food fragility compared to their younger peers (Appendix Table G7). At baseline, older children had a higher prevalence of stunting (23.3 vs 18.6 percent) and greater height-for-age deficits (-1.25 SD vs -0.97 SD). This vulnerability suggests that the nutritional component of the intervention may have had a greater effect on older children. Another possible explanation for the differential effects by age may be that younger children are less likely than older ones to consume their meals in center-based care (Andrew et al. (2018)). We also find suggestive evidence that the effects on health are larger for boys, but neither baseline differences in nutritional status (Appendix Table G8) nor food fragility or nutritional investments at home (results not reported) explain this gender heterogeneity. We speculate that gender differences in eating behaviors may contribute to these gendered effects (Keller et al. (2019)).

Finally, the fact that we find somewhat larger effects on children's health status when considering the full sample of children, regardless of their age eligibility, suggests that the health impacts might be persistent over time as children progress through primary education. Future research shall investigate long-run effects.

B. Cognitive development

We find positive program impacts on children's cognitive skills, with effect sizes as large as 0.36 SD. ToT effects are larger and close to 0.40 SD. These effects are sizable considering the cognitive development gap of nearly one standard deviation existing between high and low SES children in Colombia (Bernal, Martínez & Quintero (2015)). The magnitude of these effects is also important in relation to similar studies conducted in Colombia and other ECCE programs evaluated worldwide. In Colombia, Bernal & Fernández (2013) find positive effects on children's cognitive skills after 15 months of exposure to *Hogares Comunitarios*, with effect ranging from 0.15 to 0.30 SD. Bernal et al. (2019) report negative effects of -0.10 SD of public center-based care compared to home-based care. Andrew et al. (forthcoming) report effects of 0.15 SD on cognitive development of a teacher and coaching intervention in urban centers in Colombia. A meta-analysis of non-U.S. interventions by Nores & Barnett (2010) reports average effects of early education programs on cognitive development of about 0.25 SD. The impacts of aeioTU are at the higher end of this range of estimates and emerge early on, just eight months after the start of the program.

Girls in our study show particularly strong effects, with effect sizes close to 0.50 SD in Y2 and Y3 and positive throughout the study. These results differ from other studies' mixed evidence on gender differences from ECCE interventions (Magnuson et al., 2016). While there are no systematic differences in SES between boys' and girls' families, there is evidence of higher parental-reported interactions with boys in play and reading (Nores, Bernal & Barnett (2019)). This suggests that girls may have experienced less stimulating home environments, making them more responsive to the enhanced learning opportunities provided by the program (Garcia, Heckman & Ziff (2018)). Gender differences in early development could also contribute to these findings. Studies have consistently shown advantages for girls in early development (Barbu at al. (2015)). Nores, Bernal & Barnett (2019) – the first study of this program assessing its short-term impacts – reports that girls outperformed boys in expressive vocabulary, language, and motor skills by up to 0.16 SD at baseline. These initial developmental differences may have contributed to higher returns for girls.

We also find suggestive evidence that children with lower baseline cognitive development 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).

In its important to interpret our results in the context of high socio-economic vulnerability, limiting generalizations to low-income households. With this consideration in mind, we find suggestive evidence that the effects on cognitive skills are concentrated on children from less vulnerable households, while the effects on health are stronger for children from more vulnerable families. Lower SES children experienced higher risks such as food fragility, lower parental investments, and higher stunting rates. The concentration of nutritional effects on the most fragile households, coupled with stronger cognitive impacts for less disadvantaged children, suggests that synergies might exist between nutritional status, home environments, and high-quality ECCE in the production of cognitive skills. Our findings suggest that children facing very high levels of nutritional and home fragility may require additional investments to address their nutritional deficits and fully benefit from ECCE programs. It may be that for ECCE interventions to produce benefits on child cognitive skills, nutrition and home support need to be simultaneously addressed.

C. Socio-emotional development

We find that the effects on socio-emotional development are always small in magnitude and never statistically different from zero. These results align with previous evidence from Colombia. Bernal et al. (2019) find no effects of center-based care on socio-emotional development compared to home-based care. Andrew et al. (forthcoming) reports no effects of a teacher training and coaching program on children's socio-emotional development. Bernal & Fernández (2013) find improved interactions with peers but increased disruptive behaviors for children attending home-based care. In a different developing country context, Dean and Jayachandran (2020) also find that attending formal pre-school has no effects on children's socio-emotional skills. Our null result also aligns with findings from the broader ECCE literature, and could be due to lack of sensitivity, biases, or inaccurate measurement of underlying constructs using parental-reported measures (Renk & Phares (2004)).

A null effect could also be interpreted positively in relation to studies on center-based intervention in the U.S., reporting negative effects on socio-emotional development. Center-based care offers children opportunities to develop social skills while increasing the likelihood of disruptive behaviors occurring as children compete for limited resources (Baker, Gruber & Milligan (2008, 2019); Haskins, (1985)). The absence of negative effects of aeioTU suggests a favorable context in this regard.

D. Convergence in cognitive development over time

One perhaps puzzling finding is the convergence of cognitive effects in the last year of the program, even among age-eligible children. We explore several explanations for this result.

First, we examine enrollment patterns in primary education over time by treatment group. Figure 5 plots average enrollment rates in primary school by treatment group over time, by combining information from parental reports and administrative records from the Colombian Ministry of Education. As expected, given children's age, enrollment in elementary school is 0 percent at the beginning of the intervention. This figure gradually rises over time, reaching over 70 percent by 2016. Importantly, children assigned to the treatment group have lower enrollment rates in primary school compared to similarly aged children in the control group (these differences are particularly pronounced among older cohorts, Appendix Figure G4). This is despite the fact that children in the treatment group are on average three months older than children in the control group (Table 3) and suggests that the control group progressed earlier to primary education.[11] In Table 7 we regress primary school enrollment on a treatment indicator and confirm the findings in Figure 5: there is a small difference in primary school enrollment between treatment and control children in Y2 (0.3 percentage points). This difference increases over time, so that in the last study wave the treatment group's enrollment rate is 9.1 percentage points (22 percent) lower than that of the control group (this difference is statistically significant at the 5 percent level). The magnitude of this difference is similar when using administrative enrollment data from the government's data system instead of parental reports (columns 3 and 4 of Table 7), and when considering the full sample irrespectively of age-eligibility (Appendix Table G9).[12]

While the random assignment to the treatment group appears to relate to a later transition into primary school, this is not enough to explain the convergence in cognitive skills between treatment and control children. For such convergence to occur, children in primary education would need to be learning more of the measured cognitive skills, possibly through exposure to more advanced learning materials or through higher-achieving peers (Berlinski, Busso & Giannola, 2023; Sacerdote (2011)). We examine this hypothesis by comparing the effectiveness of aeioTU *vis-à-vis* alternative care arrangements: being cared at home, attending other ECCE programs, and being enrolled in primary school.

Random assignment to the program alone is insufficient to identify these effects, as we have more than one single endogenous variable. We therefore use instrumental variables (IV) techniques. Following the common practice in the literature, we interact experimentally assigned program offer with observed

[11] 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 the fact that parents seem to over-report enrollment into the elementary education system compared with the information collected from the administrative records.

[12] These trends occur within a community context where first grade enrollment rates were 53 percent in 2011 and increased to 59 percent in 2014. Official statistics from the government are available here: https://www.datos.gov.co/Educaci-n/MEN_ESTADISTICAS_EN_EDUCACION_EN_PREESCOLAR-B-SICA/nudc-7mev/data.

characteristics (e.g., Abdulkadiroğlu, Angrist & Pathak (2014); Garcia, Heckman & Ziff (2018); Kline & Walters (2016)). We consider alternative sets of instruments and alternative model specifications (additional details on the IV approach are reported in Appendix F).

In Table 8, we report both OLS and IV results for the effects of alternative care arrangements on children's cognitive development (the excluded category is being cared for at home). Column (1) shows that there is a positive association between primary school enrolment or enrollment in aeioTU and children's cognitive skills. At the same time, attending other childcare services is negatively associated with child cognitive development. When treating these childcare choices as endogenous, the coefficient on aeioTU becomes larger in magnitude and the coefficient on alternative childcare arrangements becomes more negative (column 2). Therefore, according to both the OLS and IV estimates, aeioTU is more effective than alternative childcare programs and home care in improving children's cognitive development. But importantly, primary school enrollment has a larger effect on child cognitive development compared to aeioTU, with effects as large as 0.93 SD (column 3). These results are robust across model specifications and for different sets of instruments (columns 2 to 7). Taken together the results in Table 7 and 8 support the hypothesis that the later transition of treated children in primary education could explain the convergence in cognitive skills that we observe in the last study wave.

One disadvantage of the IV 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 and would include over 10 endogenous regressors). To look at the dynamic of the effects, we therefore analyze compliance and the counterfactual experiences of children in the control group. In Appendix Table G10, we report OLS comparisons between children assigned to the treatment group and different subgroups of children in the control group, including control children cared for at home, control children attending other ECCE programs, and control children not enrolled in primary school. The first row the table reports the main ITT effects on cognitive development (from Table 4) as a reference. The results in Appendix Table G10 are consistent with the evidence presented in Table 8: there is a positive association between aeioTU and child cognitive skills, relative to all comparison groups and in all study waves including the last.

Finally, as mentioned earlier, aeioTU was part of the national early childhood strategy, which increased the availability of center-based early education slots by 200 percent nationwide between 2011 and 2016 (Bernal et al. (2019)) and included investments in quality. Bernal & Ramírez (2019) evaluate the effects of this expansion between 2011 and 2013, reporting positive and persistent effects on children's vocabulary. Andrew et al. (forthcoming) report positive effect of DCAS of children's cognitive development. While we lack information on access to high quality alternative childcare services in the study's communities, it is plausible that these trends influenced the availability and quality of the counterfactual programs used by children in the control group. The consistent positive impact observed when comparing treated children to control children cared for at home supports this hypothesis.

V. Conclusions

In this paper, we report the effects of a high-quality center-based early education program on the development of disadvantaged children in Colombia, that we evaluate through a five-year randomized control trial. We follow children longitudinally to study the effects of the program on their early childhood developmental trajectories in the years preceding their enrolment in primary education. The program provides full day high quality ECCE for children throughout the year and includes a nutritional component. It emphasizes structural and process quality, including project and play-based learning, meaningful adult-child interactions, and parental engagement. We take a holistic view of child development and assess the effects of the program along several domains including children' cognitive and socio-emotional skills, and their health.

We find positive effects of the program on children's cognitive development and health, and no impacts on children's socioemotional development or on their home environment. Following children over time, we then document the longitudinal effects of the intervention, finding remarkably stable effects on health. On the other hand, the impacts on children's cognitive development are mitigated in the fifth year of the program, even among those children who are still enrolled in preschool centers. We investigate possible mechanisms for this, including transitions to primary education, the crossover of control children in preschool centers, and changes in counterfactual over time.

Our findings suggest that improving ECCE quality can significantly affect child development and underscore the potential for integrated early education programs to support disadvantaged children's health in low- and middle-income contexts. The growing body of studies in Colombia, including our own, emphasize the need to further understand and measure quality in early childhood care and education, and the value of further investigating how to better support (and possibly better measure) children's socioemotional development. These studies also emphasize the importance of a cohesive evaluation agenda to understand and build quality in ECCE programming. This accumulation of national evidence is key to inform and support further expansion of ECCE and for the prioritization of quality investments in the region (Angrist & Pischke (2010). It also highlights the need for creating and sustaining continuous quality monitoring and improvement systems to inform national strategies and priorities.

Understanding this evaluation within the context of the DCAS national strategy is critical to inform public policy and to interpret the results. The multi-year nature of this evaluation and the fact that it was embedded within a larger policy context of expanding early childhood investments expose our study to the challenges that have been highlighted by previous research: the difficulties and ethics of limiting crossover (Deaton (2010)), 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. In line with the arguments by Deaton & Cartwright (2018) and Kline & Walters (2016), 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 Colombia at the time of this study.

Future research shall investigate whether the effects we find are sustained in the long-run, which will be key to compute the life-cycle benefits of high-quality ECCE program for disadvantaged children.

REFERENCES

Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. Econometrica, 82(1), 137-196.

Aboud, Frances E. "Evaluation of an early childhood preschool program in rural Bangladesh." Early Childhood Research Quarterly 21, no. 1 (2006): 46-60.

Almond, Douglas, Janet Currie, and Valentina Duque. "Childhood circumstances and adult outcomes: Act II." Journal of Economic Literature 56, no. 4 (2018): 1360-1446.

Andrew, Allison, Orazio Attanasio, Raquel Bernal, Lina Cardona, Sonya Krutikova, and Marta Rubio-Codina. "Preschool quality and child development." Journal of Political Economy, Forthcoming.

Andrew, A., O. Attanasio, R. Bernal, L. Cardona, S. Krutikova, D. M. Heredia, C. Medina, X. Pena, M. Rubio Codina, and M. Vera-Hernandez (2018). Evaluation of infant development centres: an early years intervention in Colombia. Technical Report June, International Initiative for Impact Evaluation, USA.

Angrist, Joshua D., and Jörn-Steffen Pischke. "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics." Journal of economic perspectives 24, no. 2 (2010): 3-30.

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

Attanasio, Orazio P., Vincenzo Di Maro, and Marcos Vera-Hernández. "Community Nurseries and the Nutritional Status of Poor Children." Evidence from Colombia. Econ Journal 123, no. 571 (2013): 1025-1058.

Attanasio, O. P., Fernández, C., Fitzsimons, E. O., Grantham-McGregor, S. M., Meghir, C., & Rubio-Codina, M. (2014). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial. Bmj, 349.

Attanasio, Orazio, Helen Baker-Henningham, Raquel Bernal, Costas Meghir, Diana Pineda, and Marta Rubio-Codina. "Early Stimulation and Nutrition: The Impacts of a Scalable Intervention." NBER Working Paper 25059, (2018).

Attanasio, O., de Barros, R. P., Carneiro, P., Evans, D. K., Lima, L., Olinto, P., & Schady, N. (2022). Public childcare, labor market outcomes of caregivers, and child development: Experimental evidence from Brazil (No. w30653). National Bureau of Economic Research.

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

Bailey, Martha J., Shuqiao Sun, and Brenden Timpe. "Prep School for poor kids: The long-run impacts of Head Start on Human capital and economic self-sufficiency." American Economic Review 111, no. 12 (2021): 3963-4001.

Baker, Michael, Jonathan Gruber, and Kevin Milligan. "The long-run impacts of a universal childcare program." American Economic Journal: Economic Policy 11, no. 3 (2019): 1-26.

Baker, Michael., Jonathan Gruber, and Keving Milligan. "Universal childcare, maternal labor supply, and family well-being." Journal of Political Economy 116, no. 4 (2008): 709-745.

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

B Banerjee, Abhijit V., and Esther Duflo. "Poor economics: A radical rethinking of the way to fight global poverty." Public Affairs, (2011).

Barbu, Stéphanie, Aurélie Nardy, Jean-Pierre Chevrot, Bahia Guellaï, Ludivine Glas, Jacques Juhel, and Alban Lemasson. "Sex Differences in language Across Early Childhood: Family Socioeconomic Status does not Impact Boys and girls Equally." Frontiers in Psychology 6 (2015).

Barnett, W. Steven. "Effectiveness of early educational intervention." Science 333, no. 6045 (2011): 975-978.

Barnett, W. Steven, and Cynthia E. Lamy. "Achievement Gaps Start Early. Preschool Can Help." In Closing the Opportunity Gap: What America Must Do to Give Early Child an Even Chance, edited by P. L. Carter and K. G. Welner. OUP USA: Oxford Scholarship Online, (2013).

Barnett, W. Steven and Kwanghee Jung. "Effects of New Jersey's Abbott preschool program on children's achievement, grade retention, and special education through tenth grade." Early childhood research quarterly 56 (2021): 248-259.

Bartlett, M. S. "The statistical conception of mental factors." British Journal of Psychology 28, no. 1 (1937): 97-104.

Barrera-Osorio, F., Bertrand, M., Linden, L. L., & Perez-Calle, F. (2011). Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia. American Economic Journal: Applied Economics, 3(2), 167-195.

Bassok, D., Gibbs, C. R., and Latham, S. "Preschool and children's outcomes in elementary school: Have patterns changed nationwide between 1998 and 2010?" Child Development 90, no. 6 (2019): 1875-1897.

Bastos, P., Bottan, N. L., and Cristia, J. "Access to preprimary education and progression in primary school: Evidence from rural Guatemala." Economic Development and Cultural Change 65, no. 3 (2017): 521-547.

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

Beaton, G., and Ghassemi, H. "Supplementary feeding programs for young children in developing countries." American Journal of Clinical Nutrition 35 (1982): 1177-1198.

Behrman, J., Cheng, Y., and Todd, P. "Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach." Review of Economics and Statistics 86, no. 1 (2004): 108-132.

Berlinski, Samuel, Matias Busso, and Michele Giannola. "Helping struggling students and benefiting all: Peer effects in primary education." Journal of Public Economics 224 (2023): 104925.

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

Berlinski, S., Galiani, S., and Gertler, P. "The effect of pre-primary education on primary school performance." Journal of Public Economics 93, no. 1-2 (2009): 219-234.

Bernal, R. "The impact of a vocational education program for childcare providers on children's well-being." Economics of Education Review 48 (2015): 165-183.

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

Bernal, R., & Ramírez, S. M. Improving the quality of early childhood care at scale: The effects of "From Zero to Forever". World Development, 118 (2019): 91-105.

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

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

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

Biroli, P., D. Del Boca, J. J. Heckman, L. P. Heckman, Y. K. Koh, S. Kuperman, et al. "Evaluation of the Reggio Approach to Early Education." Research in Economics 72, no. 1 (2018): 1-32.

Black, M. M., & Dewey, K. G. "Promoting equity through integrated early child development and nutrition interventions." Annals of the New York Academy of Sciences, 1308(1) (2014): 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. "Early Childhood Development Coming of Age: Science through the Life Course." The Lancet, 389(10064) (2016): 77-90.

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

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

Brinkman, S. A., A. Hasan, H. Jung, A. Kinnell, N. Nakajima, and M. Pradhan. "The Role of Preschool Quality in Promoting Child Development: Evidence from Rural Indonesia." Washington, D.C.: The World Bank, (2016).

Britto, P. R., S. Lye, K. Proulx, A. Yousafzai, S. Matthews, T. Vaivada, et al. "Nurturing Care: Promoting Early Childhood Development." The Lancet 389 (2017): 91-102.

Camilli, G., S. Vargas, S. Ryan, and W. S. Barnett. "Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development." Teach Coll Rec 112, no. 3 (2010): 579-620.

Carrell, Scott E., Mark Hoekstra, and Elira Kuka. "The long-run effects of disruptive peers." *American Economic Review* 108.11 (2018): 3377-3415.

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

Comisión Inter-Sectorial para la Primera Infancia CIPI. "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 (2013).

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg. "Who Benefits from Universal Childcare? Estimating Marginal Returns to Early Childcare Attendance." Journal of Political Economy 126, no. 6 (2018): 2356-2409.

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

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. School inputs, household substitution, and test scores. American Economic Journal: Applied Economics, 5(2), (2013): 29-57.

Dean, J. T., & Jayachandran, S. (2020). Attending kindergarten improves cognitive development in India, but all kindergartens are not equal. University of Chicago Booth School of Business.

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

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

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

De Chaisemartin, C., & Behaghel, L. "Estimating the effect of treatments allocated by randomized waiting lists." Econometrica 88.4 (2020): 1453-1477.

Dietrichson, Jens, Ida Lykke Kristiansen, and Bjørn A. Viinholt. "Universal preschool programs and long-term child outcomes: A systematic review." Journal of Economic Surveys 34.5 (2020): 1007-1043.

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

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

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

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

Duncan, G. J., Leak, J. A., Li, W., Magnuson, K., Schindler, H., & Yoshikawa, H. "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 (2011).

Elango, S., J. L. García, J. J. Heckman, and A. Hojman. "Early Childhood Education." In Economics of Means-Tested Transfer Programs in the United States, Volume 2, 235-297. University of Chicago Press, (2015).
Encuesta Longitudinal Colombiana. "Encuesta Longitudinal Colombiana (ELCA)." Bogotá, Colombia: Universidad de los Andes (2013). Available at https://encuestalongitudinal.uniandes.edu.co/en/.
Englund, M. M., B. White, A. J. Reynolds, L. Schweinhart, and F. A. Campbell. "Health Outcomes of the Abecedarian, Child-Parent Center, and HighScope Perry Preschool Programs." In Health and Education in Early Childhood: Predictors, Interventions and Policies, 257-285, (2014).
Feinstein, L. "Inequality in the Early Cognitive Development of British Children in the 1970 Cohort." Economica 70, no. 277 (2003): 73-97.
Felfe, C., and R. Lalive. "Does Early Childcare Affect Children's Development?" Journal of Public Economics 159 (2018): 33-53.
Fernald, L., E. Prado, P. Kariger, and A. Raikes. "A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries." Washington, D.C.: The World Bank, (2017).
Fernald, L., P. Kariger, M. Hidrobo, and P. Gertler. "Socioeconomic Gradients in Child Development in Very Young Children: Evidence from India, Indonesia, Peru, and Senegal." Proceedings of the National Academy of Sciences 109, no. 2 (2012): 17273-17280.
Fort, M., A. Ichino, and G. Zanella. "The Cognitive and Noncognitive Costs of Daycare at Age 0-2 for Children in Advantaged Families." Journal of Political Economy 128.1 (2020): 158-205.
Friedman-Krauss, A. H., W. S. Barnett, K. A. Garver, K. S. Hodges, G. G. Weisenfeld, and B. A. Gardiner. "The State of Preschool 2019: State Preschool Yearbook." New Brunswick, NJ: National Institute for Early Education Research, (2020).
Garcia, J. L., J. J. Heckman, and A. L. Ziff. "Gender Differences in the Benefits of an Influential Early Childhood Program." European Economic Review 109 (2018): 9-22.
Giannola, Michele. Parental investments and intra-household inequality in child human capital: Evidence from a lab-in-the-field experiment. The Economic Journal, (forthcoming).
Gorsuch, R. "Factor Analysis." Lawrence Erlbaum Associates, Hillsdale, NJ, (1983).
Gray-Lobe, Guthrie, Parag A. Pathak, and Christopher R. Walters. "The long-term effects of universal preschool in Boston." The Quarterly Journal of Economics 138, no. 1 (2023): 363-411.
Gupta, N. D., and M. Simonsen. "Non-cognitive Child Outcomes and Universal High-Quality Childcare." Journal of Public Economics 94, no. 1-2 (2010): 30-43.
Haskins, R. "Public School Aggression Among Children with Varying Day-care Experience." Child Development 56 (1985): 689-703.
Havnes, T., and M. Mogstad. "Is Universal Childcare Leveling the Playing Field?" Journal of Public Economics 127 (2015): 100-114.
Heckman, J. J. "Schools, Skills and Synapses." Economic Inquiry 46, no. 3 (2008): 289-324.
Heckman, J. J., R. Pinto, and P. Savelyev. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." American Economic Review 103, no. 6 (2013): 2052-2086.
Hojman, Andrés, and Florencia Lopez Boo. "Public childcare benefits children and mothers: Evidence from a nationwide experiment in a developing country." *Journal of Public Economics* 212 (2022): 104686.
Imbens, Guido W., and Joshua D. Angrist. "Identification and Estimation of Local Average Treatment Effects." Econometrica 62, no. 2 (1994): 467-475.
Jacoby, H. "Is There an Intrahousehold 'Flypaper Effect'? Evidence from a School Feeding Programme." The Economic Journal 112 (2002): 196-221.

Jenkins, J. M., T. W. Watts, K. Magnuson, E. T. Gershoff, D. H. Clements, J. Sarama, and G. J. Duncan. "Do High-Quality Kindergarten and First-Grade Classrooms Mitigate Preschool Fadeout?" Journal of Research on Educational Effectiveness 11, no. 3 (2018): 339-374.
Jensen, H., A. Pyle, J. M. Zosh, H. B. Ebrahim, A. Zaragoza Scherman, J. Reunamo, and B. K. Hamre. "Play Facilitation: The Science Behind the Art of Engaging Young Children." White paper, Denmark: The LEGO Foundation, (2019). https://www.legofoundation.com/media/1681/play-facilitation_the-science-behind-the-art-ofengaging-young-children.pdf.
Joo, Y. S., K. Magnuson, G. J. Duncan, H. S. Schindler, H. Yoshikawa, and Kathleen M. Ziol-Guest. "What works in early childhood education programs?: A meta–analysis of preschool enhancement programs." Early Education and Development 31, no. 1 (2020): 1-26.
Keller, K., S. Kling, B. Fuchs, A. Pearce, N. Reigh, T. Masterson, and K. Hickok. "A Biopsychosocial Model of Sex Differences in Children's Eating Behaviors." Nutrients 11 (2019): 1-21.
Kline, P., & Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of Head Start. The Quarterly Journal of Economics, 131(4), 1795-1848.
Kottelenberg, M. J., and S. F. Lehrer. "Targeted or Universal Coverage? Assessing Heterogeneity in the Effects of Universal Childcare." Journal of Labor Economics 35, no. 3 (2017): 609-653.
Leamer, E. E. "Tantalus on the Road to Asymptopia." Journal of Economic Perspectives 24, no. 2 (2010): 31-46.
Little, T. D. "Longitudinal Structural Equation Modeling." New York, NY: Guilford, (2013).
López Bóo, F., M. Dormal, and A. Weber. "Validity of Four Measures of Childcare Quality in a National Sample of Centers in Ecuador." PloS One 14, no. 2 (2019).
Magnuson, K., Kelchen, R., Duncan, G., Schindler, H., Shager, H., and Yoshikawa, H. "Do the Effects of Early Childhood Education Programs Differ by Gender? A Meta-Analysis." Early Childhood Research Quarterly 36, no. 3 (2016): 521-536.
Malaguzzi, Loris. "For an education based on relationships." Young children 49.1 (1993): 9-12.
Martinez, S., Naudeau, S., and Pereira, V. "Preschool and Child Development under Extreme Poverty: Evidence from a Randomized Experiment in Rural Mozambique." Washington, D.C.: The World Bank, (2017).
McCoy, D. C., Gonzalez, K., and Jones, S. "Preschool Self-Regulation and Preacademic Skills as Mediators of the Long-Term Impacts of an Early Intervention." Child Development 90, no. 5 (2019): 1544-1558.
Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., and Mather, N. "Batería III Pruebas de Aprovechamiento." Itasca, IL: Riverside Publishing, (2005).
Noboa-Hidalgo, G. E., and Urzua, S. "The Effects of Participation in Public Childcare Centers: Evidence from Chile." Journal of Human Capital 6, no. 1 (2012): 1-34.
Nores, M., and Barnett, W. S. "Benefits of Early Childhood Interventions Across the World: (Under) Investing in the Very Young." Economics of Education Review 29, no. 2 (2010): 271-282.
Nores, M., Bernal, R., and Barnett, W. S. "Center-Based Care for Infants and Toddlers: The aeioTU Randomized Trial." Economics of Education Review 72 (2019): 30-43.
Nores, M., Figueras-Daniel, A., López, M. A., and Bernal, R. "Implementing aeioTU: Quality Improvement Alongside an Efficacy Study. Learning while Growing." Annals of the New York Academy of Sciences 1419, no. 1 (2018): 201-217.
Nores, M., and Prayag, R. "Early Childhood Education Effects into Adulthood." In SAGE Handbook on the Sociology of Education, edited by M. Berends, B. Schneider, and S. Lamb, forthcoming. Sage.
Ogunlade, A. O., Kruger, H. S., Jerling, J. C., Smuts, C. M., Covic, N., Hanekom, S. M., ... & Kvalsvig, J. (2011). Point-of-use micronutrient fortification: lessons learned in implementing a

preschool-based pilot trial in South Africa. International journal of food sciences and nutrition, 62(1), 1-16.
Özler, B., Fernald, L. C., Kariger, P., McConnell, C., Neuman, M., & Fraga, E. "Combining pre-school teacher training with parenting education: A cluster-randomized controlled trial". Journal of Development Economics, 133, (2018): 448-467.
Padilla, E., Lugo, D., and Dunn, L. "Test de Vocabulario en Imágenes Peabody (TVIP)." Circle Pines, MN: American Guidance Service (AGS), Inc., (1986).
Ponitz, C. E., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., and Morrison, F. J. "Touch Your Toes! Developing a Direct Measure of Behavioral Regulation in Early Childhood." Early Childhood Research Quarterly 23, no. 2 (2008): 141-15.
Ramey, C. T., Campbell, F. A., Burchinal, M., Skinner, M. L., Gardner, D. M., and Ramey, S. L. "Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers." Applied Developmental Science 4, no. 1 (2000): 2-14.
Ramey, S. L., & Ramey, C. T. (2023). Early Childhood Education that Promotes Lifelong Learning, Health, and Social Well-being: The Abecedarian Project and its Replications. Medical Research Archives, 11(11).
Renk, K., and Phares, V. "Cross-Informant Ratings of Social Competence in Children and Adolescents." Clinical Psychology Review 24, no. 2 (2004): 239-254.
Reynolds, A. J., Ou, S. R., Mondi, C. F., and Hayakawa, M. "Processes of Early Childhood Interventions to Adult Well-Being." Child Development 88, no. 2 (2017): 378-387.
Meisel-Roca, A., and Ricciuli-Marín, D. (2018). La pobreza en Santa Marta: Los Estragos del Bien. Documento de Trabajo sobre Economía Regional y Urbana; No. 266.
Romano, J., and Wolf, M. "Stepwise Multiple Testing as Formalized Data Snooping." Econometrica 73, no. 4 (2005): 1237-1282.
Rubio-Codina, M., Attanasio, O., Meghir, C., Varela, N., and Grantham-McGregor, S. "The Socio-Economic Gradient of Child Development: Cross-Sectional Evidence from Children 6-42 Months in Bogota." Journal of Human Resources 50, no. 2 (2015): 464-483.
Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far?. In Handbook of the Economics of Education (Vol. 3, pp. 249-277). Elsevier.
Simeon, D. T. (1998). School feeding in Jamaica: a review of its evaluation. The American journal of clinical nutrition, 67(4), 790S-794S.
Singer, D., Golinkoff, R. M., & Hirsh-Pasek, K. "Play= Learning: How play motivates and enhances children's cognitive and social-emotional growth." Oxford University Press, (2006).
Sparrow, S. S., Balla, D. A., Cicchetti, D. V., & Harrison, P. L. "Vineland adaptative behavior scales: classroom edition." American Guidance Service, (1985).
Squires, J., Bricker, D., & Twombly, E. "Technical Report on ASQ:SE." Baltimore, Co: Paul H. Brookes Publishing, (2009).
Sudfeld, C. R., Charles McCoy, D., Danaei, G., Fink, G., Ezzati, M., Andrews, K. G., & Fawzi, W. W. (2015). Linear growth and child development in low-and middle-income countries: a meta-analysis. Pediatrics, 135(5), e1266-e1275.
Tanner, J., Candland, T., & Odden, W. "Later impacts of early childhood interventions: a systematic review." IEG Working Paper, 2015/3. World Bank Group.
United Nations. "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/, (2016).

United Nations Children's Fund (UNICEF). Early Childhood Development. UNICEF Vision for Every Child. UNICEF, New York, 2023.

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

Wolf, S. "Year 3 Follow-up of the 'Quality Preschool for Ghana' Interventions on Child Development." Accepted in Developmental Psychology, (2019).

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

Word Health Organization. "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, (2007).

World Health Organization. "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, (2006).

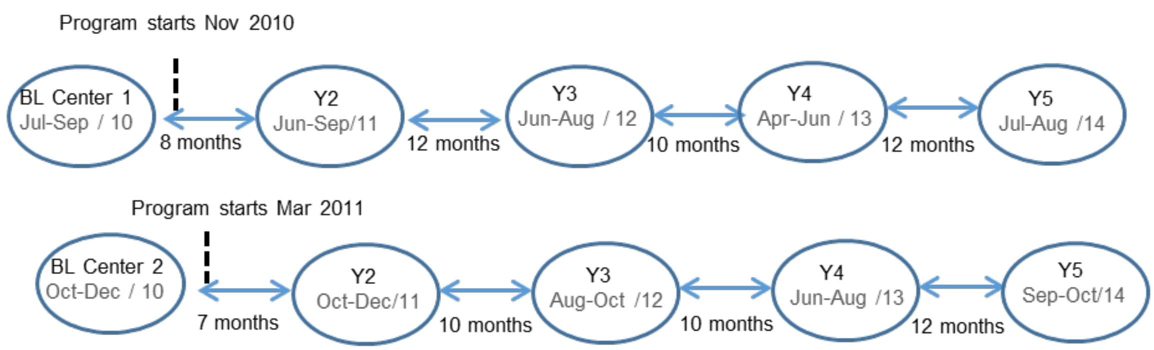
World Health Organization. "WHO guideline on the prevention and management of wasting and nutritional oedema (acute malnutrition) in infants and children under 5 years". (2023). Available at: https://files.magicapp.org/guideline/a3fe934f-6516-460d-902f-e1c7bbcec034/published_guideline_7330-1_1.pdf (last accessed December 2023)

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

Yoshikawa, H., Wuermli, A. J., Raikes, A., Kim, S., & Kabay, S. B. "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, (2018).

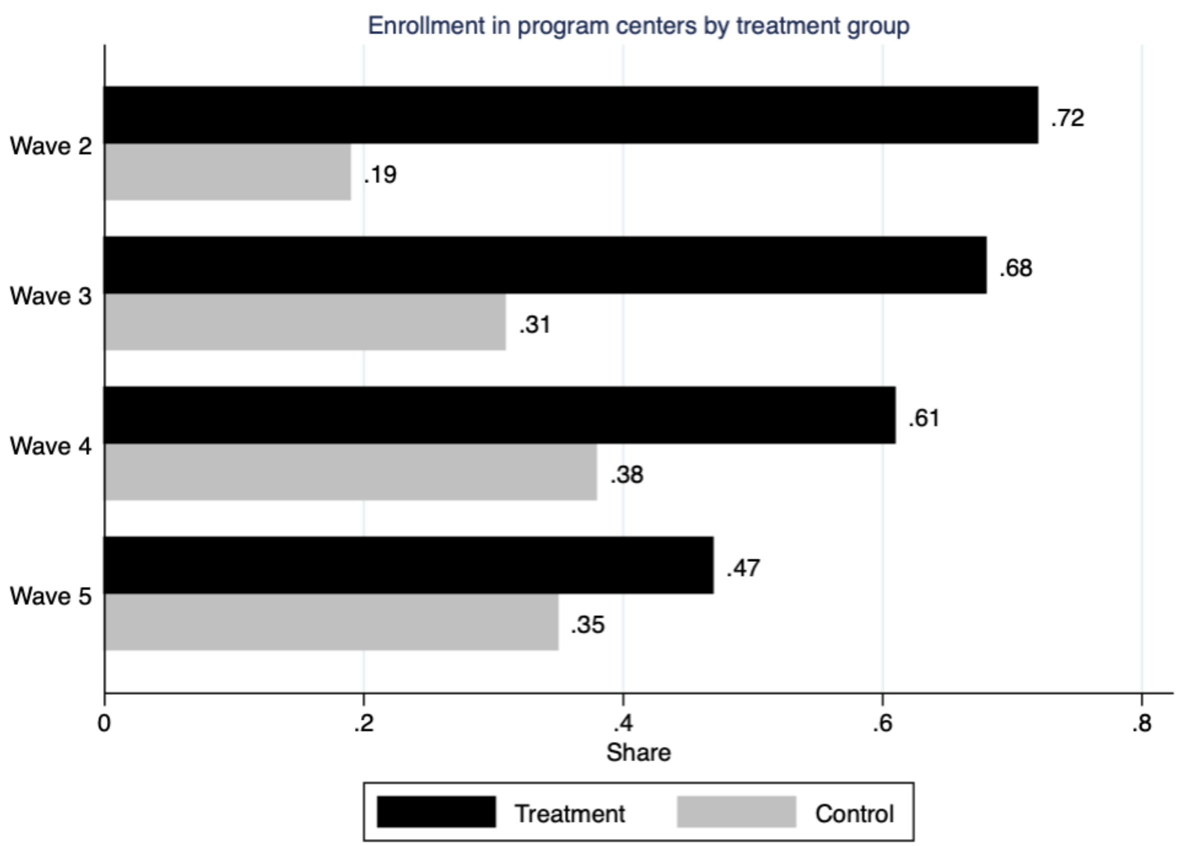
Tables & Figures

Figure 1. Study timeline



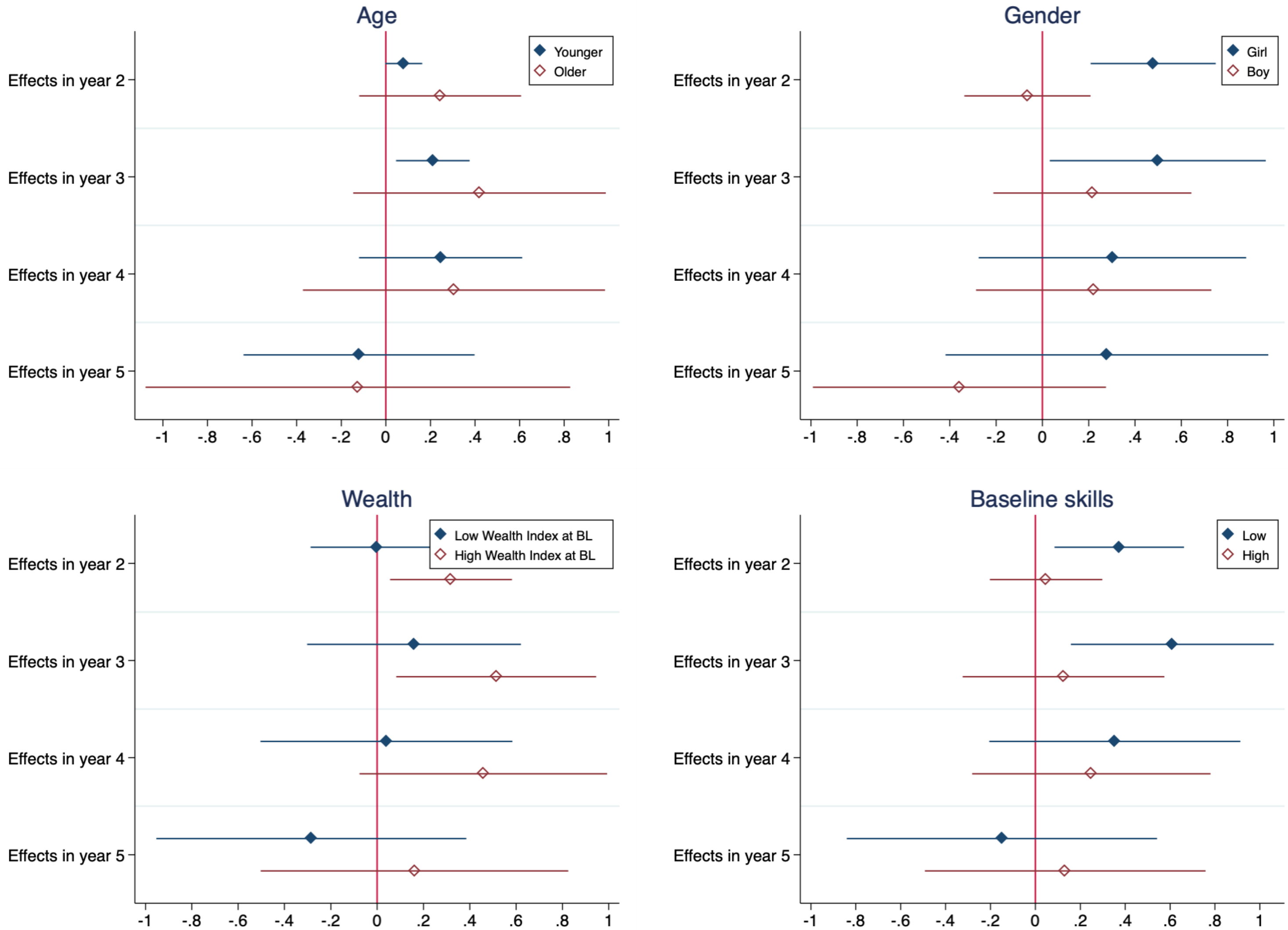
Notes: This figure plots the study timeline, including yearly assessments.

Figure 2: Enrollment by study wave and treatment group



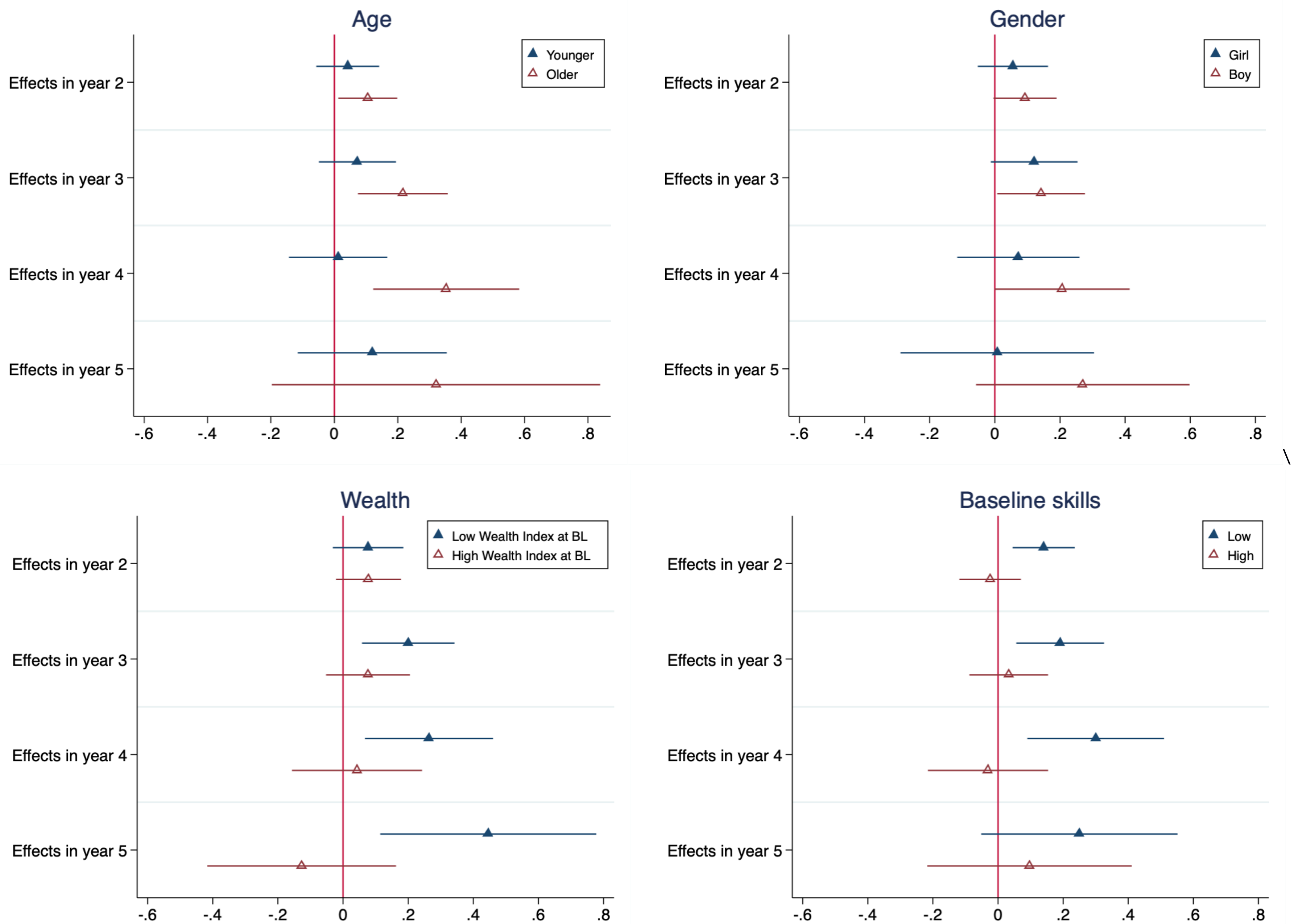
Notes: This figure plots for each study wave the share of children who were initially assigned to the treatment group and effectively enrolled in aeioTU, and the share of children who were initially assigned to the control group and were enrolled in aeioTU. That is, it depicts the share of the children 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. The sample is restricted to age eligible children as defined in the main text.

Figure 3: Heterogeneous effects on child cognitive skills



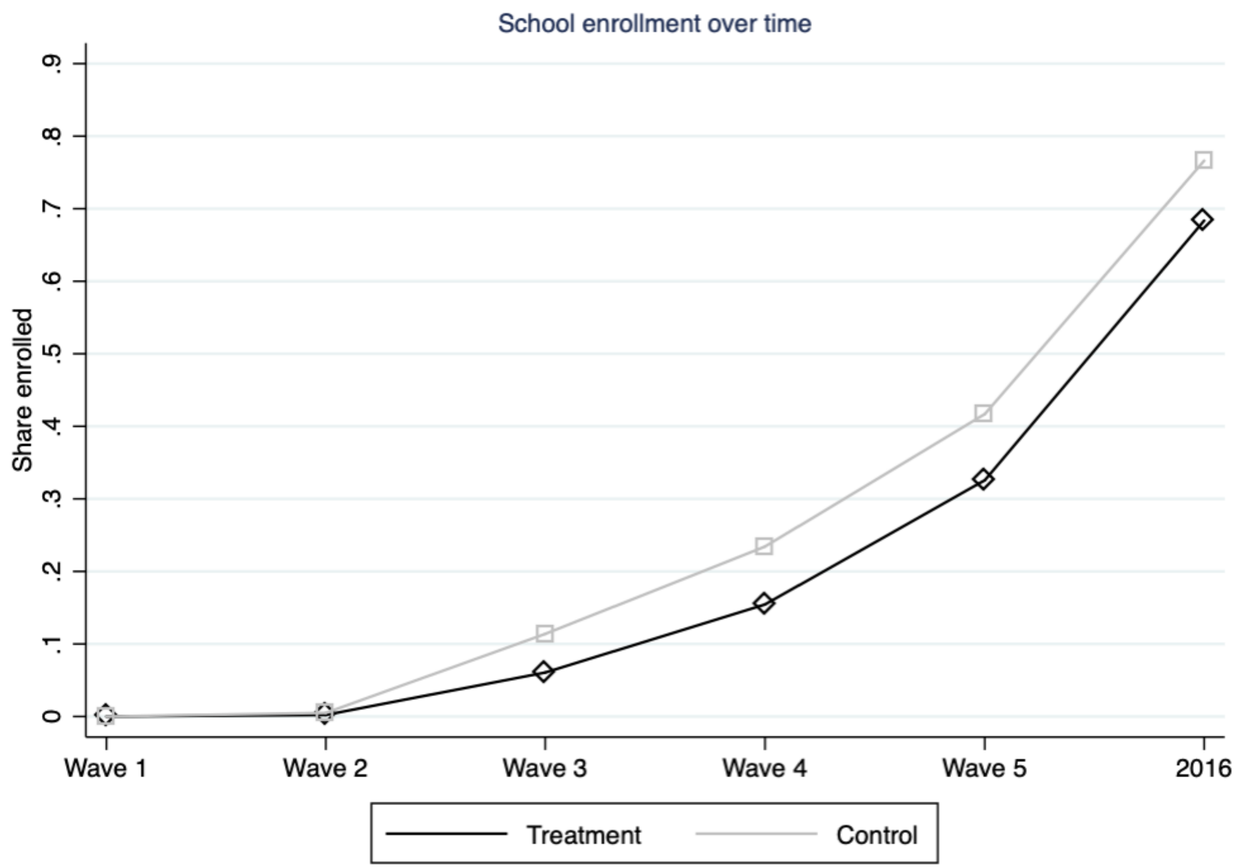
Notes: This figure plots treatment effects by subgroup. The top left panel compares younger (0-2) and older children (2-4). The top right panel compares boys and girls. The bottom left panel compares higher and lower SES children (based on whether household wealth is above or below the median in the 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 4: 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 5: Primary school enrollment over time by treatment group



Notes: This figure plots the share of children enrolled in the primary 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
Full sample	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	0+	0+	0+	0+	0+
Height					
Weight					
Arm circumference					
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 Strengths and Difficulties Questionnaire for socio-emotional development.

Table 3: Baseline characteristics of children and families

		All		Treatment		Control		
	Observations	Mean	SD	Mean	SD	Mean	SD	P-value
Panel A: Demographics								
Age in months	1,073	25.332	12.470	26.919	13.082	24.091	11.833	0.005
Gender (male)	1,073	0.523	0.500	0.522	0.500	0.523	0.500	1.000
Race (black)	1,073	0.615	0.487	0.631	0.483	0.603	0.490	0.975
Mother is single	1,073	0.276	0.447	0.268	0.443	0.282	0.451	0.994
Mother works	1,073	0.236	0.425	0.225	0.418	0.244	0.430	0.986
Mother secondary complete +	1,073	0.364	0.481	0.363	0.481	0.365	0.482	1.000
Childcare use at baseline	1,073	0.171	0.377	0.206	0.405	0.145	0.352	0.093
Father lives at home	1,069	0.688	0.463	0.685	0.465	0.691	0.462	1.000
Father secondary complete +	1,007	0.399	0.490	0.411	0.493	0.390	0.488	0.990
Wealth index	1,073	0.110	4.875	-0.094	4.772	0.269	4.952	0.918
Children books at home	1,072	1.433	2.575	1.477	2.954	1.399	2.237	0.994
Household size	1,073	5.355	2.007	5.287	1.898	5.409	2.089	0.975
No. of children <= 5 yrs	1,073	2.692	0.806	2.775	0.838	2.626	0.775	0.037
CCT beneficiary	1,073	0.341	0.474	0.340	0.474	0.342	0.475	1.000
Panel B: Child development								
Cognitive development								
Bayley Cognitive	798	48.455	14.997	49.176	15.550	47.949	14.592	0.490
Bayley Receptive	790	19.284	7.838	20.065	8.087	18.738	7.621	0.100
Bayley Expressive	794	19.618	9.326	20.214	9.988	19.201	8.819	0.419
Bayley Fine Motor	793	32.439	9.624	32.923	9.994	32.101	9.352	0.483
Bayley Gross Motor	797	46.890	13.400	47.288	14.029	46.608	12.946	0.538
TVIP	384	6.995	6.024	7.659	6.626	6.235	5.165	0.100
Woodcock-Munoz: Applied Problems	244	89.148	9.750	88.280	9.117	90.170	10.395	0.419
ELSA: Reading Comprehension	245	2.469	3.087	2.649	3.125	2.263	3.043	0.538
Health								
Weigh-for-age Z-score	1,046	-0.365	1.031	-0.427	0.990	-0.317	1.061	0.208
Height-for-age Z-score	1,048	-1.123	1.080	-1.110	0.995	-1.134	1.143	0.710
Weight-for-height Z-Score	1,038	0.339	0.976	0.268	0.955	0.395	0.989	0.104
Arm circumference	1,061	15.308	1.099	15.357	1.054	15.270	1.133	0.345
Socio-emotional development								
ASQ Total Score	1,060	46.239	29.369	48.953	30.530	44.101	28.263	0.009

Notes: The table reports baseline child and household characteristics by treatment group. Column (8) reports two-sided Romano-Wolf p-values for the difference between treatment and control.

Table 4: ITT effects on child development

	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 5: Effects of program enrollment on child development

	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value	Observations (F-statistic)
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. F-statistic is for the first stage regression of program enrolment on random assignment to treatment.

Table 6: 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 7: Enrollment in the primary education system over time

	Primary Education Enrollment			
	Parental reports		Administrative records (2016)	
	(1)	(2)	(3)	(4)
Treatment effect, Wave 2	-0.003	-0.021		
	(0.424)	(0.011)		
Treatment effect, Wave 3	-0.053	-0.071		
	(0.002)	(0.000)		
Treatment effect, Wave 4	-0.080	-0.091		
	(0.002)	(0.000)		
Treatment effect, 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 primary 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 8: Comparing aeioTU to alternative education services

Child cognitive skills	OLS	IV1		IV2		IV3	
		Levels	Interacted	Levels	Interacted	Levels	Interacted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
School	0.828	0.885	0.931	0.854	0.906	0.834	0.863
	(0.000)	(0.002)	(0.001)	(0.003)	(0.001)	(0.004)	(0.002)
aeioTu	0.145	0.289	0.276	0.303	0.291	0.250	0.229
	(0.029)	(0.040)	(0.047)	(0.033)	(0.039)	(0.084)	(0.110)
Alternative childcare	-0.049	-0.272	-0.277	-0.273	-0.285	-0.232	-0.262
	(0.682)	(0.438)	(0.439)	(0.438)	(0.430)	(0.503)	(0.446)
Observations	3218	3218	3218	3218	3218	3218	3218

Notes: This table presents the results of the control function approach described in the main text. OLS refers to the OLS results. 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 dummy variable for whether the child is enrolled in school, "aeioTU" is a variable that indicates the number of years enrolled in an aeioTU center, and "Alternative childcare" is a dummy variable for whether the child attends other childcare services different from aeioTU. The excluded category is being cared for at home. Standard errors are clustered at the child level. P-values are reported below the point estimate. The sample is restricted to age eligible children as defined in the text.

Appendix for “The Effect of Center-Based Early Education on Disadvantaged Children’s Developmental Trajectories: Experimental Evidence from Colombia”

Raquel Bernal, Michele Giannola and Milagros Nores

Appendix A. Description of child assessments

This section describes the measures used to assess children’s development and the home environment in the study.

Health: As is standard practice in early intervention studies in developing countries, 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). 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, & Dunn 1986). The TVIP is a measure of receptive language and has been used extensively in preschool studies (Early et al. (2007)) and shown sensitivity to early interventions (Leroy et al. (2008)). Receptive language has been shown to be highly predictive of later development (Pianta (2012)).

We also measured child development using the Vineland Adaptive Behavior Scales (Sparrow et al. 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 & 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 et al. (2009)).

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.

Finally, we measured executive function using the Head-Toes-Knees and Shoulders (HTKS), which examines behavioral regulation in children (Ponitz et al. (2008)). 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, & 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.

Home environment: The home environment was assessed through parent surveys on: (1) discipline strategies, (2) nutritional and feeding habits, and (3) parental engagement with children. Discipline strategies are measured with an 8-item scale asking parents to rate the frequency of certain types of discipline strategies, including physical and verbal punishments, as well as positive alternatives. This was adapted from the Fragile Families Study (Westat, 2011) and collected across all waves. Starting in Y2, questions about meal contents were used to construct a measure of a balanced diet (i.e., inclusion of all nutritional elements in each meal or each day), and of food insecurity (i.e. whether the child skipped at least one meal due to lack of resources). Parental engagement with the child was assessed using questions on the number of hours parents devoted to childcare during weekdays and weekends, and the frequency of activities with children such as reading, feeding, playing, and going on walks with them or visiting places. These questions were collected in all waves.

Appendix B. Construction of the latent factors

As described in the paper, 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 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, Pinto & Savelyev (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, Pinto & Savelyev (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, 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 $m_{i,k,t}^{j}$ 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:

$$m_{i,k,t}^{j} = \alpha_{k,t}^{j} + \lambda_{k,t}^{j}\theta_{i,k,t} + \varepsilon_{i,k,t}^{j} \qquad j = 1, \dots, N_{k,t} \tag{1}$$

where the terms $\alpha_{k,t}^{j}$ are intercepts, $\lambda_{k,t}^{j}$ are factor loadings, $\theta_{i,k,t}$ is the latent factor of skill type k, and the terms $\varepsilon_{i,k,t}^{j}$ 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 & 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.[1]

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 & 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.[2]

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. Suppose that the set of available measures available for skill k can be partitioned in two subsets: M_k^A whose generic element is a, and M_k^B whose generic element is b. These two subsets are such that $M_k^A \cap M_k^B = \emptyset$. 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 $M_k^A = \{a_1, a_2, a_3\}$, $M_k^B = \{b_1, b_2, b_3\}$. Assume further that $M_{k,1} = M_k^A$,

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

[2] Agostinelli & 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.

$M_{k,2} = M_k^A \cup M_k^B$ and $M_{k,3} = M_k^B$. Omitting the subscript k to avoid notational clutter, we can write:

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

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

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

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 $\lambda_1^{a_1} = \lambda_2^{a_1} = 1$.[3] 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).[4]

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.[5] For health, we observe the same measures in all periods, thus the

---

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

[4] 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.

[5] 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.

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, Hansen, & Heckman (2003) show that the parameters in (1), the distribution of the latent factors and the distribution of the measurement errors are non-parametrically identified. 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).[6] These factor scores, which summarize the information contained in the different measures are used in the estimation of treatment impacts, but we also report treatment effects on the individual measures as a robustness.

[6] 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 factor

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.

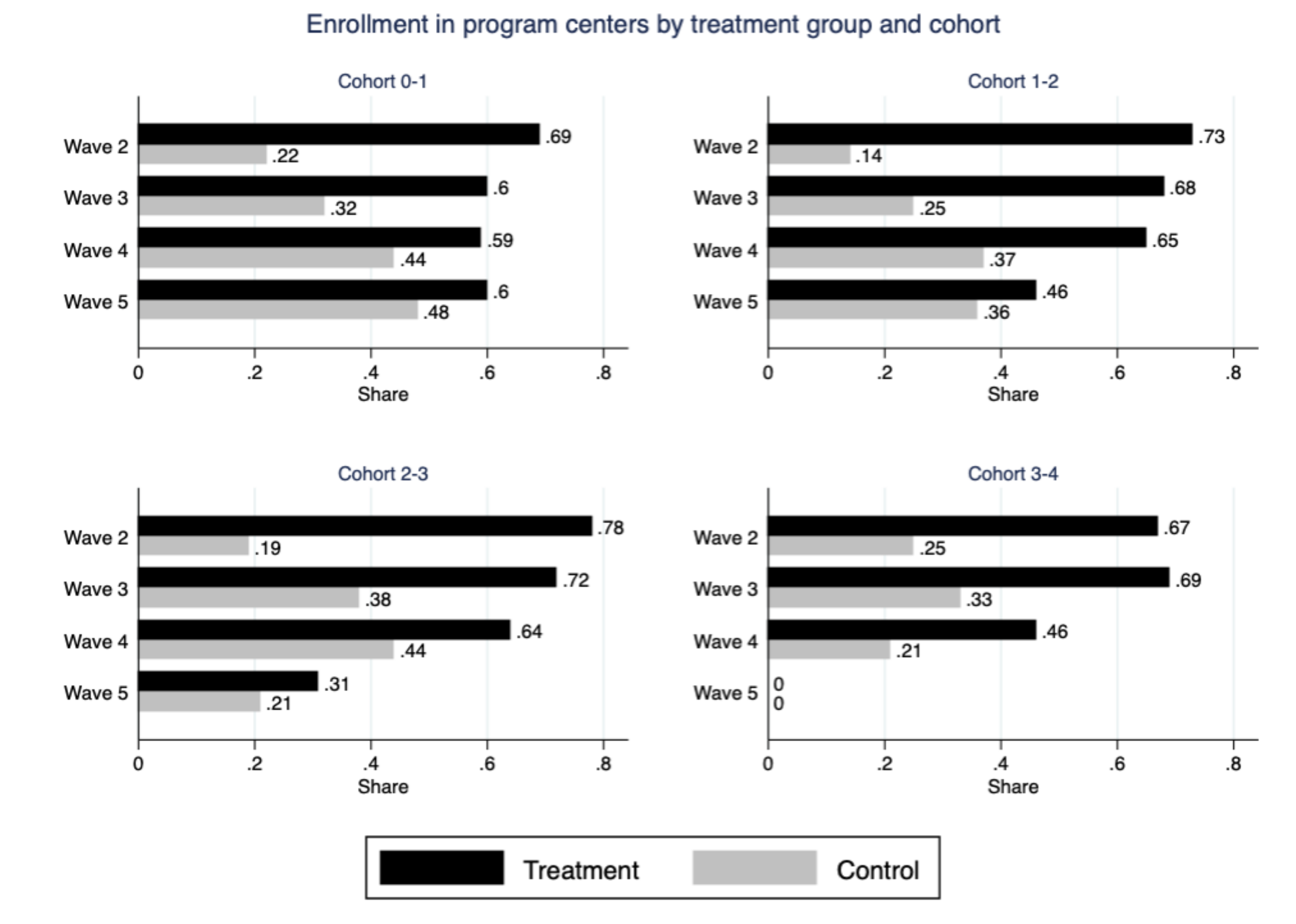
Appendix Table C3: Difference in baseline outcome variables by attrition

	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 enrolled in aeioTU, and the share of children who were initially assigned to the control group and were 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: Impact of lottery on years in aeioTU

	Years enrolled	
	(1)	(3)
Treatment	1.130	1.154
	(0.000)	(0.000)
Mean enrolment control	1.118	1.118
Observations	1073	1073
Controls	N	Y

Notes: The dependent variable is the number of years the child attends aeioTU. 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. All regressions controls for randomization strata. Standard errors are adjusted for clustering at the child level. P-values are reported below the point estimate.

Appendix Table D2: 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.

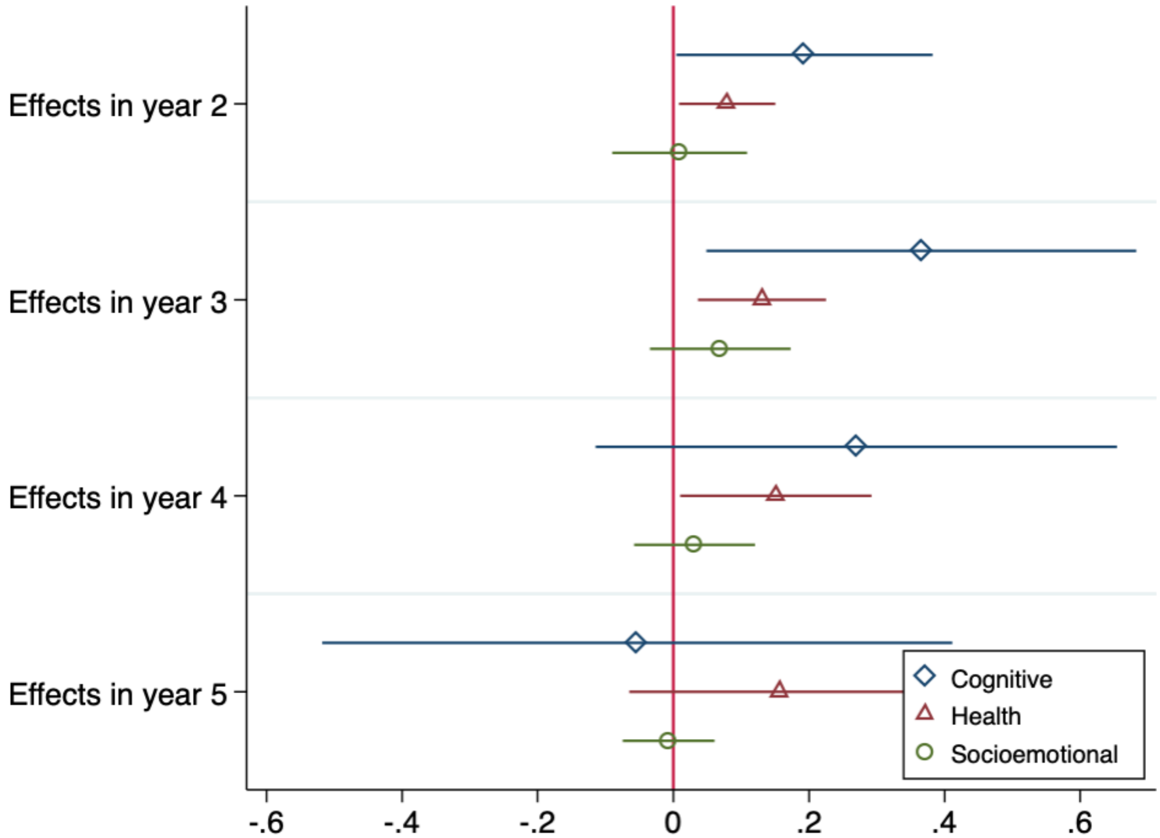
Appendix Table D3: Determinants of enrollment

	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 Figure E1: ITT effects



Notes: This figure reports OLS estimate of equation (1) with the corresponding 95 percent confidence intervals. Effects interpreted in terms of SD in the control group at baseline. 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.

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

	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 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

	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value	Observations
Cognitive	0.169	0.032	0.385	0.010	0.290	0.071	-0.106	0.671	3,418
	(0.091)	(0.091)	(0.166)	(0.043)	(0.198)	(0.133)	(0.240)	(0.691)	
Health	0.063	0.037	0.115	0.007	0.140	0.025	0.139	0.109	3,484
	(0.035)	(0.068)	(0.047)	(0.024)	(0.071)	(0.062)	(0.113)	(0.107)	
Socioemotional	0.056	0.855	0.072	0.914	0.026	0.713	-0.006	0.432	3,490
	(0.053)	(0.978)	(0.053)	(0.978)	(0.046)	(0.975)	(0.035)	(0.891)	

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. Covariates include child baseline score and tester fixed effects. 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

	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 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)

	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 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)

	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 child development without tester effects (full sample)

	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Observations
Cognitive	0.171	0.040	0.363	0.013	0.383	0.020	-0.003	0.506	3,918
	(0.098)	(0.079)	(0.162)	(0.055)	(0.186)	(0.063)	(0.207)	(0.499)	
Health	0.088	0.008	0.146	0.001	0.168	0.006	0.227	0.011	3,992
	(0.036)	(0.014)	(0.048)	(0.004)	(0.067)	(0.014)	(0.098)	(0.014)	
Socioemotional	0.033	0.737	0.044	0.790	0.013	0.599	0.001	0.511	4,001
	(0.052)	(0.940)	(0.054)	(0.940)	(0.051)	(0.940)	(0.024)	(0.940)	

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 E7: ITT effects on child development without controls (full sample)

	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value	Observations
Cognitive	0.169	0.031	0.383	0.010	0.441	0.010	0.055	0.397	3,918
	(0.091)	(0.076)	(0.165)	(0.045)	(0.188)	(0.045)	(0.210)	(0.370)	
Health	0.063	0.037	0.125	0.005	0.152	0.011	0.222	0.013	3,992
	(0.035)	(0.034)	(0.048)	(0.015)	(0.066)	(0.028)	(0.100)	(0.028)	
Socioemotional	0.018	0.645	0.074	0.922	0.032	0.766	0.010	0.673	4,001
	(0.050)	(0.987)	(0.052)	(0.987)	(0.044)	(0.987)	(0.023)	(0.987)	

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. Covariates include child baseline score and tester fixed effects.

Appendix Table E8: 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	0.060	0.043	0.001	0.496	-	-	735
	(0.028)		(0.035)		(0.053)				
Bayley receptive	0.096	0.012	0.152	0.003	0.100	0.148	-	-	726
	(0.042)		(0.055)		(0.096)				
Bayley expressive	0.096	0.015	-0.082	0.873	0.134	0.199	-	-	728
	(0.044)		(0.072)		(0.158)				
Bayley fine motor	0.040	0.072	0.061	0.087	-0.031	0.635	-	-	732
	(0.027)		(0.045)		(0.091)				
Bayley gross motor	0.037	0.078	-0.024	0.735	-0.008	0.546	-	-	726
	(0.026)		(0.038)		(0.066)				
TVIP	0.333	0.009	0.370	0.003	0.196	0.095	0.065	0.376	3,001
	(0.141)		(0.133)		(0.149)		(0.205)		
WCM	-0.141	0.914	0.061	0.243	0.068	0.227	-0.076	0.759	2,725
	(0.103)		(0.087)		(0.090)		(0.108)		
ELSA	0.286	0.001	0.195	0.001	0.061	0.108	-	-	2,087
	(0.087)		(0.064)		(0.049)				
HTKS	-0.004	0.507	-0.022	0.557	-0.038	0.654	-0.106	0.808	2,121
	(0.206)		(0.154)		(0.096)		(0.122)		

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 Table E9: ITT effects on health measures

Measure	Y2 Effect (SE)	P-Value	Y3 Effect (SE)	P-Value	Y4 Effect (SE)	P-Value	Y5 Effect (SE)	P-Value	Observations
Weight for age	0.087	0.008	0.127	0.001	0.124	0.004	0.091	0.086	3,461
	(0.036)		(0.039)		(0.047)		(0.067)		
Height for age	0.157	0.001	0.111	0.009	0.112	0.013	0.109	0.047	3,463
	(0.050)		(0.047)		(0.050)		(0.065)		
Arm circumference	-0.038	0.601	0.037	0.344	-1.828	0.856	0.082	0.275	3,479
	(0.148)		(0.092)		(1.718)		(0.137)		
BMI for age	-0.020	0.617	0.088	0.108	0.055	0.249	-0.025	0.588	3,460
	(0.067)		(0.071)		(0.081)		(0.110)		
Not Stunted	0.032	0.085	0.050	0.009	0.038	0.025	0.043	0.039	3,463
	(0.023)		(0.021)		(0.019)		(0.024)		

Notes: OLS estimate of equation (1). Each row corresponds to a separate regression for the outcome reported in the row header. 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. Following the common practice in the literature, we interact experimentally assigned program offer with observed characteristics (e.g., Abdulkadiroğlu, Angrist & Pathak (2014); Kline & Walters (2016)). Similarly to Garcia, Heckman & Ziff (2018) we further consider combinations of different instruments and alternative specifications. 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 the national Conditional Cash Transfer program Familias en Accion.

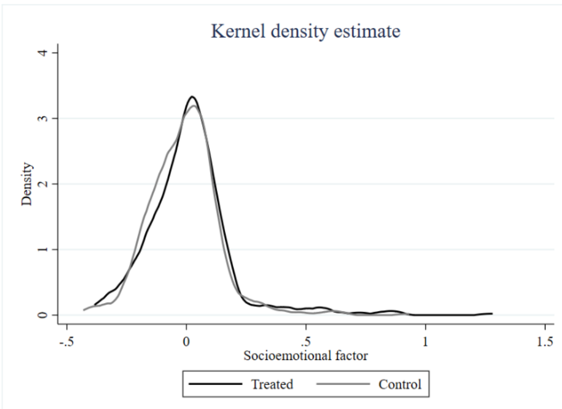
2. IV2: the second set of instruments includes, maternal marital status, an indicator for whether the child's father is a member of the household, father's education, mother's marital status, and an indicator for whether the family receives Familias en Accion.

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.

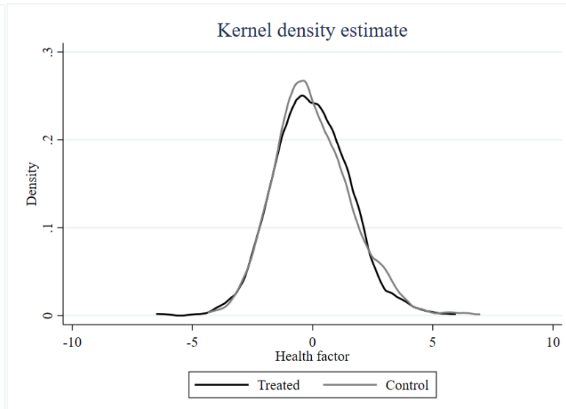
For each of these three sets of instruments we test two distinct specifications. In the first specification we use the instruments measured in levels, while in the second specification we use the instruments measured in levels and interactions of the initial randomization status with the observable characteristics.

Appendix G: Additional Figures and Tables

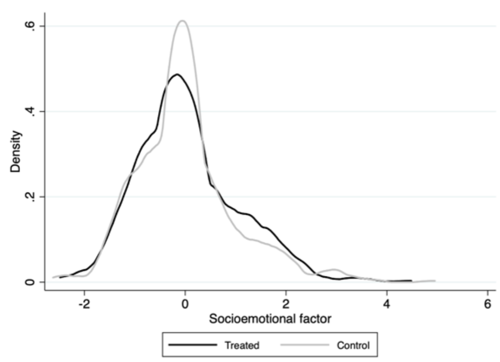
Appendix Figure G1: Distributions of latent factors



Cognitive skills



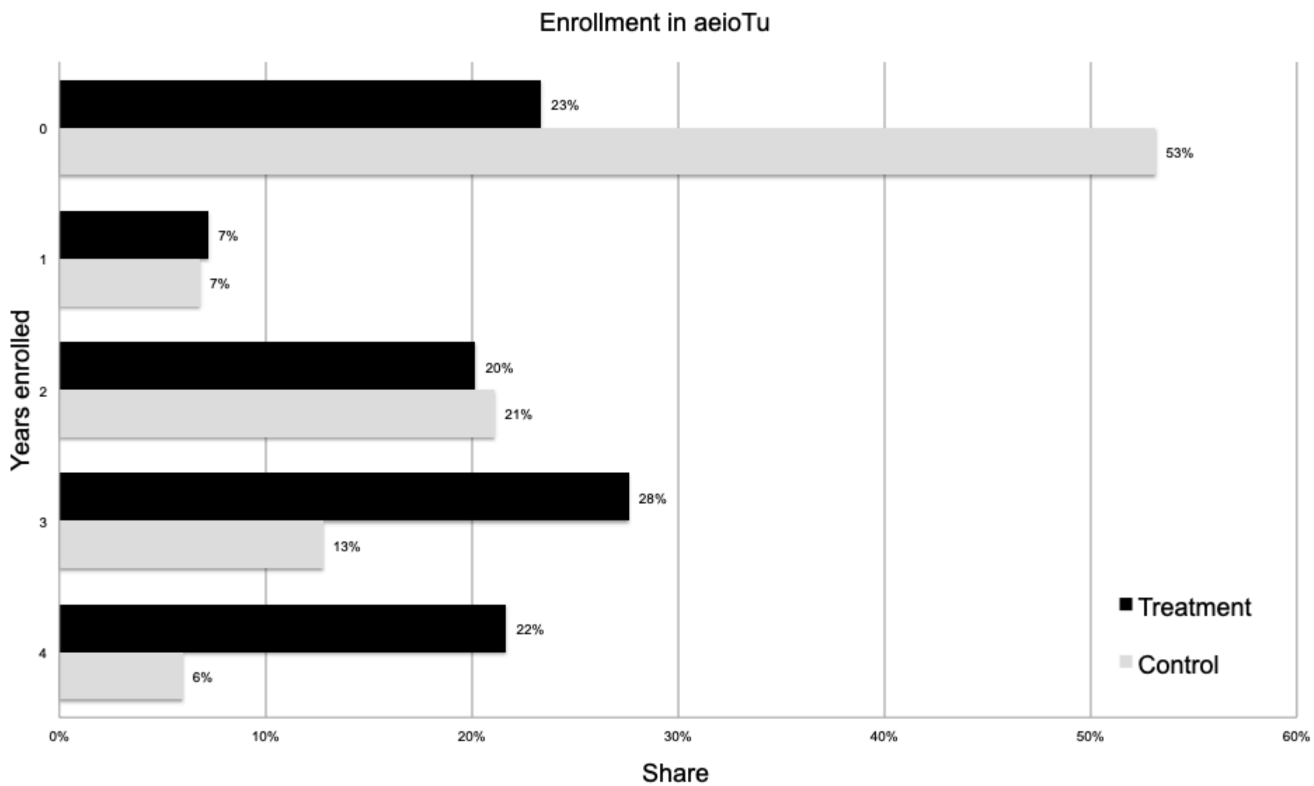
Health



Socio-emotional skills

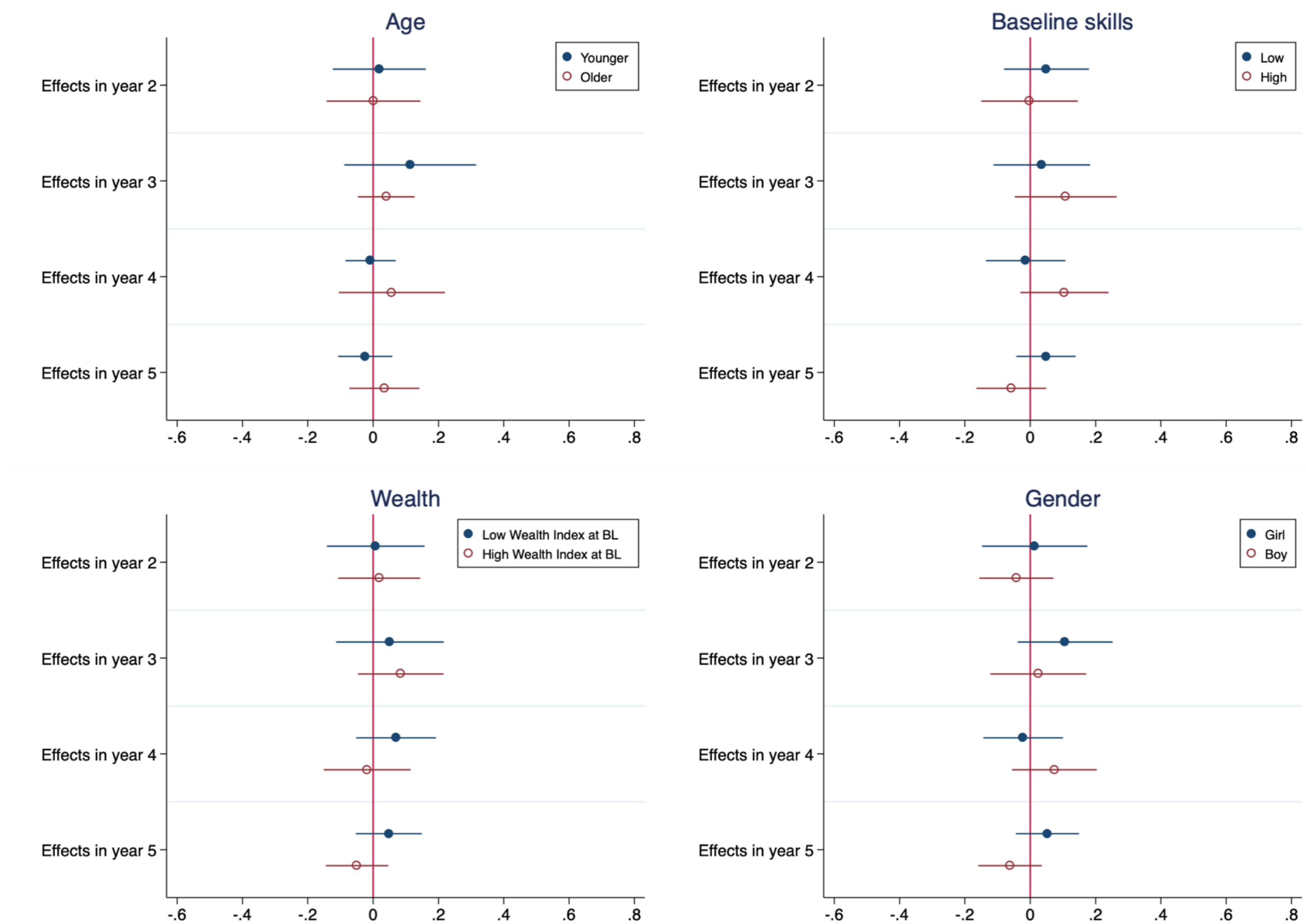
Notes: This figure plots the distributions of latent factors at baseline by treatment group.

Appendix Figure G2: Years of enrollment in aeioTU by treatment group



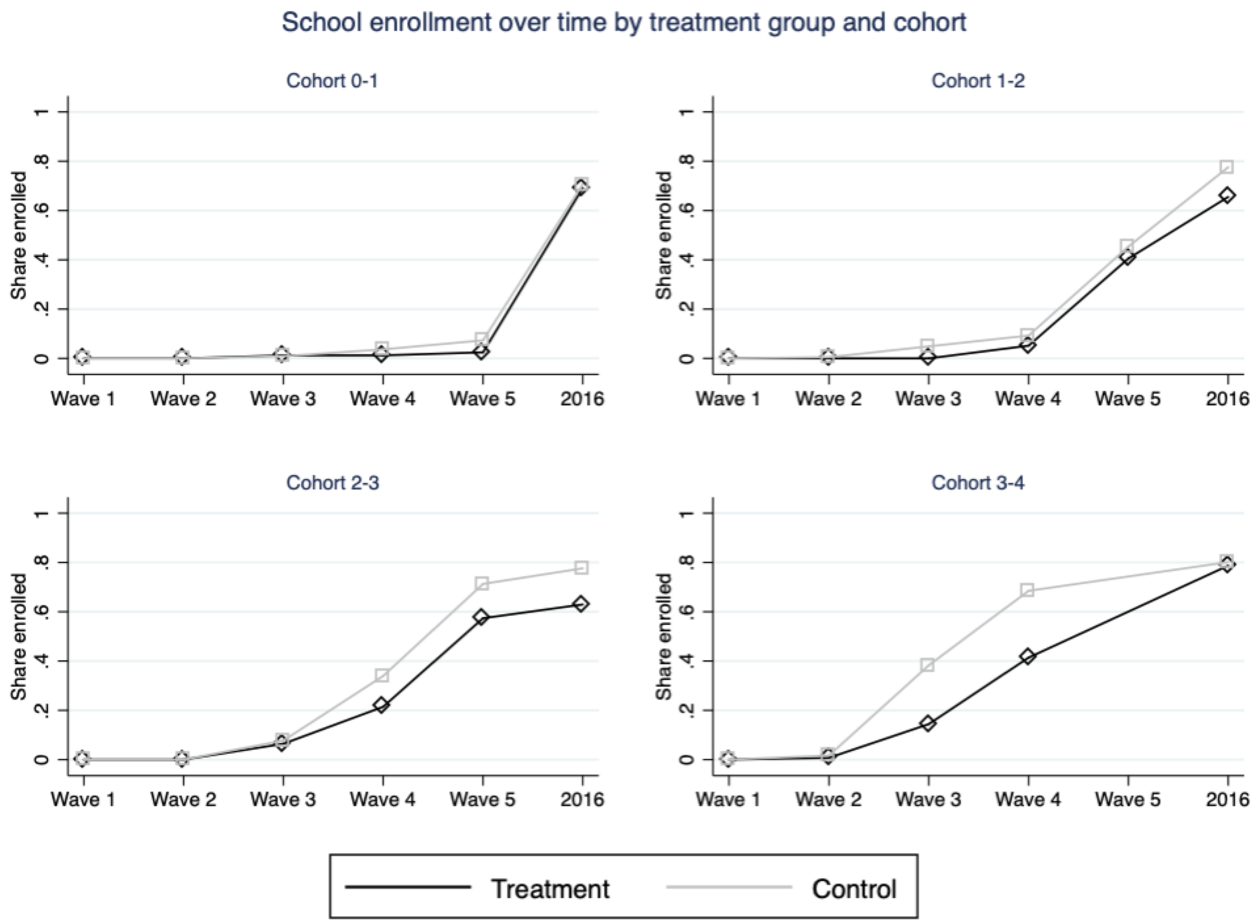
Notes: This figure plots the number of years of enrollment in an aeioTU center by treatment group. Each bar represents the share of children who was enrolled in a center for a given number of years out of the total number of children in the treatment or control group. The sample is restricted to age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

Appendix Figure G3: Heterogeneous effects on child socio-emotional skills



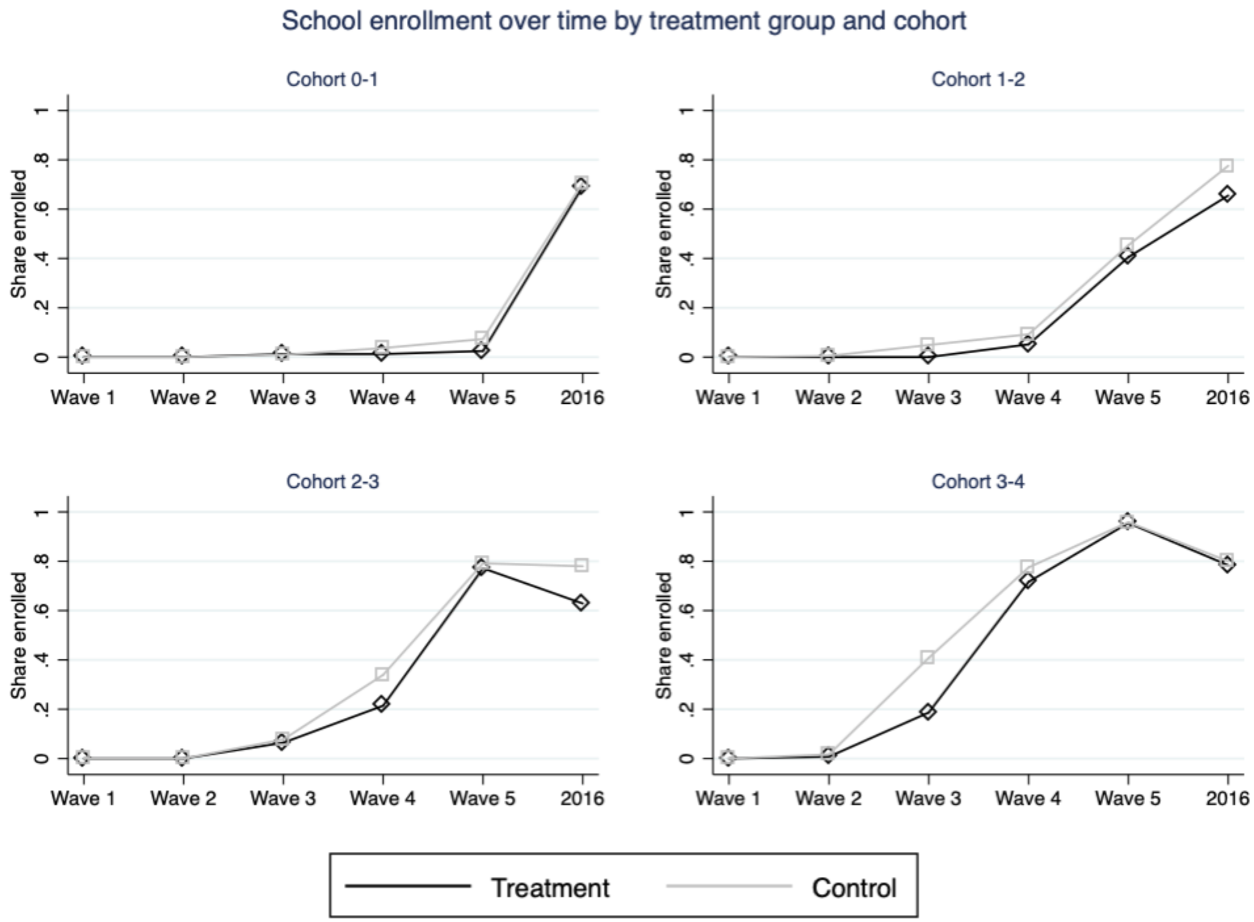
Notes: This figure plots treatment effects by subgroup. The top left panel compares younger (0-2) and older children (2-4). The bottom left panel compares higher and lower SES children (based on whether household wealth is above or below the median in the sample). The bottom right panel compares children 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.

Appendix Figure G4: Primary 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 Wave 5.

Appendix Figure G5: Primary 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**

Notes: This table reports correlation between the factor scores and demographic characteristics of the household and investments.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix Table G2: Baseline characteristics of children and families (non-attritors)

		All		Treatment		Control		
	Observations	Mean	SD	Mean	SD	Mean	SD	P-value
Panel A: Demographics								
Age in months	1,001	25.378	12.439	26.939	13.020	24.159	11.834	0.009
Gender (male)	1,001	0.519	0.500	0.515	0.500	0.523	0.500	1.000
Race (black)	1,001	0.619	0.486	0.642	0.480	0.601	0.490	0.869
Mother is single	1,001	0.276	0.447	0.271	0.445	0.279	0.449	1.000
Mother works	1,001	0.238	0.426	0.230	0.421	0.244	0.430	0.997
Mother secondary complete +	1,001	0.369	0.483	0.369	0.483	0.368	0.483	1.000
Childcare use at baseline	1,001	0.170	0.376	0.203	0.402	0.144	0.352	0.169
Father lives at home	1,000	0.690	0.463	0.686	0.465	0.693	0.461	1.000
Father secondary complete +	940	0.406	0.491	0.420	0.494	0.396	0.489	0.990
Wealth index	1,001	0.063	4.657	-0.200	4.170	0.269	4.998	0.732
Children books at home	1,001	1.421	2.609	1.478	3.032	1.375	2.224	0.994
Household size	1,001	5.327	2.009	5.262	1.900	5.377	2.090	0.978
No. of children <= 5 yrs	1,001	2.676	0.796	2.772	0.840	2.601	0.751	0.012
CCT beneficiary	1,001	0.353	0.478	0.351	0.478	0.354	0.479	1.000
Panel B: Child development								
Cognitive development								
Bayley Cognitive	743	48.573	14.741	49.360	15.180	48.016	14.414	0.466
Bayley Receptive	737	19.332	7.814	20.102	8.034	18.789	7.617	0.118
Bayley Expressive	739	19.648	9.312	20.225	9.901	19.240	8.860	0.466
Bayley Fine Motor	739	32.591	9.462	33.170	9.780	32.182	9.219	0.466
Bayley Gross Motor	742	47.085	13.127	47.485	13.697	46.799	12.712	0.509
TVIP	358	7.176	6.118	7.837	6.760	6.429	5.220	0.134
Woodcock-Munoz: Applied Problems	229	89.100	9.814	88.236	9.256	90.104	10.378	0.466
ELSA: Reading Comprehension	229	2.485	3.130	2.686	3.157	2.259	3.098	0.509
Health								
Weigh-for-age Z-score	977	-0.362	1.036	-0.438	0.999	-0.302	1.061	0.100
Height-for-age Z-score	978	-1.112	1.082	-1.113	1.005	-1.112	1.141	0.995
Weight-for-height Z-Score	969	0.332	0.971	0.253	0.961	0.394	0.976	0.072
Arm circumference	989	15.309	1.104	15.349	1.066	15.277	1.133	0.494
Socio-emotional development								
ASQ Total Score	988	46.335	29.502	49.391	30.628	43.932	28.383	0.010

Notes: The table reports baseline child and household characteristics by treatment group. Column (8) reports two-sided Romano-Wolf p-values for the difference between treatment and control.

Appendix Table G3: Heterogeneity by age

	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

	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

	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

	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

	Observations	Cohort 0-2		Cohort 2-4		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

Notes: This table describe the nutritional status and food fragility of the sample at baseline by age of the children at program enrollment. Cohort 0-2 refers to children who were younger than 2 at the start of the program. Cohort 2-4 refers to children who were between 2 and 4 at the start of the program.

Appendix Table G8: Nutritional differences at baseline by gender

	Observations	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

Notes: This table describe the nutritional status and food fragility of the sample at baseline by child gender.

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

	Primary school enrollment			
	Parental reports		Administrative records (2016)	
	(1)	(2)	(3)	(4)
Treatment effect, Wave 2	-0.003	-0.047		
	(0.424)	(0.000)		
Treatment effect, Wave 3	-0.047	-0.092		
	(0.009)	(0.000)		
Treatment effect, Wave 4	-0.005	-0.058		
	(0.857)	(0.011)		
Treatment effect, 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

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.

REFERENCES

Agostinelli, Francesco, and Matthew Wiswall. Estimating the technology of children's skill formation. No. w22442. National Bureau of Economic Research, 2016.

Anderson, T. and Rubin, H. (1956). Statistical inference in factor analysis. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, volume 5, 111–150. University of California Press, Berkeley

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

Bartlett, M. S. "The statistical conception of mental factors." British Journal of Psychology 28, no. 1 (1937): 97-104.

Blaga, Otilia M., et al. "Structure and continuity of intellectual development in early childhood." Intelligence 37.1 (2009): 106-113.

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.

Carneiro, P., Hansen, K., & Heckman, J. (2003). Estimating Distributions of Counterfactuals with an Applicaton to the Returns to Schooling and Measurement of the Effects of Uncertainty on Schooling Choice. International Economic Review

Cheadle, Jacob E. "The Early Literacy Skills Assessment (ELSA)."

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

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.

Early, Diane M., et al. "Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs." Child development 78.2 (2007): 558-580.

Fernald, L. C., Kariger, P., Engle, P., & Raikes, A. (2009). Examining early child development in low-income countries: a toolkit for the assessment of children in the first five years of life. World Bank.

Garcia, J. L., J. J. Heckman, and A. L. Ziff. "Gender Differences in the Benefits of an Influential Early Childhood Program." European Economic Review 109 (2018): 9-22.

Gorsuch, R. "Factor Analysis." In Handbook of Psychology: Research Methods in Psychology, Volume 2, 143-164. John Wiley & Sons, Inc., Hoboken, NJ, 2003.

Gorsuch, R. "Factor Analysis." Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.

Heckman, J. J., R. Pinto, and P. Savelyev. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." American Economic Review 103, no. 6 (2013): 2052-2086.

Leroy, Jef L., et al. "The Oportunidades program increases the linear growth of children enrolled at young ages in urban Mexico." The Journal of Nutrition 138.4 (2008): 793-798.

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

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

Pianta, Robert C., et al., eds. Handbook of early childhood education. Guilford Press, (2012).

Ponitz, C. E., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., and Morrison, F. J. "Touch Your Toes! Developing a Direct Measure of Behavioral Regulation in Early Childhood." Early Childhood Research Quarterly 23, no. 2 (2008): 141-15.

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

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

Squires, J., Bricker, D., & Twombly, E. "Psychometric studies of ASQ." ASQ-3 user's guide (2009).

Westat, Inc. "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, 2011.

Word Health Organization. "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, 2007.

World Health Organization. "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, 2006.