HCEO WORKING PAPER SERIES

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

Raquel Bernal
Michele Giannola
Milagros Nores

Working Paper

2022-027

09/2022; Revised 10/2024

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

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

www.hceconomics.org

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

September 2024

Raquel Bernal, Michele Giannola and Milagros Nores *

Early childhood development is a global priority, yet questions remain about how best to support disadvantaged children in low- and middle-income countries. We study the impacts of high-quality, center-based early childhood education on low-income children in Colombia through a randomized controlled trial, tracking participants longitudinally over five years. The program has sustained, positive effects on children's health - including a 22% reduction in stunting - and positive effects on their cognitive development, partly driven by improvements in children's health and nutrition. These effects operate through the educational and nutritional components of the program, rather than through parental behavioral responses to the intervention. The cognitive gains do not persist in the last year of the study, as children from the control group progress earlier to primary school and catch-up, highlighting the importance of considering counterfactual care alternatives to understand the effects of preschool programs in low- and middle-income settings. (JEL: J13, I10, I20, H43).

---

* *Bernal*: Universidad de Los Andes, Bogotá, Colombia (email: rbernal@uniandes.edu.co). *Giannola*: University of Naples Federico II, Naples, Italy, Center for Studies in Economics and Finance, and the Institute for Fiscal Studies (email: michele.giannola@unina.it) *Nores*: National Institute for Early Education Research, Graduate School of Education, Rutgers, The State University of New Jersey (email: mnores@nieer.org). 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 most grateful to aeioTU for opening their doors to our evaluation team, and to iQuartil for their excellent work in managing data collection on site. We also gratefully acknowledge the valuable research assistance provided by Santiago Lacouture and Diana Pérez. The views expressed are those of the authors and do not necessarily represent those of the funders. The authors confirm contributions 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 Rutgers IRB (IRB Approval Number E09-568). Trail Registry # AEARCTR-0001903, https://www.socialscienceregistry.org/trials/1903.

Poverty hampers the development of millions of children in low- and middle-income countries (LMICs), at a great cost to individuals and society (Black et al. (2016)). Socio-economic disparities in cognitive and language development emerge as early as age three and worsen over time (Barnett & Lamy (2013); Duncan & Magnuson (2013); Fernald, Kariger, Hidrobo & Gertler (2012); Heckman (2008)). High-quality early interventions have shown the potential to alter children's developmental trajectories and reduce opportunity gaps (Almond, Currie & Duque (2018); Black & Dewey (2014); Daelmans et al. (2017)). Today, high-quality comprehensive programming for early childhood development is a global priority (Berlinski & Schady (2016); Nores & Barnett (2010)), recognized as a target under the United Nations' Sustainable Development Goals for 2030 (United Nations (2016)). Nevertheless, there remain questions about the most effective strategies to achieve these goals in LMICs.

A substantial body of evidence from the U.S. and other high-income countries points to positive effects of center-based preschool programs on child development, that are sustained in the long run (Bailey, Sun & Timpe (2021); Barnett & Jung (2021); Campbell & Ramey (1994); Currie & Thomas (1995); Deming (2009); Drange & Havnes (2019); Englund et al. (2014); Gray-Lobe, Pathak & Walters (2023)). These studies find that even in the presence of convergence in the cognitive outcomes of treated and untreated children, there are long-term effects in other domains, such as health and schooling (Nores & Prayag (2024); Reynolds, Ou, Mondi & Hayakawa (2017)), thanks to improvement in non-cognitive outcomes (Heckman, Pinto & Savelyev (2013)). However, it is unclear that findings for the U.S. can be generalized to low- and middle-income countries. First, the quality of care – which is essential to program effectiveness (Araujo & Schady (2015)) – varies considerably across contexts and is typically poorer in developing countries (Yoshikawa et al. (2018)). Second, the counterfactual experiences of children in the control group – which are crucial to understanding program's impacts (Kline & Walters (2016)) – are also likely to be very different in the U.S. than in low- and middle-income countries. Therefore, extrapolating the demonstrated benefits of center-based education to LMICs could well be misleading.

Despite the substantial growth of center-based care across the developing world (according to the World Bank, the share of children enrolled in childcare in LMICs rose from 15% in 1980 to 58% in 2020)[1], to date most studies have focused either on home-based intervention for children aged 0-3, or on center-based interventions for older children (Attanasio et al. (2022)); Evans, Jakiela & Acosta (2024)).

This paper fills this knowledge gap by presenting findings from a five-year Randomized Controlled Trial (RCT) studying the effects of a high-quality, center-based preschool program on the development of

[1] Official statistics from the World Bank can be found here: https://data.worldbank.org/indicator/SE.PRE.ENRR?end=2022&name_desc=false&start=1970&view=chart

disadvantaged children aged between zero and four in Colombia. We randomly assigned over 1,000 children from two disadvantaged communities in the north of the country to a treatment group and a control group using an admission lottery, and followed them and their families longitudinally, conducting a baseline and four yearly follow-ups. We study the program's effects on children's developmental trajectories while they are still age-eligible for pre-primary education – that is, in the years before their enrollment in primary school. We leave the study of longer-run effects for future research based on administrative data being collected on children's standardized test scores on national achievement tests.

The program we study is named "aeioTU". Similar to the widely studied Head Start and Abecedarian programs in the U.S., aeioTU envisions a holistic approach for the developing child, and aims to cultivate children's cognitive, social, emotional, and physical capabilities. For this reason, the program integrates an educational component with health and nutritional components and encourages the active participation of families. Integrated programs of this kind, which are common in higher-income countries, are the current focus of the new 2023 WHO and UNICEF guidelines and goals for children under five in LMICs. Therefore, research on these types of programs is fundamental to inform early childhood care and education (ECCE) programming in the developing world.

Children in aeioTU centers receive full-day care, 11 months a year, and are provided with 70 percent of their daily nutritional requirements through breakfast, lunch, and two snacks. The program's educational philosophy is inspired by the Reggio Emilia approach (Malaguzzi (1993)), featuring key elements of process quality such as project- and play-based learning, rich adult-child interaction, and a balanced mix of teacher-directed and child-initiated activities. Program centers are characterized by low child-to-teacher ratios and highly qualified teachers with extensive pre- and in-service training. Like Head Start, aeioTU encourages parental engagement through workshops, regular communication, and meetings. We evaluate aeioTU comprehensively and do not disentangle the distinct effect of each program component.

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

The program has positive effects on children's cognitive development, with effect sizes as large as 0.36 standard deviation (SD) relative to the control group. The magnitude of these effects corresponds to about one-third of the socio-economic gradient in children's cognitive development in urban Colombia (Rubio-Codina et al. (2015)). 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 decrease of 5 percentage points 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)). We find no effect on children's socio-emotional development. We also show that there are no significant changes in children's home environments, suggesting that the effects operate through the program itself rather than through parental behavioral responses to the intervention.

Research on the effects of preschool programs in both higher- and lower-income countries emphasizes impacts heterogeneity by gender, age, and socio-economic status (Cascio (2021); Evans, Jakiela & Acosta (2024); Garces, Thomas & Currie (2002); Garcia, Heckman & Ziff (2018)). We study whether aeioTU has heterogeneous effects along these dimensions, showing that the impact on cognitive skills is larger for girls than boys, while the health effects are more pronounced for boys. In line with recent evidence (Britto et al. (2017)), we also find that children with a lower baseline level of development benefit the most, both in cognitive development and in health. Finally, we find that the impacts of the treatment on health are stronger for children who were older upon enrollment.

Following children over time, we then study the longitudinal effects of the intervention. We find that health impacts are remarkably stable, whereas the effects on cognitive development are mitigated in the last study year, even among children who are still age-eligible for pre-primary education. Investigating the reasons for this convergence, we show that, compared with the treatment group, children in the control group transition to primary school earlier. Following the common practice in the literature (e.g., Kling, Liebman & Katz (2007); Abdulkadiroğlu, Angrist & Pathak (2014)), we use an instrumental variable strategy exploiting excluded interactions between experimental assignments and household characteristics to account for the selection of children into alternative care arrangements and show that, compared with aeioTU attendance, early enrollment in primary school improves children's cognitive outcomes by 0.7 SD. Using these estimates and enrollment patterns in alternative care arrangements observed in the data, we perform a simple back-of-the-envelope calculation to show that earlier enrollment of control group children in primary school can fully explain the convergence in cognitive outcomes. These findings highlight the importance of considering counterfactual alternatives to understand the effects of preschool programs in low- and middle-income settings (Kline & Walters (2016)). Our results further underscore the significant role that primary school might play in sustaining and furthering the early developmental benefits of pre-primary education, and suggest that a timely transition to high quality primary education might be essential to maximize the long-term benefits of early childhood programs (Bailey et al. (2020); Reynolds & Temple (2019)).

This study makes several contributions to the literature. First, it examines the dynamic impacts of aeioTU over a five-year period. Longitudinal evaluations of early childhood programs in developing countries are rare (Tanner, Candland & Odden (2015)) but necessary to fully understand their effects on disadvantaged children's developmental trajectories. Only a few experimental studies outside the U.S. have followed children for more than a year. Bernal, Attanasio, Peña & Vera-Hernández (2019) follow children transitioning from home-based to center-based care in Colombia over an 18-month period and report neutral or negative effects on their development. Dean & Jayachandran (2020) consider the effects of pre-primary education in India for older children (aged 3.5 to 4.5), finding positive effects on cognitive development that persist in first grade, but no effect on socio-emotional development. Attanasio et al. (2022) study the effect of public childcare in Brazil for children aged 0-3 and find that childcare attendance has a positive short-run effect on cognitive development, a sustained effect on health, and no effect on socio-emotional development. We add to this growing body of work on the effects of center-based care in LMICs, analyzing children's cognitive, socio-emotional and health development over a five-year period.

The longitudinal nature of the study further allows us to answer questions related to program duration; in particular, we examine variations in the effects for different lengths of childcare exposure. Exploiting treatment-induced variations in enrollment duration, we assess how exposure length influences the development of children in younger cohorts. Our analysis shows that each additional year of childcare improves health by between 0.09 and 0.146 SD. At the same time, our findings caution against the linear extrapolation when considering cognitive development: Assuming linearity in the effects, we find that each additional year improves children's cognitive skills by 0.227 SD; however, a non-linear model reveals that total returns follow and inverse U-shaped pattern, with the marginal effect being large and positive for the first two years, null for the third year and negative for the fourth year of childcare enrollment.

Second, our results highlight the importance of considering child health in assessing the benefits of center-based early education. While childcare centers worldwide often also feature health and nutritional components, evaluations typically focus only on children's cognitive and socio-emotional development (e.g., Berlinski, Galiani & Gertler (2009); Dean & Jayachandran (2020); Berkes, Bouguen, Filmer & Fukao (2024)). This omission is noteworthy for several reasons. First, integrated programs – combining nutritional and educational components – are the focus of the current WHO and UNICEF guidelines for children under five, so evaluating their effects on health is crucial to understanding their full potential. Second, the health and nutrition services of these programs are likely to play an important role in reducing nutritional deficiencies, which in turn can improve children's ability to learn (Bailey, Sun & Timpe (2021); Frisvold (2015); Simeon (1998)) and thus help explain the measured impact on cognitive development. Drawing on production function estimates from the same context (Attanasio, Bernal, Giannola & Nores (2020)), we predict that the children's health gains facilitated by the program contribute to improvements in cognitive skills, with an estimated magnitude ranging from 0.019 to 0.038 SD, thus explaining between 5 to 10 percent of the treatment effects.

Third, leveraging extensive and detailed data on parental behavior and children's home environments, we can assess whether the effects of center-based care for disadvantaged children are mediated by parental responses. Evaluations of center-based programs in developed and developing countries typically ignore the mechanisms behind the measured impacts, but recent evidence indicates that the parents of older school-aged children respond to public programs by changing the level and type of investments they privately provide to their children (Das et al. (2013); Pop-Eleches & Urquiola (2013); Greaves, Hussain, Rabe & Rasul (2023)). We show that in our study such behavioral responses are not present, indicating that the program's positive effects on children's development are not mediated by changes in parental investments at home.

Finally, the aeioTU leadership engaged in a nationwide effort to promote comprehensive changes in early childhood care and education in Colombia and so became part of a broader national early childhood strategy aimed at increasing access and improving the quality of childcare services (Mesa, Nores & Vega (2021)). Over time, the program grew to serve over 13,000 children throughout Colombia (3.5 percent of all children enrolled in center-based care in 2016), making it a noteworthy case study and speaking to its scalability potential, particularly given the financial and human capital constraints present in Colombia, which are typical of low- and middle-income contexts.

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

I. Background and intervention

A. Early childhood education in Colombia

Colombia's economic growth rate declined from 7 percent in 2010 to 1.7 percent in 2018, while inequality eased 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 early childhood care and education (ECCE) from six months to five years of age, with the expectation that they then transition to primary school. Primary education includes a first transitional year (grade 0) called *Transición*, although enrolment in this grade is not universal (OECD 2016). This grade is followed by five years of compulsory primary school.

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); Rubio-Codina et al. (2015)). The Colombian Institute of Family Welfare (*Instituto Colombiano de Bienestar Familiar*, ICBF) has led efforts to mitigate these inequalities through investment in ECCE, but the rate of enrollment in public childcare centers for socioeconomically vulnerable children is still no better than 38 percent (Bernal & Camacho (2014); World Bank, (2013)). Another 20 percent of children nationwide attend home-based care services known as *Hogares Comunitarios de Bienestar* (HCB). Although the original purpose of HCB was to promote female labor supply, and despite the low quality of the service, evaluations have shown its positive effects on children's 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"). The strategy aimed 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 380,000 children by 2016. The increased access to integrated center-based care facilitated by the

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

[3] Comprehensive childcare services provide a holistic approach encompassing nutrition, healthcare, nurturing, and early education. These services incorporate pedagogical content designed to foster cognitive and socio-emotional development, not merely offering a safe environment for children while their parents' work.

implementation of DCAS resulted in improvements in children's language and cognitive development (Andrew et al. (2024); Bernal & Ramírez (2019)).

The aeioTU organization was a strong advocate for DCAS, and the aeioTU program thus ended up forming part of this national strategy (Mesa, Nores & Vega (2021)). The present study should accordingly be interpreted in the context of Colombia's large-scale investment in ECCE beginning in the early 2010s.

B. The aeioTU program

The program we study is called "aeioTU", after the NGO that runs the preschool centers. Like the famous Head Start program in the U.S., aeioTU is an integrated center-based early education intervention, including educational and health components as well as family participation.

At the time this evaluation was planned in 2010, the program offered full-day (9 hours) center-based care for children under the age of five, for 11 months a year, delivered via a public-private partnership between aeioTU and DCAS. At the outset, the Colombian government provided a stipend of USD 1,500 per child per year to the centers, which aeioTU supplemented out of its own resources. This changed in 2014, when aeioTU ended the supplement after careful planning and cost-structure monitoring.

Program centers had distinct classes for infants and toddlers. The centers had low child-to-teacher ratios (4:1 for infants and 6:1 for toddlers), highly qualified teachers (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 other public center-based programs in Colombia, which had inferior teacher qualifications, higher child-to-teacher ratios (25:1 for toddlers) and lacked teacher training or coaching strategies (Bernal, Attanasio, Peña & Vera-Hernández (2019)). Teacher training, curricular support, classroom materials and quality monitoring improved over time. Nores, Figueras-Daniel, López & Bernal (2018) report increasing levels of process quality comparable to other high-quality programs in the region. However, child-teacher ratios and teacher qualifications deteriorated owing to public requirements and funding constraints (for instance, the government required providers to hire personnel from the HCB as some of those programs were phased out).

Another key component of the program is health and nutrition: aeioTU provides 70 percent of children's daily nutritional requirements through breakfast, lunch, and two snacks. This became mandatory in all public child-care centers in 2012, but the requirement was not in place at the beginning of this study.[4]

[4] We did not collect data to monitor nutritional intakes in program centers.

As a comparison, Head Start provides between one-half and two-thirds of daily nutritional needs for full-day services (U.S. Department of Health & Human Services (2024)). Moreover, aeioTU centers are staffed with an on-site nutritionist who periodically monitors children's nutritional status. Finally, aeioTU 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 may be an important pathway to support disadvantaged children's development (Ogunlade et al. (2011); Simeon et al. (1998)).

The educational philosophy of the program, inspired by the Reggio Emilia approach (Malaguzzi (1993)), emphasizes project- and play-based learning, rich adult-child interaction, the intentional integration of learning across multiple domains, and a balanced mix of teacher-directed and child-initiated activities.[5] The program is organized around projects and study themes that emerge from the children's play and investigations. Daily activities are guided and structured through pedagogical guidelines and group planning 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. In line with the Reggio Emilia philosophy, aeioTU centers also operate an open-door policy with families, encouraging their active participation through workshops, regular communication, and meetings (including active encouragement of mothers to come to the center to breastfeed their infants).

Finally, aeioTU uses data on children and classrooms to monitor and improve quality, a feature that is crucial to ECCE quality (Bernal (2015); Bowman, Donovan & Burns (2001); Yoshikawa, Weiland & Brooks-Gunn (2016)). These characteristics align with the essential features of high-quality ECCE as highlighted in the literature (Jensen et. al (2019); Singer, Golinkoff & Hirsh-Pasek (2006); Yoshikawa et al. (2018)). By contrast, other DCAS programs are free to define their own curricula, following only broad national standards. Consequently, comparable public center-based programs are not required to use research-based curricula or include pedagogical guidelines for teachers' activities (Bernal, Attanasio, Peña & Vera-Hernández (2019)). We do not have comparative assessments to confirm that aeioTU had better quality than other programs targeting the same communities. We can only qualify that the inputs often

[5] 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 environment through explorations, which belong spontaneously to children's everyday experiences, their play, their speaking, thinking, and negotiating (Malaguzzi (1993)). Adults are mentors and guides in this process, not mere caregivers or knowledge providers. That is, they offer opportunities for children to explore their own interests. The approach recognizes many ways of understanding the world and expressing thoughts, and seeks to promote these communication channels within the educational experience, including art, music, dance, movement, play and exploration.

linked to center-based quality (e.g., children to teacher ratios, training, qualifications) were higher on average in the aeioTU centers.

In sum, much like center-based programs evaluated in the U.S. and other higher-income countries, aeioTU is an integrated ECCE program including an educational and nutritional component, and family participation. For ethical and logistical reasons, it was not feasible to evaluate the specific contribution of these distinct components (for example, by providing only the educational component and not the health and nutritional components to some children). Therefore, we follow the literature and focus our analysis on the overall effects of this integrated intervention on children's development. Integrated programs are currently the focus of the 2023 WHO and UNICEF guidelines and goals for children under five in LMICs (WHO (2023); UNICEF (2023)), which highlight the importance of understanding their potential for ECCE planning in the developing world.

By 2016, aeioTU operated 20 centers and provided comprehensive ECCE to about 13,300 low-income children nationwide, or 3.5% of all Colombian children enrolled in center-based care that year.

II. Methods

A. Study timeline

We evaluate the effects of aeioTU using a Randomized Controlled Trial (RCT) with families of young children in two disadvantaged communities in northern Colombia. Figure 1 shows the study timeline, including yearly assessments. We conducted baseline assessments in late 2010 (Y1), prior to randomly assigning children and before the start of the program. The baseline data collection for the first community took place between July and September 2010, and the program started in November. In the second community, data collection took place between October and December, and the program began in March 2011. We then tracked children longitudinally over five calendar years, through the end of 2014. Assessments took place 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 randomized controlled trial using an oversubscription model. The evaluation sites were selected among early childhood 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 to detect program effects and (ii) sufficient demand (i.e., oversubscription), to allow for an assignment lottery among applicants. Two eligible communities, where centers were under construction and scheduled to open in 2010, were identified in northern Colombia. These communities were deemed suitable given the scarcity of alternative ECCE services. Location choices for the centers also included the political will of a mayor who supported the initiative (by funding the infrastructures) and the study, as well as the ICBF's approval, which prioritized underserved areas, and community support, which was garnered through meetings with community members.

The two study communities are part of the Caribbean coastal region, the second most populous in the country, which is known for its racial and ethnic diversity. In 2010, the poverty rate in these communities was above the national average (at about 40%) and comparable to that of other mid-size cities in the country. The high poverty rate is due in part to the large influx of displaced populations during the preceding decade and their dependency on informal labor (Meisel-Roca & Ricciuli-Marín (2018)). As in the rest of the country, poverty rates in the region decreased in the early 2010s. In Section III.A below, we describe how families and children in our sample compare to 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 in these communities. A total of 1,288 eligible children were identified. All the families met the income-eligibility requirements and expressed interest in enrolling their child in aeioTU if offered a slot. From the initial sample, 70 children were excluded from the randomization: 66 were offered a slot for such reasons as being related to center staff, 2 moved out of the communities before randomization, and 2 exceeded the program's age eligibility before the start of the services. The final sample thus consisted of 1,218 children under the age of five. All the families gave their active consent to participate in the study.

Random assignment to the program was stratified by age group (five cohorts, corresponding to children aged 0-1, 1-2, 2-3, 3-4, and 4-5 at baseline), gender and neighborhood. We used computer-generated randomly ordered lists to assign children to the treatment and control groups. The randomization took place in a public in-person lottery event, where both treated and control families were able to witness the slots in aeioTU being assigned through the lottery. Following the lottery, the centers reached out to the winners for enrollment. Where the winners declined the offer, the centers selected substitutes from the control group using the publicly generated random-order lists.

As is common in the experimental literature studying the effects of early childhood interventions, we consider the initial random assignment to treatment in an Intention-To-Treat framework (ITT) for the first set of results. We also use an instrumental variable approach exploiting the lottery random assignment to study the effects of actual program enrollment on children's development (Gray-Lobe, Pathak & Walters (2023)). Therefore, our estimates do not experience issues related to imperfect compliance that arise in randomized waitlist designs (Chaisemartin & Behagel (2020)).

Child assessors and parent interviewers were blind to treatment status. However, parents may have revealed their status at post-testing, so that over time the assessors may have learned this information.

C. Study sample

From a policy perspective, it is important to distinguish the effects of aeioTU on children's development while they are eligible for pre-primary education, from the effects once children transition to primary school. This paper focuses on the former question only, estimating the impact of aeioTU on pre-school children, and leaves the study of longer-run outcomes for future research (based on administrative data being collected on children's standardized test scores on national achievement tests). Consequently, our analysis is restricted to children who are *age-eligible* for pre-primary education in any 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 (*Transición*) is mandated. Since data collection took place during the second half of each school year, we determined children's eligibility in a particular year based on their eligibility for at least 30 percent of the time between the third week of January of the previous year and wave *t*'s data collection (we show below that our findings are robust to alternative eligibility definitions).

The analytical sample consists of 1,073 children aged zero to four at baseline, ensuring that we observe at least one follow-up in which the child is still eligible for the program. Given the longitudinal nature of the data, older children are progressively excluded from the analysis as they are expected, based on their age, to transition to primary education. If we were to pool *eligible* and *non-eligible* children, the results on the program's impact would be confounded with the effects of primary school. As a robustness check, we also report results when eligible and non-eligible children are pooled, understanding that these estimates are likely to confound program effects with subsequent schooling experience.

Table 1 describes the sample, reporting the changes in sample size at each wave. At baseline 1,073 children under four were randomized into a treatment group (471 children were offered a slot) and a control group (602 children were not offered a slot). By Y3, 6 percent of children in the oldest cohort (three to four years old at baseline) had lost eligibility; and by Y4, 48 percent of children in this age cohort was no longer eligible. By Y5, all children in the oldest cohort had lost eligibility, and 47 percent of 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 α=0.05, a 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 controlling for baseline covariates). These magnitudes are 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 aeioTU would affect children's development positively, particularly in view of the limited supply of ECCE alternatives in the two low-income communities in the study (as we describe below, at baseline only 17.1 percent of the children had used childcare services). Specifically, we expected the program's quality components, infrastructure, and operational supports to enhance learning opportunities, improving children's language, cognitive and socio-emotional development. We also anticipated that the health monitoring and nutritional components of the program would benefit children's health, which could in turn improve their cognitive and socio-emotional development (Frisvold (2015); Sudfeld et al (2015)).

In addition, we predicted that the program could influence the children's home environment and parental engagement, but the direction of such changes was ambiguous, a priori. On the one hand, the emphasis on active family participation could enhance parental knowledge of child development and improve parenting; further, if the marginal return to parental investment increases with improvements in the quality of early education (because of complementarities between these two), parents might be induced to invest additional resources in their child. On the other hand, parents might respond to the program by redistributing resources, including food and time, to non-treated children or to themselves, potentially attenuating program impacts (Barrera-Osorio et al. (2011); Giannola (2024)).

E. Measures and outcomes

Given 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 development. To study the effects on children's outcomes, we used well-established instruments with strong psychometric properties, recommended for ECCE evaluations, including those in developing countries (Fernald, Prado, Kariger & Raikes (2017)). Child assessments were administered by trained psychology graduates and seniors, who achieved high standards of reliability (100 percent agreement with the trainer) through a two-week training that included live reliability with children. Assessors were offered refresher training every year, and any new assessors were fully trained in similar conditions. The data collection was conducted in rented and adapted spaces, ensuring identical conditions for treatment and control children. Information on children's caregivers and home environments was collected through direct interviews with the caregivers, conducted in separate rooms alongside the assessment of the child. Parents and children received small incentives for participation, and an on-site snack.

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

Health: We measured children's health by measurements of height, weight, and arm circumference, following the standards and procedures set out by the World Health Organization (WHO (2006, 2007)). We use these to construct age-normed indicators of health and nutritional status (e.g., height-for-age, HAZ, and weight-for-age, WAZ) that describe children's development by comparison with a healthy population.

Cognitive skills: We measured children's cognitive development using a battery of assessments including the Bayley Scales of Infant Development III (BSID-III; Bayley (2005)), the Peabody Picture Vocabulary Test in Spanish (TVIP) (Padilla, Lugo & Dunn (1986)), the Early Literacy Skills Assessment (ELSA; DeBruin-Parecki (2005)), the Woodcock-Muñoz III Tests of Achievement (WM; Muñoz et al. (2005)) and the Head-Toes-Knees and Shoulders (HTKS) task (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) (Bracken, Keith, & Walker (1998); Doyle et al. (1997)) and the Vineland Adaptive Behavior Scales (Sparrow, Balla, Cicchetti & Harrison (1985)).

Home environment: We assessed children's home environments through direct survey-structured interviews with their primary caregivers. We collected information on the caregivers' discipline strategies and on nutritional and feeding habits (e.g., whether the child skipped a meal in the week preceding the interview, and the nutritional content of the meals consumed). We further measured time investment both in terms of total number of hours spent by each parent with the child and in terms of whether the parents performed specific activities with the child such as playing, singing, or reading.

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

Given the longitudinal nature of the study, a central issue is how to compare treatment effects across different points in time. Little (2013) emphasizes the importance, when measures change throughout a study, of making them comparable in order to capture the evolution of the same underlying developmental domain. Unfortunately, longitudinal studies often compare scores from different measures over time, even where it is not clear how these scores translate between different assessments.[6] When age-appropriate measures for younger children are replaced by measures that are appropriate at later ages, there is limited scope for comparing program effects over time. We exploit the availability of a common subset of measures 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 a single 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 factor model. Our main results report treatment effects on these latent factors, but we also report results using individual measures in our analysis.[7]

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

[7] Appendix Figure G1 plots the distributions of the factor scores, and Appendix Table G1 reports the correlations between the factor scores and baseline sociodemographic characteristics. The distributions are well behaved, and the correlations have the expected signs. Cognitive development is positively correlated with maternal education, household wealth, and the number of books in the home. The correlation between cognitive development and maternal education strengthens with the child's age. Health is correlated only with maternal education. The correlations between socio-emotional skills and household characteristics are weaker, the most important factor here being the number of children's books at home.

F. Empirical strategy

We estimate Intent-To-Treat (ITT) effects on child development for age-eligible children based on the following Ordinary-Least-Squares (OLS) specification:

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

where $Y_{i,s,t}$ is the outcome of child *i* in strata *s* at time *t*. W_t 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 the treatment group in the initial lottery and 0 otherwise. δ_s are randomization strata fixed effects, and X_i is a vector of background characteristics included to improve efficiency and to account for potential imbalances between groups in the distribution of pre-random assignment characteristics (Kling, Liebman & Katz (2007)). We control for tester effects in our main specification but also conduct robustness checks excluding them. We cluster standard errors at the individual level, the unit of randomization.

β_t in equation (1) identifies the ITT program impact at each survey wave, that is, the effect of being offered a slot in an aeioTU center. To benchmark the magnitude of the effects, we report the impacts in terms of standard deviation units of the outcome variable in the control group at baseline. Variations of this model inquire into heterogeneous effects by age at baseline (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 several outcomes. To address issues of multiple inference, we compute Romano & Wolf (2005) step-down p-values, testing four hypotheses within each developmental domain, one for each study wave. We report both unadjusted and adjusted p-values in the analysis below. These p-values correspond to one-tailed tests for the intervention's impact, reflecting the assumption that the program would not harm the children, as detailed by our theory of change, and as in other evaluations of ECCE interventions (e.g., Attanasio et al. (2014)). We also report two-sided p-values below.

We also report two-stages-least-squares (2SLS) estimates for the impact of attending an aeioTU centers on child outcomes, using the initial lottery assignment as an instrument for actual program participation. Initial lottery assignment is a valid instrument as it was randomly allocated, and it significantly predicts enrollment (we show below that a regression of program enrollment on random assignment yields a large and strong first stage). We report two sets of 2SLS results. First, using information on effective enrollment we construct total enrollment and estimate the effect of *cumulative* participation in the program on children's development. Under the assumption of linearity, this is the effect of attending a program center for one *additional* year. Second, relaxing the linearity assumption, we report the 2SLS estimates for *duration* of attendance, exploiting treatment-induced variation in length of exposure.

These 2SLS effects are local average treatment effects for the "compliers", that is, those children whose program participation (or duration of exposure) was influenced by the lottery (Imbens & Angrist (1994)). Kline & Walters (2016)) show that the interpretation of these LATEs depends on the prevalence of alternative childcare options available (e.g., other center-based care, home care) and demonstrate that when those alternatives are not rationed, the LATEs are the policy-relevant parameters for evaluating expansions of aeioTU centers, irrespective of the range of possible alternatives.

The results of Kline and Walter (2016) imply, further, that in the presence of substitution, these LATE parameters are a weighted average of counter-factual specific sub-LATEs. In Section IV, we use 2SLS to identify these sub-LATEs, exploiting excluded interactions between experimental assignments and household characteristics to account for the selection of children in alterative care arrangements (Kling, Liebman & Katz (2007); Abdulkadiroğlu, Angrist & Pathak (2014)).

III. Results

A. Baseline characteristics

Table 3 describes baseline characteristics of children and families in the sample and reports balance by lottery assignment. Panel A reports demographic characteristics of children and families. At baseline, children were on average 25 months old. They lived in households with on average 2.7 children under age five, and 27 percent lived in single-headed households. Mothers had on average 8.5 years of schooling; only 36 percent had a high school degree. There are some baseline statistically significant differences between lottery winners and losers. Treatment families have slightly more children under 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 infants than toddlers and preschoolers in the treatment group by design – the class sizes for younger children were smaller and the slots were correspondingly 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). Replicating the analysis for the subsample of children re-interviewed in each wave, these main facts are confirmed (Appendix Table G2). We control for these baseline imbalances in our main specification, but we also run robustness checks without controlling for these covariates.

Panel B of Table 3 describes children's development at baseline. To assess their vulnerability in cognitive development, we compute standardized BSID III scores at baseline (these are not shown in Table 3, which reports raw scores).[8] The scores for cognitive, language and motor development are 90.4 (SD=13.3), 88.9 (SD=13.2), and 93.6 (SD=13.6), respectively. That is, children in the sample were approximately 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)). By comparison, children between 18 and 36 months of age in middle-income households in Bogotá 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 those of children in low socio-economic urban households in the ELCA (Colombian Longitudinal Household Survey (2013)). At baseline, children were also highly vulnerable nutritionally, with height-for-age one standard deviation below WHO standards (WHO (2006, 2007)). About 21.6 percent of children in our sample were stunted, compared to a national rate of 15.2 percent in rural areas and 12 percent in urban areas (ELCA (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 with wealthier children in the country and worldwide.

There are no significant differences between treatment and control children at baseline in cognitive development or health, but the treatment group have slightly higher ASQ:SE scores. We control children's baseline outcomes in all specifications.

B. Attrition

A detailed account 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 follow-up varied slightly from wave to wave, since some who were not found in one wave were found in a later one. We estimate the probability of non-attrition as a function of treatment status and household characteristics (Appendix Table C2). We find that: (i) attrition rates do not differ between the treatment and control groups (the point estimate is -0.017, p-value=0.426), (ii) less vulnerable families are less likely to be surveyed in all the waves, and (iii) beneficiaries of the national

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

conditional cash transfer program (CCT) 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

Appendix Table D1 reports the effect of winning the lottery on enrollment in an aeioTU center. The first two columns report whether the child *ever* attends a center and columns 3 and 4 report the effect on the *number of years* of attendance. 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 40 percent of the control group children attended a program center at some point during the study period. In the treatment group, that share rises to 70.4 percent (or an increase of 76 percent, p-value for the difference < 0.001). Turning to number of years enrolled in aeioTU, column 3 shows that, on average, children who did not win the lottery attended the childcare centers for 1.12 years, and the winners attended for 2.56 years, (an increase of more than 100% increase, p-value for the difference < 0.001). In columns 2 and 4, we further control for baseline covariates and find similar effects. Appendix Figure D1 plots the distribution of cumulative enrollment by lottery status: 47 percent of control children enrolled for at least one year and 19 percent for three or more years during the study period. For the treatment group, the figures are 76 and 50 percent.

Following the patterns found in other longitudinal evaluations of center-based programs in other settings (e.g., Attanasio et al. (2022); Clements et al. (2013); Dean & Jayachandran (2020)), compliance with treatment assignment decreased over time, since for ethical reasons the program was not rationed and control children were allowed to fill center slots as these became available. Figure 2 reports aeioTU enrollment rates by randomization status at each study wave. In Y2, 72 percent of lottery winners were enrolled in centers and 81 percent of losers were not. By the last year, 53 percent of the lottery winners were not enrolled, while the enrollment rate of control group children increased from 19 percent in Y2 to 36 percent in Y5. These patterns are heterogeneous between cohorts, with higher enrollment rates for the youngest and lower for the oldest, consistent with the progressive transition of the latter to elementary school (Appendix Figure D2).

These results show that in the course of the study there was cross-over from control to treatment, particularly for children enrolled as babies or toddlers. As noted above, we address this by considering ITT estimates in a first set of results and using lottery assignment as an instrument for program enrollment to estimate a Local Average Treatment Effect (Imbens & Angrist (1994)). As in Kline & Walters (2016)), these LATEs are treatment effects in relation to a mix of counterfactual childcare alternatives: center-based childcare options other than aeioTU, home-based care, and primary school enrollment (we return to this point in Section IV.D).

We also analyze the determinants of compliance for the entire sample and separately for the treatment and control groups (Appendix Table D2), finding that: (i) compliance is unrelated to lottery assignment and (ii) on average, compliers and non-compliers do not differ significantly in observable characteristics.

D. ITT effects on child development

Figure 3 and Table 4 present intention-to-treat estimates by child development domain. The estimates are based on equation (1), and each coefficient represents the impact of winning the lottery on a given outcome in year t for children who are still age-eligible in that year. The effects are reported as fractions of a standard deviation of the outcome in the control group at baseline. In Table 4, we include unadjusted p-values and Romano Wolf p-values adjusted for multiple hypothesis testing. For each domain, the p-values test the significance of the treatment effects across the four follow-ups.

We find a positive and significant impact of winning the lottery on cognitive skills. At the first follow-up (Y2), just eight months into the program, treated children's skills are 0.193 SD higher than those of the control group (step-down p-value=0.068). This effect increases to 0.366 SD in Y3 (step-down p-value=0.051) and is 0.270 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 convergence between treated and control children in Section IV.D. The effects in Y2, Y3 and Y4 are not statistically different from one another, while the effect in Y5 is statistically different from that in Y3 at the 10 percent level, but not from those in the other study periods.

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) and increases over time, to 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 (effect sizes between -0.007 and 0.069 SD) and never statistically significant.

We run several robustness checks in Appendix E. Appendix Table E1 shows the robustness of the results when excluding tester effects. Appendix Table E2 checks the sensitivity of our estimates to the exclusion of household characteristics that were unbalanced at baseline, again confirming the main findings. To determine whether the changes in treatment effects over time for cognitive skills are influenced by changes in the estimation sample (given that we progressively exclude older children who are no longer age-eligible), Appendix Table E3 reports the treatment effects only on the children who remain age-eligible until the final year of the study. The patterns are consistent with those in Table 4, but less precisely estimated owing to the smaller sample size. Appendix Table E4 reports the effects for children who are age-eligible for ECCE for 50 percent of the time between two survey dates (rather than 30 percent as in our main specification). These coefficients are very similar to those reported in Table 4, with a slight increase in the estimated impact on child health with the more conservative 50 percent eligibility threshold (the Y5 effect is 0.21 SD as against 0.16 SD in the main specification). Appendix Tables E5-E7 similarly report the results for the full sample of children who were originally randomized in the lottery, regardless of their age-eligibility in each study wave. Appendix Table E5 shows that for the full sample the effects on cognitive skills are similar in Y2 and Y3 and larger in Y4 (0.40 compared to 0.27 in our main specification), while the health effects are larger in Y5 (0.231 compared to 0.157 in our main specification), suggesting that the impact of aeioTU is likely to persist even after children start primary school.

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 positive effects on the cognitive, language and motor sub-scales of the Bayley Scales of Infant Development in the first two years of the program (fewer than 40 children were in the age-range of the Bayley in Y4, and none in Y5). We also find positive effects on language skills measured by the TVIP and ELSA tests, which fade out over time, in line with the results in Table 4. The results further show that the effects on the Woodcock-Muñoz III Tests of Achievement, measuring math skills (applied problems), and on children's executive functions assessed using the HTKS test are never statistically significant.

Finally, in line with Table 4, Appendix Table E9 shows that children in the treatment group have higher weight-for-age (WAZ) and height-for-age (HAZ) than those in the control group. The effect size on HAZ ranges between 0.109 and 0.157 SD, that on WAZ between 0.087 and 0.177 SD. We also find that the program leads to a decrease in the probability of being stunted (length-for-age Z-score < -2) – a primary manifestation of chronic nutritional deprivation – of 3 to 5 percentage points, a 13-22 percent reduction relative to the control group at baseline. At the same time, Appendix Table E9 shows there is no effect either on children's arm circumference or on their Body Mass Index (BMI).

E. Effects of enrollment on child development

Table 5 reports the 2SLS results instrumenting *cumulative* program enrollment with random assignment to treatment; thus, the coefficients correspond to the effect of one *additional* year of attendance at an aeioTU center. As noted above, we use the lottery as an exogenous shifter for program participation: a regression of enrollment on winning the lottery yields a highly significant first stage (the F-statistic is close to 30, as reported in Table 5).

The 2SLS estimates mirror the ITT coefficients in Table 4, but the magnitudes are generally larger, as expected. Column 1 pools all the waves together and assumes linearity in the effects of program exposure (e.g., the effect of one additional year for a child already enrolled for one year is the same as that of one year for a child previously not enrolled at all). The estimates indicate that one additional year improves children's cognitive skills by 0.227 SD (p-value=0.038) and health by 0.129 SD (p-value=0.009). Extrapolating, this suggests that the difference between a full four years of pre-school attendance and none is 0.904 SD for cognitive skills and 0.516 SD for health.

Table 5 further reports 2SLS estimates by study wave, showing that one additional year of aeioTU increases children's cognitive skills by 0.359 SD in Y2 (step-down p-value=0.062), 0.389 SD in Y3 (step-down p-value=0.046), and 0.227 SD in Y4 (step-down p-value=0.154) and decreases it in Y5 (-0.039 SD, 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 impact of enrollment on socio-emotional skills (the point estimates are always small in magnitude and never statistically different from zero).

Exploiting treatment-induced variation in the total length of program enrollment, we relax the linearity assumption and investigate how effects vary by duration in the program ("dosage"). Table 6 reports these 2SLS estimates for age-eligible children, showing that the linear approximation captures well the longitudinal effects of aeioTU on children's health. The estimates show that, compared to not attending, attending aeioTU for one year improves child health by 0.146 SD (step-down p-value=0.108). Attending for two, three and four years improves health by 0.274 SD, 0.38 SD and 0.47 SD respectively (the step-down p-values are 0.019, 0.019 and 0.087). These estimates imply marginal effects ranging from 0.09 to 0.146 SD, in line with the estimate in column 1 of Table 5.

At the same time, the results caution against the linear extrapolation for cognitive development. Compared to non-attendance, one year of exposure improves children's cognitive skills by 0.358 SD (step-down p-value = 0.066), and two and three years improve them by 0.777 and 0.762 SD respectively (step-down p-value, 0.036 and 0.098). But four years of program exposure produce an effect of -0.104 SD (step-down p-value= 0.559). These coefficients imply that the total returns for the younger children are non-linear, following an inverse U-shaped pattern. In terms of marginal effects, they suggest that one additional year is beneficial if the child was not enrolled or was enrolled for one year (the marginal effect for the first year is 0.358 SD, and for the second year it is 0.777-0.358 = 0.419 SD). If a child was already enrolled for two years, the marginal effect is close to zero (0.762-0.777=-0.015 SD). Finally, for a child who has already spent three years in the program, the effect of one additional year is negative (-0.104-0.762=-0.866 SD).

F. Effects on children's home environment

The program's effects on children's home environments and parental behavior are reported in Table 7. As is specified above in our theory of change, we expected the program to affect various aspects of the child's home environment. Since the program provided 70 percent of children's daily nutritional requirements, we anticipated that parents might alter the level and/or type of food provided at home – potentially cutting back the child's food consumption or giving them less nutritious food. Instead, we find no effect on an overall index of nutritional investment, the magnitude of the effect ranging from -0.017 SD to 0.046 SD, and never statistically significant. Beaton & Ghassemi (1982) and Jacoby (2002) discuss the issues that arise concerning in-kind transfers to children if parents do alter their child's nutritional intake in response. Our findings indicate instead that the program's nutritional transfer "sticks" to the child.

Recent evidence shows that parents respond to public programs by changing the level and type of investments they privately provide to their children (Das et al. (2013); Pop-Eleches & Urquiola (2013); Greaves, Hussain, Rabe & Rasul (2023)). We explore whether the educational component of the program or the fact that centers actively seek to engage parents in their children's development affected this investment, focusing on parental time and discipline strategies. The results in Table 7 show no effect on these two intermediate outcomes. As for nutritional investments, the estimates are consistently close to zero and not statistically significant, indicating that the program has no discernible impact on children's home environments.

As a whole, the evidence in Table 7 suggests that the program neither crowded-in nor crowded-out parental investments. The fact that parents did not alter their behavior suggests that the positive effects on children's cognitive and health development are due to the intervention itself, not to changes induced at home.

G. Heterogeneous effects

Research on preschool programs in both higher- and lower-income countries often finds heterogeneity of impacts by gender, age, and socio-economic status (Cascio (2021); Evans, Jakiela & Acosta (2024); Garces, Thomas & Currie (2002); Garcia, Heckman & Ziff (2018)). We study whether aeioTU has heterogeneous effects along these dimensions.

Figures 4 and 5, along with Appendix Tables G3-G6, present heterogeneous ITT effects on children's cognitive and health development along four dimensions: age at enrollment (younger children (0-2) vs. older (2-4)), gender, baseline development (splitting the sample at the median level of the outcome variable at baseline), and household wealth (splitting the sample at the median). Note, however, that except for age, these heterogeneity analyses were not pre-specified, and we may be underpowered to detect small differences across groups.

We find positive effects on children's cognitive development for both younger and older children from Y2 to Y4 (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 larger for girls than for boys. The effect sizes for girls in each year are 0.479 (p-value=0.019), 0.499 (p-value=0.017), 0.303 (p-value=0.155) and 0.279 (p-value=0.174). For boys, the effects in Y3 and Y4 are positive but not statistically different from zero (coefficients=0.216 and 0.221, p-values=0.152 and 0.158). In Y2 and Y5 they are negative (coefficients = -0.065 and -0.359, p-values=0.620 and 0.913). 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).

We find suggestive evidence of a stronger impact on cognitive development for children from relatively less disadvantaged households than from the most disadvantaged (Figure 4 and Appendix Table G5), although we can reject the equality of coefficients only 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), while for more vulnerable children the effects are smaller and never statistically different from zero. The point estimates also suggest that the effects are larger for children with lower cognitive development at baseline in all waves but the last (Figure 4 and

Appendix Table G6), although the null hypothesis of equality of coefficients across the two groups cannot be rejected.

Figure 5 and Appendix Tables G3-G6 show heterogeneity in the treatment impacts on health. Older children benefitted the most from the intervention, with consistently larger point estimates. The effects for older children increase over time, with an 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). For younger children, the effects are consistently smaller, although the equality of coefficients is only rejected in Y4 (Appendix Table G3). We observe large effects for boys, increasing over time, with effect sizes ranging from 0.093 in Y2 (p-value=0.129) to 0.27 SD in Y5 (p-value=0.004). The impacts for girls are always smaller, with effect sizes between 0.008 and 0.121 SD.

We also find evidence that the positive effects on health are driven by the most socioeconomically vulnerable children, with effect sizes of 0.264 SD in Y4 (p-value=0.002), and 0.446 SD in Y5 (p-value < 0.001), while the effects on children from less vulnerable households are small and never statistically different from zero (Appendix Table G5). Finally, children with poorer baseline health benefited the most from the intervention in all waves but the last (the null hypothesis of equality of coefficients across the two groups is rejected in Y2 and Y3 but not in later waves; see Appendix Table G6).

There are no meaningful heterogeneous effects on socio-emotional development. The results are reported in Appendix Figure G2 and Appendix Tables G3-G6. We discuss these results in detail in the next section.

IV. Discussion and Mechanisms

We start by discussing how our results compare with the few existing experimental evaluations studying the effects of increasing access to center-based care in LMICs, as well as with the broader international evidence.[9] Table 8 reports the characteristics and estimated ITT effects of nine experimental studies in Africa, Asia, and Latin America. Before turning to the estimated impacts, we highlight two patterns that emerge from the comparison of our study with those reported in Table 8. First, most evaluations (6 out of 9) consider the effects only for older children (aged 3-6). Second, only a minority report health impacts, instead focusing on cognitive or socio-emotional development.

[9] Other studies have considered improvements in the quality of existing childcare centers in LMICs,; see Andrew et al. (2024) and Özler et al. (2018).

A. Health

We find sustained positive effects of aeioTU on a composite index of children's health, with effect sizes ranging from 0.08 SD in Y2 to 0.16 SD in Y5, and a reduction of 13-22 percent in the incidence of stunting. The effects are large compared with the other experimental evaluations of center-based care in Table 8 (ranging from 0.05 to 0.138 SD). The magnitudes of these estimated impacts are also important in the context of the international literature. The meta-analysis of Nores & Barnett (2010) reports average effects of 0.31 SD for nutrition-specific interventions, and 0.23 SD for interventions combining nutritional and educational components.

In the Colombian context, these findings contrast with those of Bernal & Fernández (2013), who find no gains from a home-based care program with a similar nutritional component. Similarly, Bernal, Attanasio, Peña & Vera-Hernández (2019) report minor effects (0.05 SD) on children's nutritional status in public center-based care relative to home-based care. Andrew et al. (2024) find that improvements in the quality of center-based care through DCAS produced no significant health gain. By contrast, we find that aeioTU has a substantial impact on children's health development that increases over time.

Turning to the heterogeneity of effects, the greater effect on older children may be due to their higher baseline vulnerability on anthropometric indicators compared with 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). Another possible explanation for the differential effects by age might be that younger children are less likely to consume their meals in center-based care (Andrew et al. (2018)).

In contrast with Attanasio et al. (2022), we find suggestive evidence that the effects on health are larger for boys than for girls. This gender heterogeneity cannot be explained either by baseline differences in nutritional status (Appendix Table G8) or by food fragility or nutritional investment at home (results not reported). We speculate that gender differences in eating behavior may contribute to these gendered effects (Keller et al. (2019)). We also show that children with poorer baseline health benefited the most from the intervention. Finally, we find suggestive evidence of larger effects on children's health status for the full sample of children, regardless of age eligibility, which suggests that the health effect may be persistent. Future research should investigate the longer-run effects.

B. Cognitive development

Consistent with six of the nine ITT estimates reported in Table 8, we find a positive effect of center-based care on children's cognitive skills. The effect sizes we estimate are as large as 0.36 SD in Y3, which puts aeioTU at the higher end of the range of estimates relative to the studies reported in Table 8 (effect sizes between -0.160 and 0.390 SD). These effects are also sizable considering the cognitive development gap of nearly one standard deviation between high- and low-SES children in Colombia (Bernal, Martínez & Quintero (2015)).

The magnitude of these effects is also substantial by comparison with other studies in Colombia. Bernal & Fernández (2013) find positive effects on children's cognitive skills after 15 months of exposure to *Hogares Comunitarios*, with effects ranging from 0.15 to 0.30 SD. Bernal, Attanasio, Peña & Vera-Hernández (2019) report negative effects (-0.10 SD) of public center-based care compared to home-based care. Andrew et al. (2024) report an effect of 0.15 SD on cognitive development resulting from quality improvements in public childcare centers. Putting these figures in an international perspective, the 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 evidence on the gendered impacts of ECCE is mixed (see Magnuson et al. (2016) for a recent meta-analysis). In this study, we find larger effects for girls. While there are no socio-economic differences between boys' and girls' families, there is evidence of more parent-reported interaction with boys in play and reading in our setting (Nores, Bernal & Barnett (2019)). This suggests that girls may have experienced less stimulating home environments, making them more responsive to the enhanced developmental opportunities offered by the program (Garcia, Heckman & Ziff (2018)). The stronger effects for girls appear to be common in LMICs (Evans, Jakiela & Acosta (2024)).

While there is evidence that ECCE has larger effects on children from less affluent backgrounds in higher income countries (Cascio (2023); Cornelissen, Dustmann, Raute & Schönberg (2018); Felfe & Lalive (2018)), the evidence of such heterogeneity in developing countries is inconclusive (Evans, Jakiela & Acosta (2024)). In our setting, the results need to be interpreted in the light of high socio-economic vulnerability, thus limiting generalization to low-income households. With this in mind, we find suggestive evidence that the cognitive effects are concentrated on children from less vulnerable households. Finally, there is evidence suggesting that children with lower baseline cognitive development benefited the most from the intervention, strongly coinciding with other studies in the early childhood literature (see Britto et al. (2017) for a review).

There is also evidence that improvements in children's health and nutrition increase their capabilities to learn and improve their achievement (Frisvold (2015); Simeon (1998)). It is posited that one mechanism through which center-based care improves cognitive skills is improvement in children's health status (Bailey, Sun & Timpe (2021)), but there is little empirical evidence for this. We use human capital production function estimates based on data from this RCT (Attanasio, Bernal, Giannola & Nores (2020)) to shed light on this question. Attanasio, Bernal, Giannola & Nores (2020) estimate that improvement in health at age 3 translates into a gain in cognitive development of approximately 0.24 SD. Based on this, we predict that the estimated health impacts of 0.08-0.16 SD (Table 5) should translate into improvements in cognitive development ranging between 0.019 and 0.038 SD. Therefore, approximately 5 to 10 percent of the treatment impact on cognitive skills can be attributed to improved health and nutrition.[10]

C. Socio-emotional development

Table 8 shows that one consistent pattern across studies is the lack of statistically significant effects on children's socio-emotional skills (point estimates are typically smaller than 0.05 SD). Consistently, we find that the effects on socio-emotional development are always small in magnitude and never statistically significant. These results also align with previous evidence from Colombia. Bernal, Attanasio, Peña & Vera-Hernández (2019) find no effects of center-based care on socio-emotional development compared with home-based care. Andrew et al. (2024) report no effects of a teacher training and coaching program on children's socio-emotional development. Bernal & Fernández (2013) find improved interactions with peers but an increase in disruptive behavior for children attending home-based care. 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 parent-reported measures (Renk & Phares (2004)).

A null effect could also be interpreted positively in relation to studies on center-based intervention in higher income countries, which report negative effects on socio-emotional development. Center-based care offers children opportunities to develop social skills while increasing the likelihood of disruptive behavior, since the children compete for limited resources (Baker, Gruber & Milligan (2008, 2019);

[10] We estimate this as: 0.08 x 0.24 = 0.019. In a different context, Attanasio, Meghir and Nix (2020) estimate that at age 5 a 1-SD increase in health improves cognitive skills by 0.05 SD. Such estimates would imply an improvement in children's cognitive development ranging between 0.004 and 0.008 SD.

Haskins (1985)). The absence of negative effects of aeioTU suggests a comparatively favorable context in this regard.

D. Convergence

We find convergence in the cognitive skills of children in the treatment and control groups in the last year of the program, even among children who are still age-eligible for pre-primary education. In this section we explore explanations for this result.

We start by examining enrollment in primary school over time according to treatment status. Figure 6 plots primary school enrollment rates by treatment status and over time, combining information from the parent survey and administrative records from the Colombian Ministry of Education (Appendix Figure G3 shows the results by child cohort of birth). As expected, given the children's age, the rate is null at the beginning of the study and rises progressively to 70 percent by 2016. Importantly, children in the treatment group have lower primary enrollment rates than children of the same age in the control group, even though the former are on average three months older (Table 3), indicating that the control group progressed earlier to elementary school.[11]

In Table 9 we regress primary school enrollment on a treatment indicator and confirm these findings: there is a small difference in primary school enrollment between treatment and control children in Y2 (0.3 percentage points). The difference increases over time, and in the last study wave the treatment group's enrollment rate is 9.1 percentage points (22 percent) lower (this difference is statistically significant at the 5 percent level). The magnitude of this difference is similar when using administrative enrollment records rather than parental reports (columns 3 and 4 of Table 7) and for the full sample of children regardless of age-eligibility (Appendix Table G9).[12] These results are similar to Behrman et al. (2024), who find that increasing access to center-based preschool in Mexico delayed children's primary-school entry.

---

[11] Similar results for the whole sample (irrespective of age-eligibility) are presented in Appendix Figure G4 and follow a similar pattern in terms of differences between treatment and control children. We also note that parent-reported enrollment rates in primary education are higher than administrative records would indicate. These trends occur in a community context where first grade enrollment rates were 53 percent in 2011 and 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 (last accessed in September 2024).

[12] Appendix Figure G5 breaks down enrollment patterns by care arrangements, treatment group and study wave. Consistently with Figure 2, across study waves lottery winners are more likely to be enrolled in aeioTU. As a result, the share of children cared for at home or in alternative preschool programs is higher in the control group. As is shown in Appendix Figure G5, enrollment in home or center-based care decreased and primary school attendance increased over time. The substitution between home or center-based care and primary education is stronger in the control group.

While assignment to the treatment group appears to be correlated with later transition to primary school, this is not sufficient to explain the convergence in cognitive skills between treatment and control children in the last year of the study. Such convergence would require that children in primary school learn 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: home care, other childcare programs, and elementary school.

Random assignment to the program alone is insufficient to identify these effects, as we have more than one endogenous variable. We therefore use a 2SLS approach treating aeioTU enrollment and enrollment in other childcare or schooling arrangements as separate endogenous variables. Specifically, we consider the following endogenous choices: aeioTU center, home care, a different childcare or pre-school center, and primary school. In our setting, these are the only empirically relevant choices (Appendix Figure G5).

Following the common practice in the literature, we generate instruments for these endogenous variables by interacting experimentally assigned program offer with observed characteristics (e.g., Abdulkadiroğlu, Angrist & Pathak (2014); Garcia, Heckman & Ziff (2018); Kline & Walters (2016); Kling, Liebman & Katz (2007)), running two sets of 2SLS estimates. First, we use as instruments the interactions between offer and site (neighborhood) indicators and the interactions between offer and birth cohort. Second, we add interactions between offer and household level characteristics (including an indicator for whether the child's father lives in the same home and one for whether the mother works).[13] Under the assumption that the subLATEs do not vary across the interacting groups (constant effects), the 2SLS strategy recovers the causal effects of interest (Kirkeboen, Leuven & Mogstad (2016); Kline & Walters (2016)).

Table 10 reports the 2SLS estimates for the effects of alternative care and education arrangements on cognitive development, where the excluded category is attending aeioTU. Relative to aeioTU, home care and alternative center-based childcare have a negative effect on children's cognitive skills of approximately -0.3 SD. By contrast, primary school has a large positive effect of approximately 0.7 SD compared to aeioTU. These results are robust across the two sets of instruments we consider (Table 10, first and second rows).

[13] We account for the fact that the endogenous variables are binary using Procedure 21.1 in Wooldridge (2010).

Using the estimates in Table 10 and the enrollment patterns in our data (Appendix Figure G5), we perform a simple back-of-the-envelope calculation to assess the extent to which differences in childcare arrangements and their changes over time can explain the observed dynamic in treatment impacts, and specifically the convergence in cognitive development in the last year of the study. Multiplying the estimates from Table 10 by the corresponding share of children in each arrangement in the treatment and control groups, we compute the implied treatment effects and contrast these with the estimates from Table 4. Further details on this exercise are reported in Appendix F. The results, reported in Appendix Table G10, show that this accurately replicates the patterns and magnitudes of the treatment effects observed. In the second study wave, the estimated treatment effect is 0.193 SD, against our prediction of 0.159 SD. Consistent with the estimates in Table 4, our simple calculations further predict that treatment effects should increase in Y3, diminish in Y4, and turn negative (and small) in Y5 (for this year the estimated treatment effect is -0.053 and our prediction is -0.057).

One disadvantage of the 2SLS approach we use in Table 10 is that it does not allow us to track the effects over time (this would be too demanding on the data as it would include 12 endogenous regressors, one for each care arrangement and wave). To further explore the contribution of primary school attendance to the convergence observed in the last study year, we report ITT effects excluding in each year the children who are enrolled in school (Appendix Table G11). While these effects are not causal, since we estimate them on a potentially selected sample of children, they are in line with our previous analysis. Focusing on cognitive skills only, we find positive and significant effects of the program throughout the five years. In the sample of children not enrolled in school, the intention-to-treat effects are 0.172 SD in Y2, 0.328 in Y3, 0.322 in Y4, and 0.374 in Y5.

Taken together, the results in this section support the hypothesis that the later transition of treated children to primary school explains the observed convergence in cognitive skills. We conclude this discussion by highlighting that convergence in cognitive outcomes (or fade-out) is rather common in early childhood interventions when children progress through primary school. However, the benefits re-emerge later in the life cycle in terms of schooling trajectories, earnings, employment, health, and reduced crime rates (Chetty et al. (2011); Jenkins et al. (2018); Heckman, Pinto & Savelyev (2013); Weiland et al. (2019)).

V. Conclusions

This paper studies the effects of high-quality center-based early childhood education on the development of disadvantaged children in Colombia. We study the aeioTU program which provides all-day ECCE eleven months a 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 evaluate aeioTU through a five-year randomized control trial, following children longitudinally to study the effects of the program on their developmental trajectories in the years leading up to primary school. We take a holistic view of child development, measuring children's health, cognitive and socio-emotional skills.

We find positive effects of the program on children's cognitive skills and health, but no discernible impact on their socio-emotional skills or home environment. Our findings demonstrate that improving access to high-quality ECCE can improve the development of disadvantaged children in low- and middle-income countries. The findings further underscore the potential benefits of integrated early education programs for child health. The results are important in light of the current WHO and UNICEF guidelines for children under five in LMICs. In addition, our results suggest that, because of cross-productivity between health and cognitive skills in the child human capital production function, the nutritional and health components of early care and education programs can amplify the total returns to cognitive development.

Following children over time, we document the effects of the intervention over time, finding remarkably stable effects on health. On the other hand, the impacts on children's cognitive development are significantly attenuated in the fifth year of the study, even among children who are still age-eligible for pre-primary education. We show that this convergence is explained by the earlier transition to primary school for children in the control group and on the positive effect that primary education has on children's cognitive development. These findings highlight the significant role that primary school might play in sustaining and furthering the early developmental benefits of pre-primary education, and suggest that a timely progression to high quality primary education might be essential to maximize the long-term benefits of early childhood programs (Bailey et al. (2020); Reynolds & Temple (2019)).

The growing body of work in Colombia and other low- and middle-income countries emphasizes the need for better understanding and measurement of quality in early childhood care and education and further investigating how to support children's socio-emotional development. This work also stresses the importance of a cohesive evaluation agenda to understand and build quality in ECCE programming. This accumulation of within- and across-country evidence is essential to inform and support further expansion of ECCE in the developing world.

The multi-year nature of this evaluation and the fact that it was embedded in the context of the DCAS national strategy exposes our study to the same challenges highlighted by previous research: the difficulties and the ethical problem of limiting crossover, and the plausibility of interactive cofounders. This is due to the program (and program leaders) and the longitudinal evaluation informing DCAS, and DCAS expanding access to, and increasing the quality of center-based care and education nationwide. In line with the arguments of Deaton & Cartwright (2018) and Kline & Walters (2016), we explicitly analyze the impact of aeioTU *vis-à-vis* alternative care and educational arrangements to understand and interpret treatment effects.

Future research shall investigate whether the effects we observe are sustained in the long run. This is the key to illuminating the life-cycle benefits of high-quality ECCE programs for disadvantaged children in low- and middle-income countries.

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.

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

Andrew, A., Attanasio, O. P., Bernal, R., Sosa, L. C., Krutikova, S., & Rubio-Codina, M. (2024). Preschool quality and child development. Journal of Political Economy, 132(7), 000-000.

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.

Araujo, M. C and N. Schady. "Daycare Services: It's All about Quality." In The Early Years: Child Well-being and the Role of Public Policy, edited by Samuel Berlinski and Norbert Schady, chap. 4. New York: Palgrave Macmillan US, 91–121. (2015)

Attanasio, O. P., Maro, V. D., & Vera-Hernández, M. (2013). Community nurseries and the nutritional status of poor children. Evidence from Colombia. The Economic Journal, 123(571), 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, O., Bernal, R., Giannola, M., & Nores, M. (2020). Child development in the early years: Parental investment and the changing dynamics of different dimensions (No. w27812). National Bureau of Economic Research.

Attanasio, O., Meghir, C., & Nix, E. (2020). Human capital development and parental investment in India. The review of economic studies, 87(6), 2511-2541.

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.

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. Psychological Science in the Public Interest, 21(2), 55-97.

Bailey, M. J., Sun, S., & Timpe, B. (2021). Prep School for poor kids: The long-run impacts of Head Start on Human capital and economic self-sufficiency. American Economic Review, 111(12), 3963-4001.

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

Baker, M., Gruber, J., & Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. Journal of political Economy, 116(4), 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).

Barnett, W. S., & Lamy, C. (2013). Achievement gaps start early: Preschool can help in Carter, PL, & Welner, KG (Eds.). Closing the opportunity gap: what America must do to give every child an even chance.

Barnett, W. S., & Jung, K. (2021). 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, 248-259.

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

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.

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

Beaton, G. H., & Ghassemi, H. (1982). Supplementary feeding programs for young children in developing countries. The American Journal of Clinical Nutrition, 35(4), 864-916.

Behrman, J. R., Gomez-Carrera, R., Parker, S. W., Todd, P. E., & Zhang, W. (2024). Starting Strong: Medium-and Longer-run Benefits of Mexico’s Universal Preschool Mandate. Mimeo

Berkes, J., Bouguen, A., Filmer, D., & Fukao, T. (2024). Improving preschool provision and encouraging-demand: Evidence from a large-scale construction program. Journal of Public Economics, 230, 105050.

Berlinski, S., Busso, M., & Giannola, M. (2023). Helping struggling students and benefiting all: Peer effects in primary education. Journal of Public Economics, 224, 104925.

Berlinski, S., & Schady, N. (2015). The early years: child well-being and the role of public policy (p. 288). Springer Nature.

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

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

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

Bernal, R., & Ramírez, S. M. (2019). Improving the quality of early childhood care at scale: The effects of "From Zero to Forever." World Development, 118: 91-105.
Bernal, R., and Camacho, A. (2014). 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.
Bernal, R., Attanasio, O., Peña, X., & Vera-Hernández, M. (2019). The Effects of the Transition from Home-Based Childcare to Center-Based Childcare in Colombia. Early Childhood Research Quarterly, 47, 418-431.
Bernal, R., M. A. Martínez, and C. Quintero (2015). Situación de Niñas y Niños Colombianos Menores de Cinco Años entre 2010 y 2013." Bogotá, Colombia: Editorial Kimpress.
Black, M. M., & Dewey, K. G. (2014) Promoting equity through integrated early child development and nutrition interventions. Annals of the New York Academy of Sciences, 1308(1): 1-10.
Black, M. M., Walker, S., Fernald, L., Andersen, C., DiGirolamo, A., Lu, C., McCoy, D., Fink, G., Shawar, Y., Shiffman, J., Devercelli, A., Wodon, Q., Vargas-Barón, E., & Grantham-McGregor, S. (2016) Early Childhood Development Coming of Age: Science through the Life Course. The Lancet, 389(10064): 77-90.
Bowman, B., Donovan, M., & Burns, M. (2001). Eager to Learn: Educating our Preschoolers. Washington, DC: National Academy Press.
Bracken, B. A., Keith, L. K., & Walker, K. C. (1998). Assessment of preschool behavior and social-emotional functioning: A review of thirteen third-party instruments. Journal of Psychoeducational Assessment, 16(2): 153-169.
Britto, P. R., S. Lye, K. Proulx, A. Yousafzai, S. Matthews, T. Vaivada, et al. (2017). Nurturing Care: Promoting Early Childhood Development. The Lancet 389: 91-102.
Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: a follow-up study of children from low-income families. Child development, 65(2), 684-698.
Cascio, E. U. (2021). Early childhood education in the United States: What, when, where, who, how, and why 1. In The Routledge handbook of the economics of education (pp. 30-72). Routledge.
Cascio, E. U. (2023). Does Universal Preschool Hit the Target?: Program Access and Preschool Impacts. Journal of Human Resources, 58(1), 1-42.
Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. The Quarterly journal of economics, 126(4), 1593-1660.
Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year. American Educational Research Journal, 50(4), 812-850.
Comisión Inter-Sectorial para la Primera Infancia CIPI. (2013) De Cero a Siempre: Atención Integral a la primera infancia. Estrategia de Atención Integral a la Primera Infancia. Fundamentos políticos, técnicos y de gestión. Colombia: Presidencia de la República.
Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg. (2018). Who Benefits from Universal Childcare? Estimating Marginal Returns to Early Childcare Attendance." Journal of Political Economy 126, no. 6: 2356-2409.
Currie, J., & Thomas, D. (1995). Does Head Start make a difference?. The American Economic Review, 85(3), 341-364.

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

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. School inputs, household substitution, and test scores (2013). American Economic Journal: Applied Economics, 5(2): 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., & Cartwright, N. (2018) Understanding and misunderstanding randomized controlled trials." Social Science & Medicine, 210: 2-21.

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

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. American Economic Journal: Applied Economics, 1(3), 111-134.

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

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

Drange, N., & Havnes, T. "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.

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).

Evans, D. K., Jakiela, P., & Acosta, A. M. (2024, May). The Impacts of Childcare Interventions on Children's Outcomes in Low-and Middle-Income Countries: A Systematic Review. In AEA Papers and Proceedings (Vol. 114, pp. 463-466). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.

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.

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).

Frisvold, D. E. (2015). Nutrition and cognitive achievement: An evaluation of the School Breakfast Program. Journal of public economics, 124, 91-104.

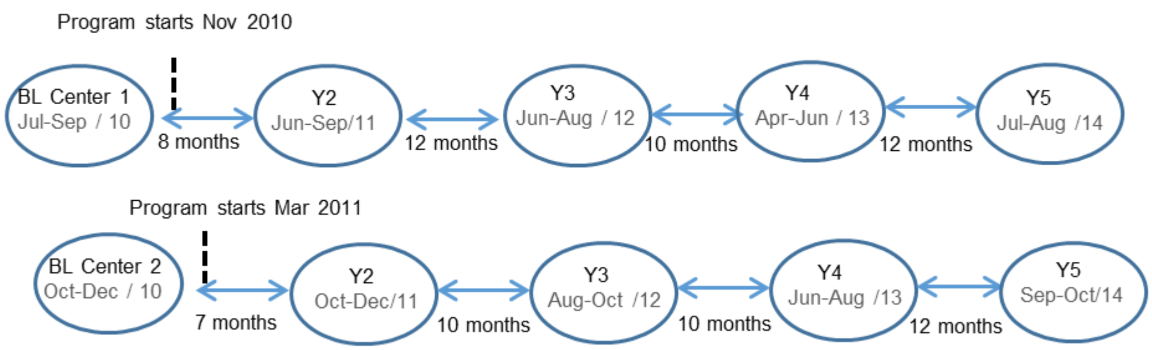
Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. American economic review, 92(4), 999-1012.
Garcia, J. L., J. J. Heckman, and A. L. Ziff. "Gender Differences in the Benefits of an Influential Early Childhood Program." European Economic Review 109 (2018): 9-22.
Giannola, Michele. "Parental investments and intra-household inequality in child human capital: evidence from a survey experiment." The Economic Journal 134.658 (2024): 671-727. 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.
Greaves, E., Hussain, I., Rabe, B., & Rasul, I. (2023). Parental responses to information about school quality: Evidence from linked survey and administrative data. The Economic Journal, 133(654), 2334-2402.
Haskins, R. "Public School Aggression Among Children with Varying Day-care Experience." Child Development 56 (1985): 689-703.
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.
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.
Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). Field of study, earnings, and self-selection. The Quarterly Journal of Economics, 131(3), 1057-1111.
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.
Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. Econometrica, 75(1), 83-119.
Little, T. D. "Longitudinal Structural Equation Modeling." New York, NY: Guilford, (2013).
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).
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.
Mesa, N., Nores, M., & Vega, H. (2021). Reaching thousands of children in low income communities with high-quality ECED services: a journey of perseverance and creativity. Frontiers in Public Health, 9, 637031.
Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., and Mather, N. "Batería III Pruebas de Aprovechamiento." Itasca, IL: Riverside Publishing, (2005).
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. & Prayag, R. (2024). Early Childhood Education Effects into Adulthood. In Berends, M., Lamb, S., & Schneider, B. (Eds.). SAGE Handbook of Sociology of Education. Sage: Chapter 14, pp. 221-234.
OECD (2016), Education in Colombia, Reviews of National Policies for Education, OECD Publishing, Paris, https://doi.org/10.1787/9789264250604-en.
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.
Pop-Eleches, C., & Urquiola, M. (2013). Going to a better school: Effects and behavioral responses. American Economic Review, 103(4), 1289-1324.
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.
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.
Reynolds, A. J., & Temple, J. A. (Eds.). (2019). Sustaining early childhood learning gains: Program, school, and family influences. Cambridge University Press.
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.
Weiland, C., Unterman, R., Shapiro, A., Staszak, S., Rochester, S., & Martin, E. (2020). The effects of enrolling in oversubscribed prekindergarten programs through third grade. Child Development, 91(5), 1401-1422.
Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.
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).

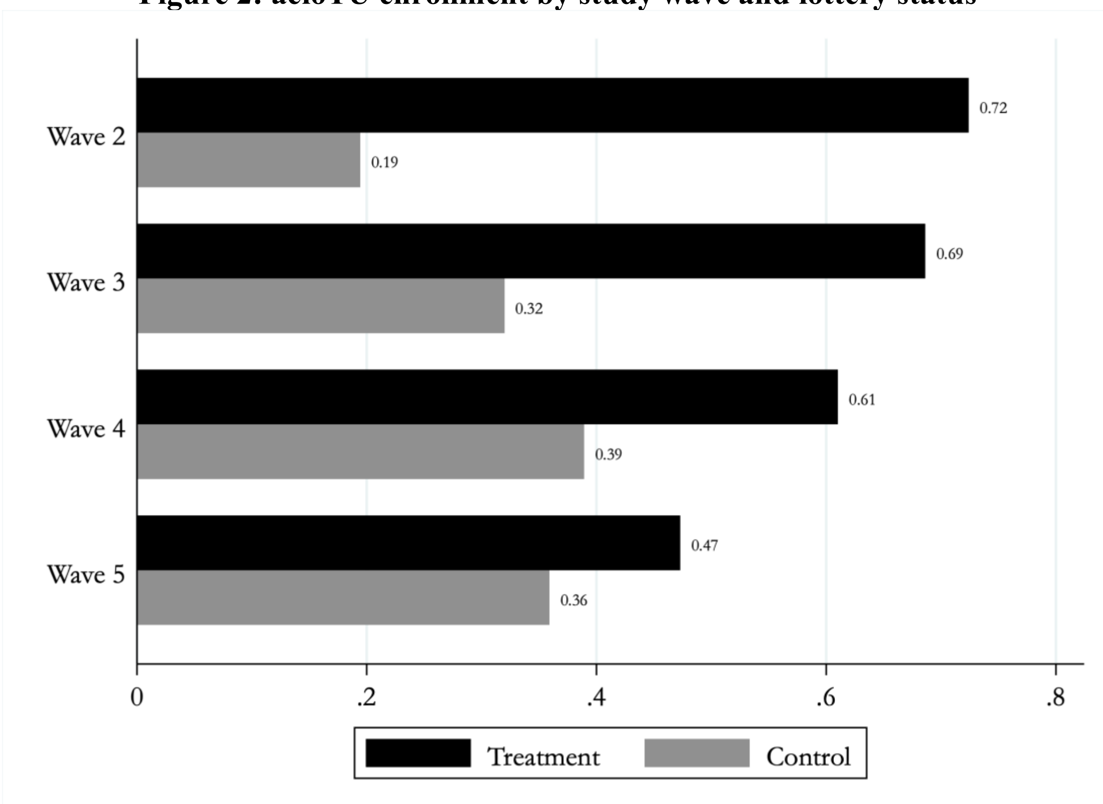
Tables & Figures

Figure 1. Study timeline



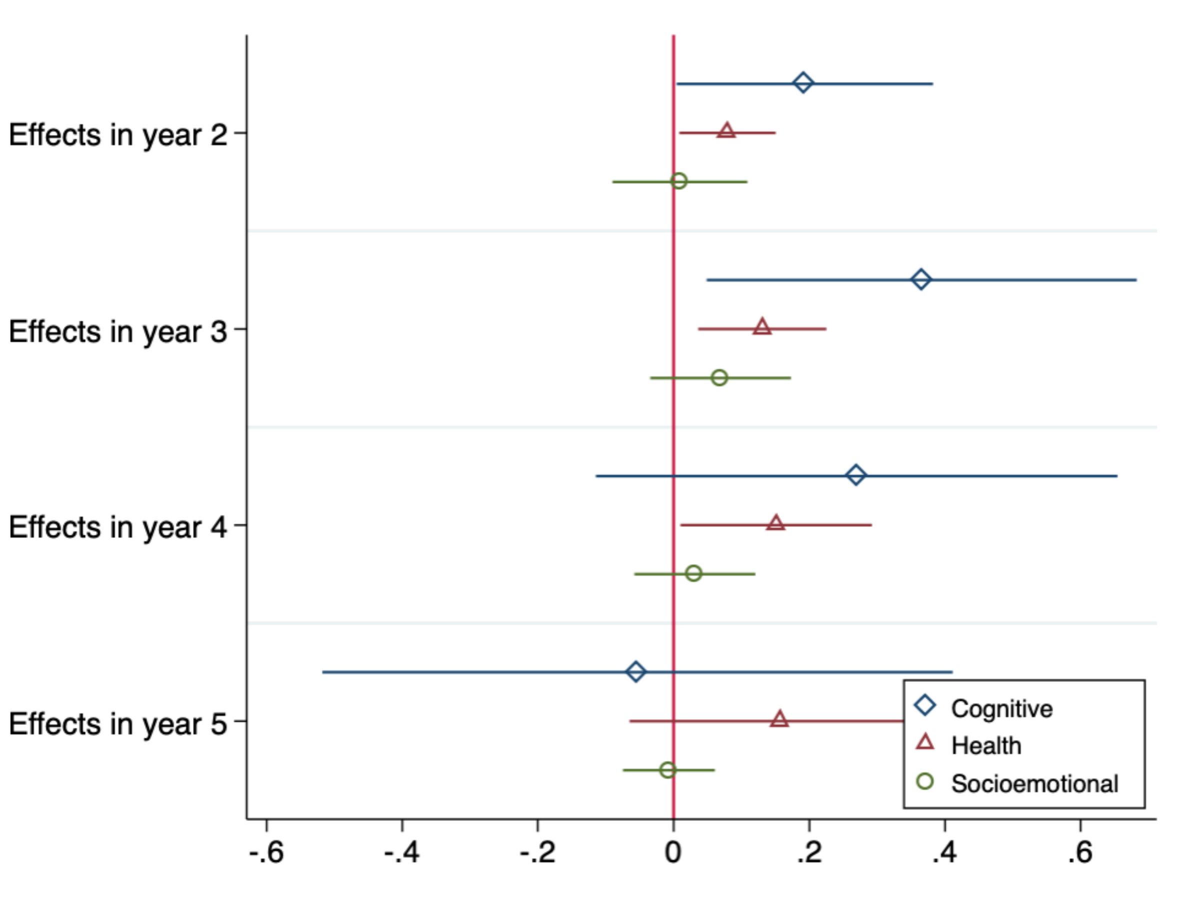
Notes: This figure plots the study timeline and yearly assessments.

Figure 2: aeioTU enrollment by study wave and lottery status



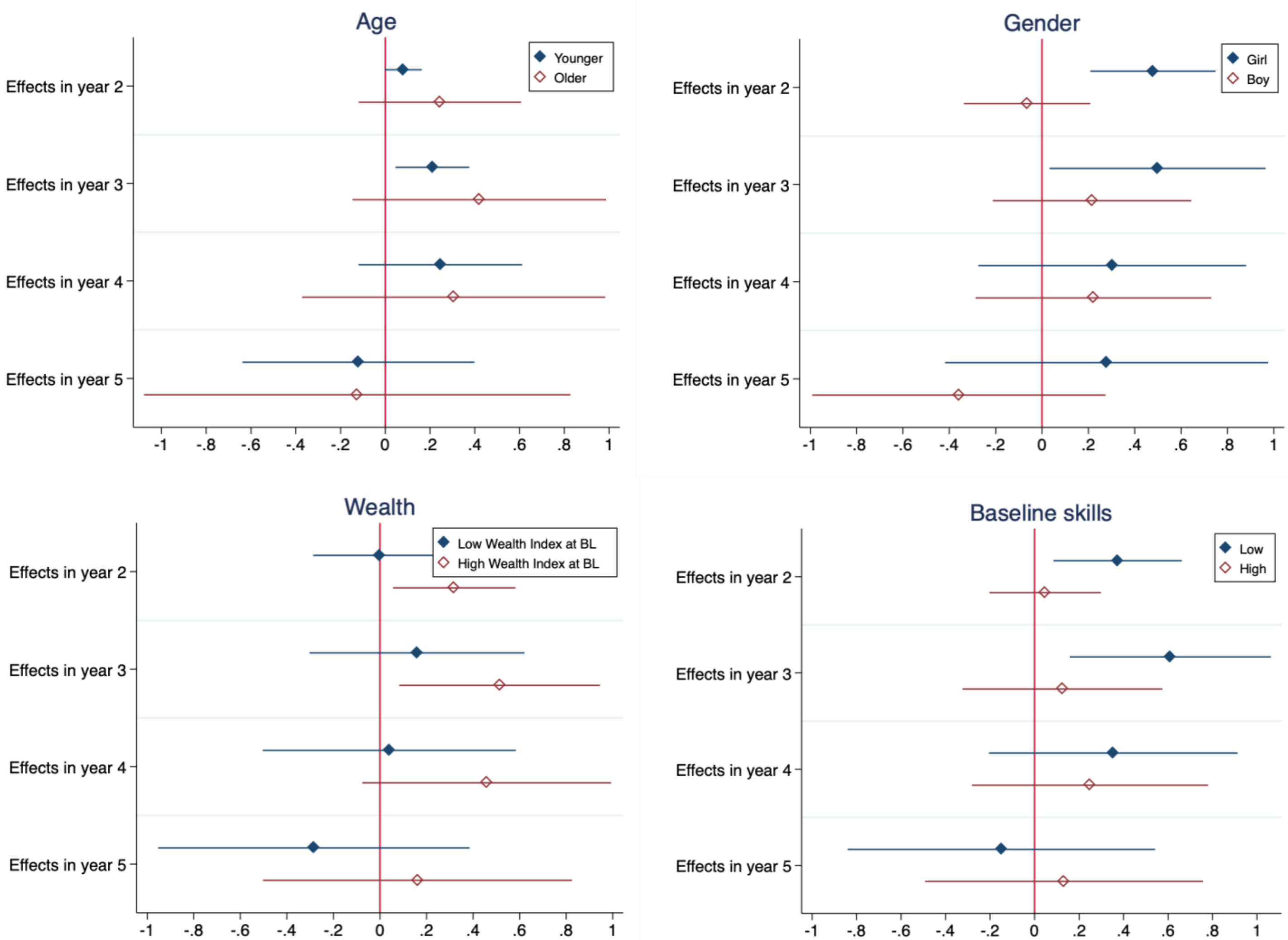
Notes: This figure plots the share of children enrolled in aeioTU centers by lottery assignment and study wave. The sample is restricted to age-eligible children as defined in the main text.

Figure 3: ITT effects



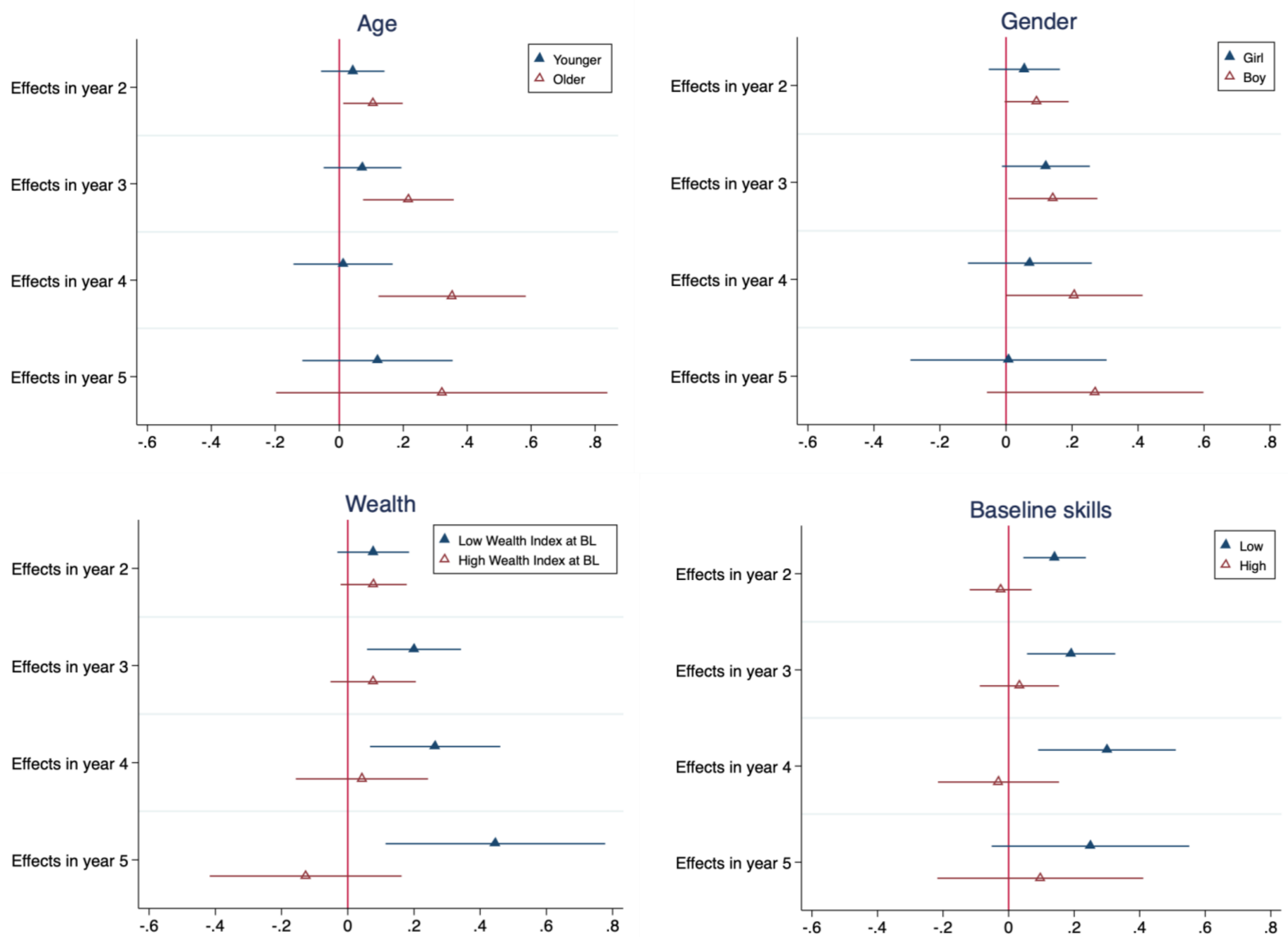
Notes: The figure reports ITT effects on child development with the corresponding 95 percent confidence intervals. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for whether the child had attended childcare prior to baseline. The sample is restricted to age-eligible children as defined in the main text.

Figure 4: Heterogeneous effects on child cognitive skills



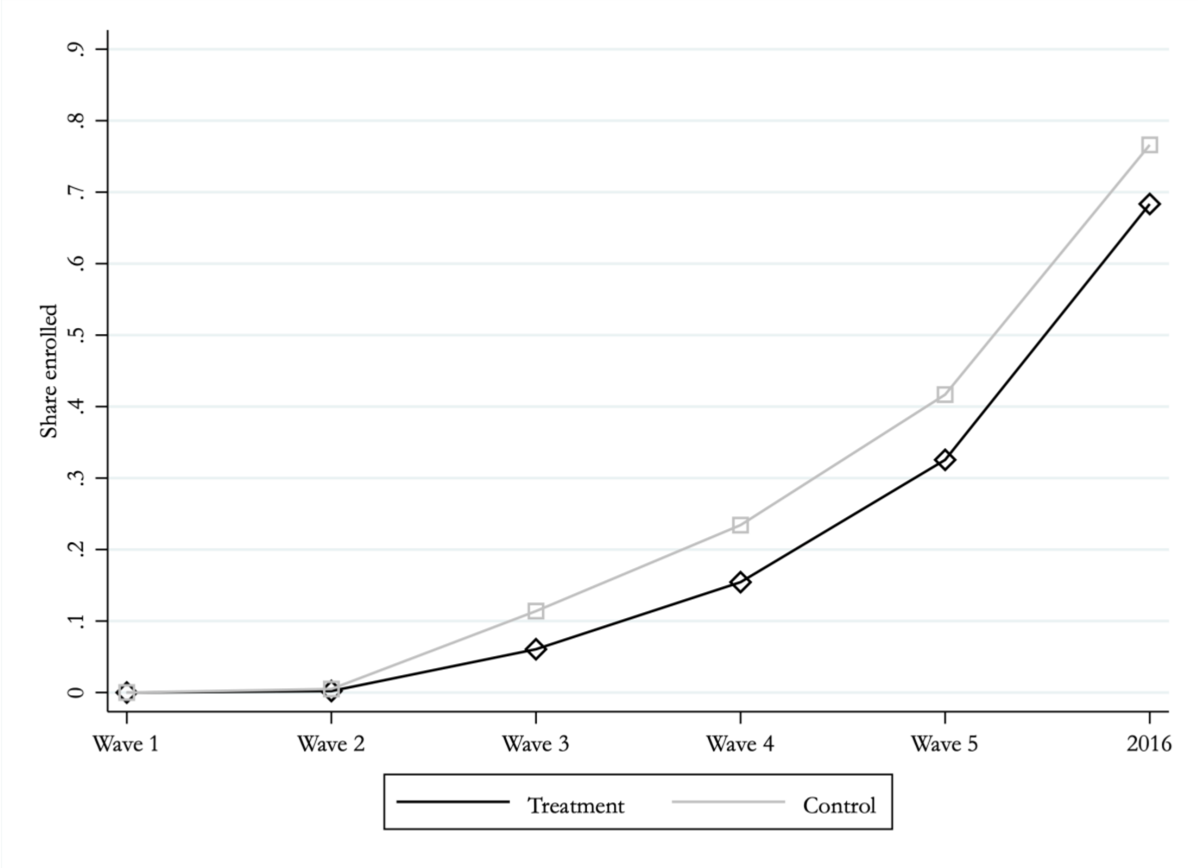
Notes: This figure plots treatment effects by subgroup. The top left panel compares younger children (0-2) and older (2-4). The top right panel compares boys and girls. The bottom left panel compares higher and lower SES children (household wealth above or below the sample median at baseline). The bottom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

Figure 5: Heterogeneous effects on child health



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

Figure 6: Primary school enrollment over time by treatment group



Notes: This figure plots the share of children enrolled in primary school by treatment group. 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 (Total)	Eligible in Y2 (%)	Eligible in Y3 (%)	Eligible in Y4 (%)	Eligible in Y5 (%)
Full sample	1,073	100%	99%	89%	63%
Cohort at baseline:					
0-1	190	100%	100%	100%	100%
1-2	321	100%	100%	100%	100%
2-3	308	100%	100%	100%	53%
3-4	254	100%	94%	52%	0%

Notes: The table reports the share of children that are age-eligible at each follow-up wave, separately for the full sample and by cohort.

Table 2: Measures of child development

	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	36+	36+	36+	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: The table reports the measures used to measure child development separately by domain and shows the age range (in months) for which the measure is available.

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	12.47	26.919	13.082	24.091	11.833	0.005
Gender (male)	1,073	0.523	0.5	0.522	0.5	0.523	0.5	1
Race (black)	1,073	0.615	0.487	0.631	0.483	0.603	0.49	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	0.43	0.986
Mother secondary complete +	1,073	0.364	0.481	0.363	0.481	0.365	0.482	1
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
Father secondary complete +	1,007	0.399	0.49	0.411	0.493	0.39	0.488	0.99
Wealth index	1,073	0.11	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	0.34	0.474	0.342	0.475	1
Panel B: Child development								
<u>Cognitive development</u>								
Bayley Cognitive	798	48.455	14.997	49.176	15.55	47.949	14.592	0.49
Bayley Receptive	790	19.284	7.838	20.065	8.087	18.738	7.621	0.1
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	46.89	13.4	47.288	14.029	46.608	12.946	0.538
TVIP	384	6.995	6.024	7.659	6.626	6.235	5.165	0.1
Woodcock-Munoz	244	89.148	9.75	88.28	9.117	90.17	10.395	0.419
ELSA	245	2.469	3.087	2.649	3.125	2.263	3.043	0.538
<u>Health</u>								
Weigh-for-age Z-score	1,046	-0.365	1.031	-0.427	0.99	-0.317	1.061	0.208
Height-for-age Z-score	1,048	-1.123	1.08	-1.11	0.995	-1.134	1.143	0.71
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	15.27	1.133	0.345
<u>Socio-emotional development</u>								
ASQ Total Score	1,060	46.239	29.369	48.953	30.53	44.101	28.263	0.009

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

Table 4: ITT estimates on child outcomes

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Observations
Cognitive skills	0.193	0.022	0.366	0.012	0.27	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)	(0.01)	(0.072)	(0.038)	(0.113)	(0.081)	
Socio-emotional sills	0.009	0.571	0.069	0.904	0.031	0.753	-0.007	0.42	3,490
	(0.051)	(0.92)	(0.053)	(0.937)	(0.045)	(0.937)	(0.035)	(0.886)	

Notes: The table reports ITT effects on child development. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parentheses below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for 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: 2SLS estimates for the effect of one additional year of program exposure on child outcomes

Study year:	*Pooled*	Y2		Y3		Y4		Y5		
	Effect (SE)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Observations (F-statistic)
Cognitive skills	0.227	0.359	0.022	0.389	0.012	0.227	0.084	-0.039	0.588	3,418
	(0.127)	(0.177)	(0.062)	(0.171)	(0.046)	(0.164)	(0.154)	(0.176)	(0.600)	(29.851)
Health	0.129	0.146	0.015	0.136	0.004	0.124	0.018	0.116	0.084	3,484
	(0.054)	(0.067)	(0.085)	(0.05)	(0.019)	(0.059)	(0.020)	(0.084)	(0.074)	(30.976)
Socio-emotional sills	0.034	0.016	0.57	0.073	0.905	0.026	0.756	-0.005	0.42	3,490
	(0.030)	(0.093)	(0.63)	(0.056)	(0.989)	(0.038)	(0.989)	(0.025)	(0.528)	(31.423)

Notes: The table reports the results of a 2SLS model instrumenting cumulative program enrollment at each wave with interactions between lottery status and survey wave indicators. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parentheses below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for 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 for the first stage regression.

Table 6: 2SLS estimates for the effects of duration of program exposure on child outcomes

Years of program exposure:	1 year		2 years		3 years		4 years		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Observations (F-statistic)
Cognitive skills	0.358	0.022	0.777	0.009	0.762	0.05	-0.104	0.557	3,418
	(0.178)	(0.066)	(0.331)	(0.036)	(0.462)	(0.098)	(0.723)	(0.559)	(25.062)
Health	0.146	0.015	0.274	0.003	0.38	0.012	0.47	0.082	3,484
	(0.067)	(0.108)	(0.101)	(0.019)	(0.169)	(0.019)	(0.337)	(0.087)	(26.879)
Socio-emotional skills	0.016	0.568	0.139	0.904	0.096	0.827	0.003	0.512	3,490
	(0.094)	(0.577)	(0.107)	(0.989)	(0.101)	(0.989)	(0.114)	(0.554)	(28.762)

Notes: The table reports the results of a 2SLS model instrumenting years of program exposure with interactions between lottery status and survey wave indicators. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parentheses below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for 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 for the first stage regression.

Table 7: ITT effects on children's home environment

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Observations
Parental time	-0.004	0.912	0.021	0.571	-0.025	0.62	-0.037	0.517	3,510
	(0.04)	(0.946)	(0.038)	(0.946)	(0.05)	(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	0.01	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: The table reports ITT effects on the home environment. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parentheses below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for whether the child had attended childcare prior to baseline. The sample is restricted to age-eligible children as defined in the main text.

Table 8: Comparison of ITT effects from experimental evaluations of ECCE programs in LMICs

Study	Program	Country	Child age	ITT impact estimate (SE)		
				Cognitive	Socio-emotional	Health
	(1)	(2)	(3)	(4)	(5)	(6)
Attanasio et al. (2022)	Free childcare	Brazil	0-3	0.071 (0.037)*	0.004 (0.068)	0.138 (0.048)***
Berkes et al (2024)	Community-based preschool	Cambodia	3-5	0.054 (0.025)**	0.048 (0.038)	Not reported
Bernal et al. (2019)	Public childcare	Colombia	0.5-5	−0.110 (0.050)	−0.02 (0.050)	0.05 (0.020)
Bjorvatn et al. (2024)	Free childcare	Uganda	3-5	0.120 (0.06)**	0.040 (0.070)	Not reported
Blimpo et al (2022)	Community-based preschool	The Gambia	3-6	-0.160 (0.120)	Not reported	Not reported
Bouguen et al. (2018)	Community-based preschool	Cambodia	3-5	-0.064 (0.075)	-0.026 (0.051)	-0.017 (0.063)
Dean and Jayachandran (2020)	Private preschool	India	3-5	0.390 (0.061)***	-0.041 (0.140)	Not reported
Hojman and Lopez Boo (2022)	Free childcare	Nicaragua	0-4	0.040 (0.080)	0.18 (0.100)*	Not reported
Martinez et al. (2017)	Community-based preschool	Mozambique	3-5	0.171 (0.050)***	0.300 (0.222)	0.064 (0.062)

Notes: The table compares effects from experimental evaluations of center-based ECCE programs in LMIC. Column (1) describes the type of program, column (2) reports the country where the study took place, column (3) reports the age of target children. Columns (4) to (6) shows each program's estimated estimates and standard errors (SE) for the effect of the program on child cognitive skills, socio-emotional skills and health. When a study reports multiple effects for the same outcomes we report the largest effect. * denotes 10% significance, ** denotes 5% significance, and *** denotes 1% significance.

Table 9: Enrollment in primary school over time

	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	-0.08	-0.091		
	(0.002)	(0.000)		
Treatment effect, Wave 5	-0.091	-0.088		
	(0.016)	(0.013)		
Treatment, 2016			-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, 2016			0.734	
Controls	N	Y	N	Y
Observations	3755	3678	1073	1050

Notes: The table presents the results for a model of enrollment in primary education school. The outcome variable in columns (1) and (2) is primary 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 10: Comparing aeioTU to counterfactual alternatives

	Cared for at home	Alternative center-based childcare	Enrolled in school
Instruments	(1)	(2)	(3)
Offer x site, offer x cohort	-0.346	-0.310	0.714
	(0.003)	(0.018)	(0.000)
First stage F	1697.107	1272.774	2631.152
Offer x site, offer x cohort, offer x household characteristics	-0.329	-0.304	0.710
	(0.006)	(0.022)	(0.000)
First stage F	1527.574	1121.117	2295.157

Notes: The table presents 2SLS estimates of alternative care arrangements on child cognitive development. The excluded category is being enrolled in aeioTU. The model in the first row instruments the alternative care arrangements (home care, childcare center other than aeioTU, primary school) with interactions of the offer and site (neighborhood) indicators and interactions of the offer and child-cohort indicators. The model in the second row instruments the alternative care arrangements with interactions of the offer and site (neighborhood) indicators, interactions of the offer and child-cohort indicators, and interactions of the offer and household variables (whether the mother works, whether the mother is married, and whether the father lives with the child. All models control for main effects of the interacting variables. Standard errors are clustered at the child level. P-values for the coefficients and First-stage F-statistics 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 $m^{j}_{i,k,t}$ the j-th measure of skill of type k for child i at time t. We specify a linear relationship between the individual measures and the factors and write:

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

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

Because the latent factors have no natural location or scale, some normalizing restrictions are necessary for the identification of the factor model (Anderson & 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.

Appendix Table B1: Measures used to construct the cognitive latent factor

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

Notes:† Bayley scales for Infant Development 3rd edition. Subscales: cognitive, receptive language, expressive language, gross and fine motor.§ Woodcock-Muñoz III Tests of Achievement.‡ ELSA.

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

Appendix Table B2: Measures used to construct the socio-emotional latent factor

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 to construct the home environment latent factors

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				
Cohort	C		T		P-value	C		T		P-value	C		T		P-value	C		T		P-value
Full-sample	40	7%	32	7%	0.923	29	5%	31	7%	0.196	32	6%	39	10%	0.018	21	5%	17	7%	0.403
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%	0.42	4	3%	9	7%	0.203	2	3%	6	10%	0.071	0	0%	0	0%	1

Notes: The 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-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 table reports attrition from the study sample. 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 beneficary.

Appendix Table C3: Difference in baseline outcome variables by attrition

	Cognitive	Health	Socio-emotional
Interviewed in all waves	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: The table reports the difference in terms of baseline outcomes between attriting and non attriting children. P-values are reported below the point estimate.

Appendix D: Compliance

Appendix Table D1: Impact of lottery on enrolment in aeioTU

	Ever enrolled		Years enrolled	
	(1)	(2)	(3)	(4)
Treatment	0.304	0.307	1.130	1.154
	(0.000)	(0.000)	(0.000)	(0.000)
Mean enrolment control	0.4	0.4	1.118	1.118
Observations	1073	1073	1073	1073
Controls	No	Yes	No	Yes

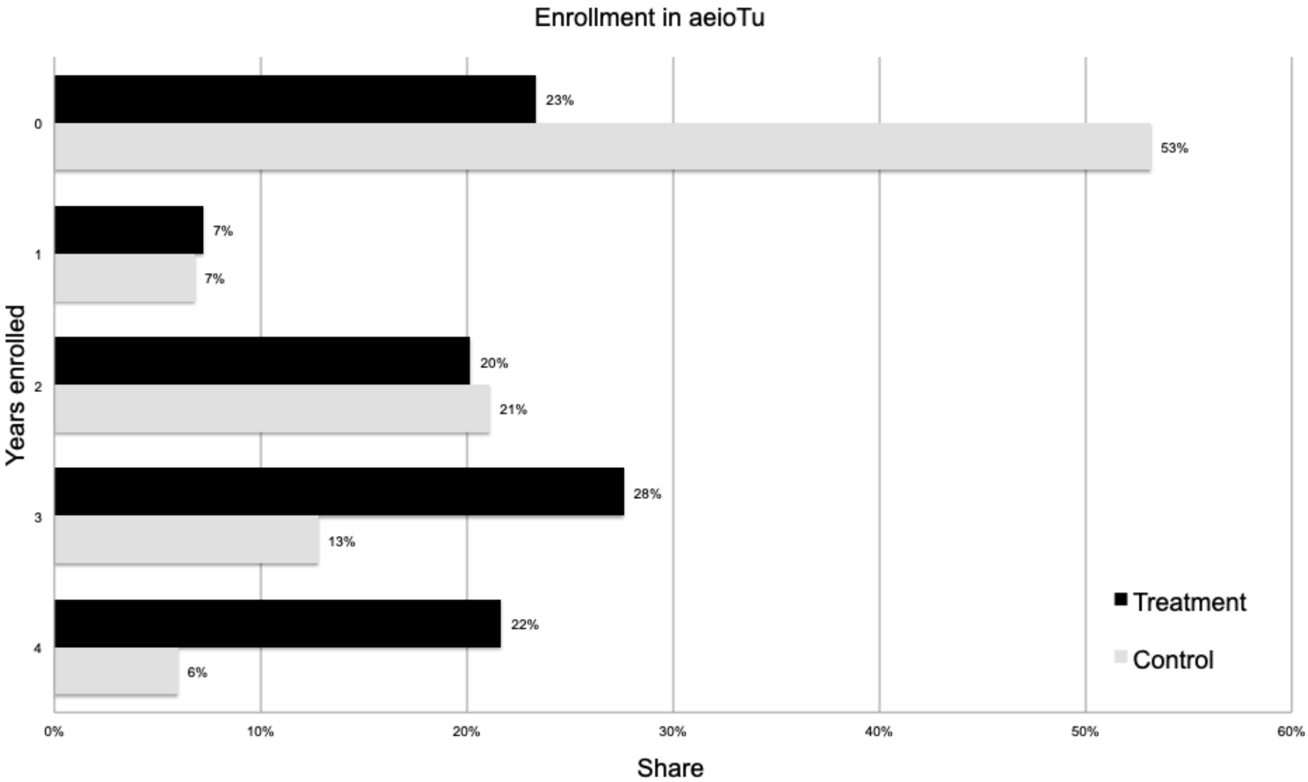
Notes: The table reports the effect of winning the lottery on enrolment in an aeioTU center. The dependent variable in columns 1 and 2 in an indicator for whether the child was ever enrolled in an aeioTU center. The dependent variable in columns 3 and 4 is the number of years the child attends aeioTU. Columns 1 and 3 only include randomization strata fixed effects. Columns 2 and 4 include the following additional child controls: 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. 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

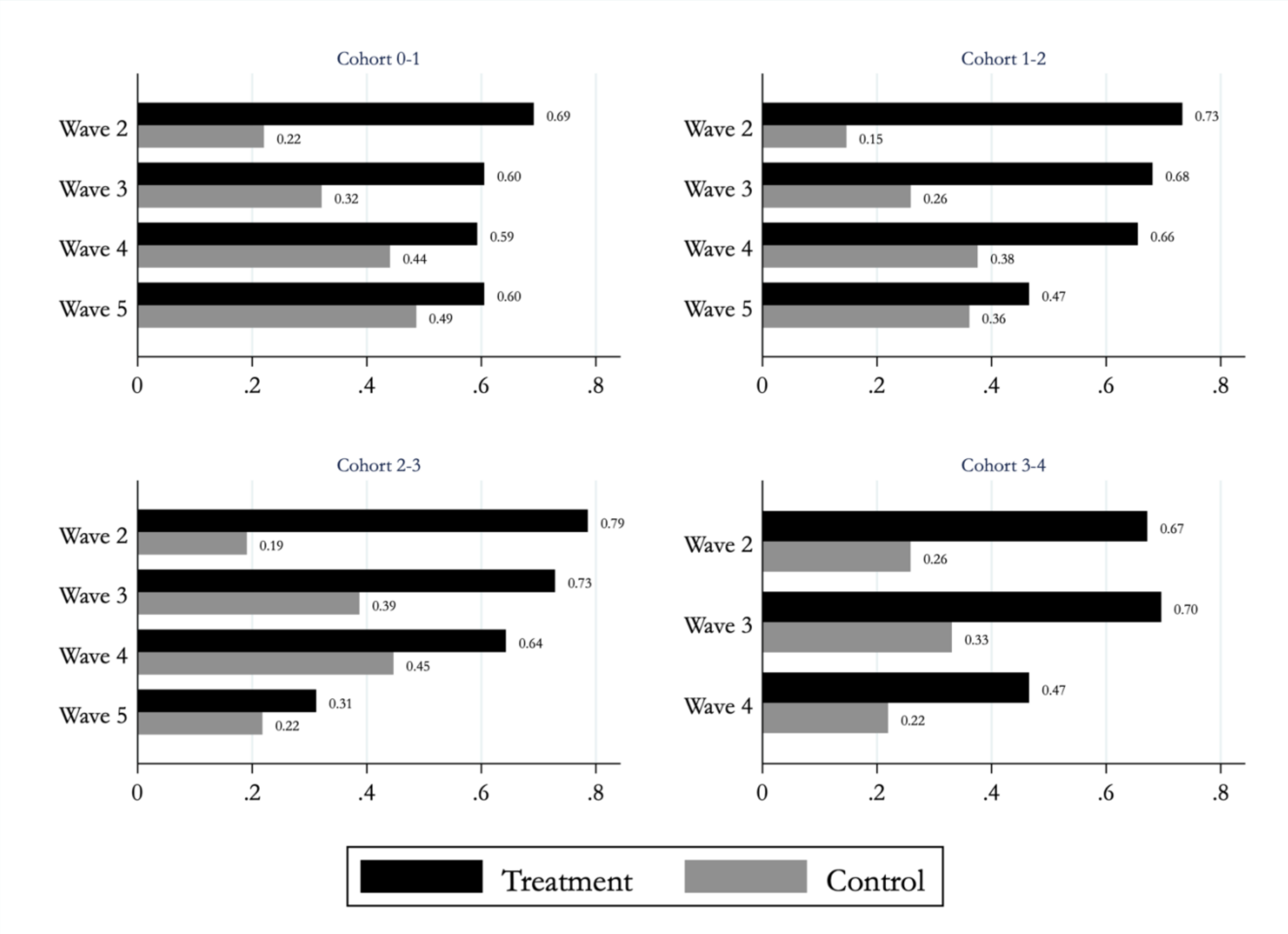
Notes: The table examines the determinants of compliance with lottery assignment. The dependent variable is a dummy for whether the child complies to initial random assignment. Standard errors are adjusted for clustering at the child level. P-values are reported below the point estimate. * CCT is Familias en Accion.

Appendix Figure D1: 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 treated (black) or control (grey) children who was enrolled in a center for a given number of years. The sample is restricted to age eligible children as defined in the main text.

Appendix Figure D2: Enrollment by study wave, lottery status and cohort



Notes: This figure plots the share of children enrolled in aeioTU centers by lottery assignment and study wave. Each subplot is a different cohort. 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 E: Robustness of ITT estimates

Appendix Table E1: ITT estimates on child outcomes without tester effects

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Observations
Cognitive skills	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	0.01	0.131	0.003	0.148	0.02	0.142	0.105	3,484
	(0.036)	(0.028)	(0.048)	(0.008)	(0.072)	(0.033)	(0.113)	(0.104)	
Socioemotional sills	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: The table reports ITT effects on child development. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for 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 estimates on child outcomes without controls

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Observations
Cognitive skills	0.169	0.032	0.385	0.01	0.29	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	0.14	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 sills	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: The table reports ITT effects on child development. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score. The sample is restricted to age eligible children as defined in the main text.

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

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Observations
Cognitive skills	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	0.05	0.128	0.092	0.06	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 sills	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: The table reports ITT effects on child development. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for whether the child had attended childcare prior to baseline. The sample is restricted to children that remain age eligible children throughout the study period.

Appendix Table E4: ITT estimates on child outcomes (50% eligible sample)

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Observations
Cognitive skills	0.179	0.03	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 sills	0.009	0.57	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: The table reports ITT effects on child development. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for whether the child had attended childcare prior to baseline. The sample is restricted to age eligible children (based on 50% of the time between two study waves) as defined in the main text.

Appendix Table E5: ITT estimates on child outcomes (full sample)

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Observations
Cognitive skills	0.175	0.035	0.366	0.012	0.4	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 sills	0.008	0.563	0.065	0.894	0.027	0.73	0.008	0.627	4,001
	(0.051)	(0.959)	(0.052)	(0.959)	(0.044)	(0.959)	(0.024)	(0.959)	

Notes: The table reports ITT effects on child development. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for whether the child had attended childcare prior to baseline. The sample includes all children who were initially in the lottery regardless of their age eligibility in any given wave.

Appendix Table E6: ITT estimates on child outcomes without tester effects (full sample)

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Observations
Cognitive skills	0.171	0.04	0.363	0.013	0.383	0.02	-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 sills	0.033	0.737	0.044	0.79	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: The table reports ITT effects on child development. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for whether the child had attended childcare prior to baseline. The sample includes all children who were initially in the lottery regardless of their age eligibility in any given wave.

Appendix Table E7: ITT estimates on child outcomes without controls (full sample)

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Observations
Cognitive skills	0.169	0.031	0.383	0.01	0.441	0.01	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 sills	0.018	0.645	0.074	0.922	0.032	0.766	0.01	0.673	4,001
	(0.050)	(0.987)	(0.052)	(0.987)	(0.044)	(0.987)	(0.023)	(0.987)	

Notes: The table reports ITT effects on child development. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score. The sample includes all children who were initially in the lottery regardless of their age eligibility in any given wave.

Appendix Table E8: ITT effects on individual tests

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	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)		

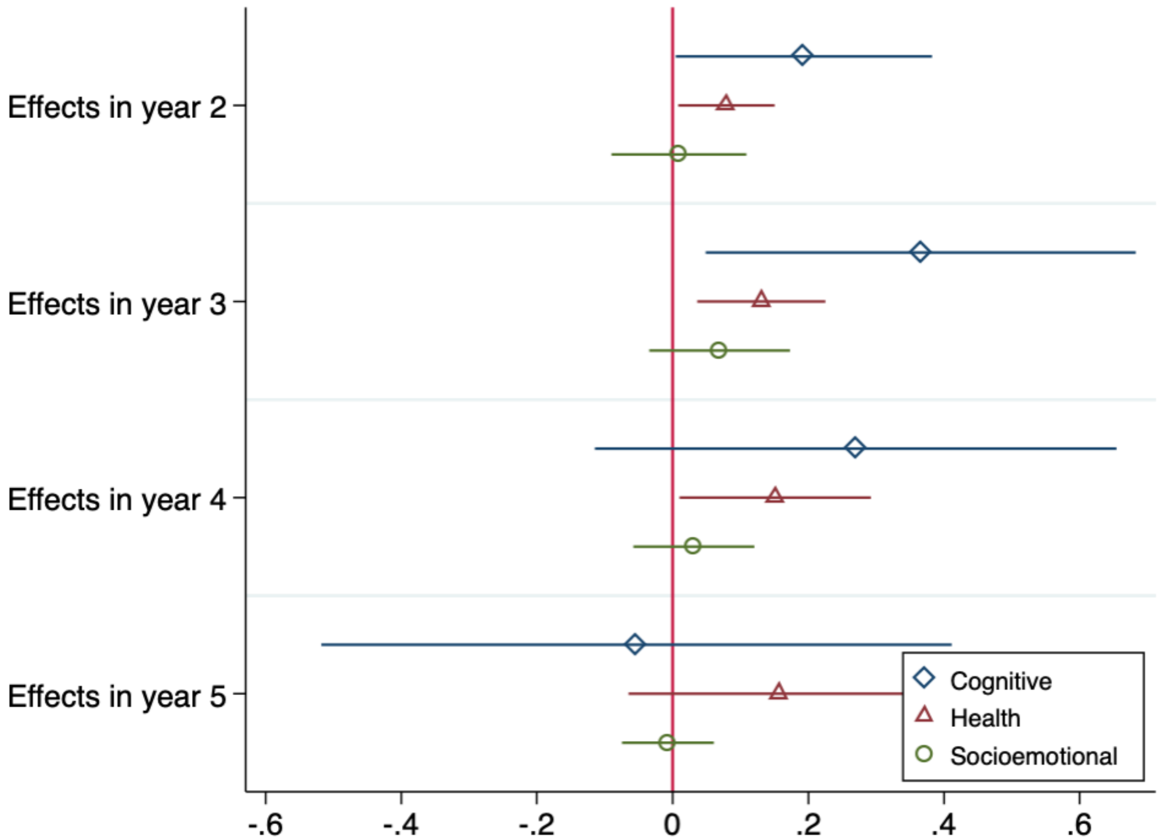
Notes: The table reports ITT effects on child development. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for 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 E9: ITT effects on health measures

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	Effect (SE)	P-Value (RW)	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: The table reports ITT effects on child development. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for whether the child had attended childcare prior to baseline. The sample is restricted to age eligible children as defined in the main text.

Appendix Figure E1: ITT effects



Notes: The figure reports ITT effects on child development with the corresponding 95 percent confidence intervals. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for whether the child had attended childcare prior to baseline. The sample is restricted to age eligible children as defined in the main text.

Appendix F: Back of the envelope calculation

We perform a simple back of the envelope calculation to understand the extent to which differences in childcare arrangements between treatment and control children, and their changes over time, can explain the observed dynamic in treatment impacts, and specifically the convergence in cognitive development in the last study wave.

We assume that outcome of children cared for at home has a mean (μ) 100 and standard deviation (σ) 10. Given this assumption we compute the implied outcomes of children in alternative care arrangements using the estimates from Table 10. For example, for children enrolled in aeioTU the outcome is $\mu + \sigma$ x 0.3 = 100 + 10 x 0.3 = 103, while for children enrolled in school this would be $\mu + \sigma$ x 1 = 100 + 1 x 10 = 110. Where -0.3 is the ITT estimate for the effect of being cared for at home compared to being in aeioTU, expressed in standard deviation (σ) units of the outcome variable from Table 10.

Using the observed enrollment patterns in different childcare and schooling arrangements over time (reported in Appendix Figure G5), we compute the difference in average outcomes between lottery winners and lottery losers, and re-express this difference in standard deviation units to compute ITT effects. We then compare these predicted effects with the estimated effects reported in Table 4. For example, in year 5 the average outcome of lottery losers is 0.1 x 100 + 0.36 x 103 + 0.13 x 100 + 0.42 x 110, and the average outcome of lottery winners is 0.12 x 100 + 0.47 x 103 + 0.09 x 100 + 0.33 x 110. So, the difference in average outcomes is 106.28 - 105.71 = -0.57, which is -0.057σ.

Appendix G: Additional Tables and Figures

Appendix Table G1: Correlations of factor scores with sociodemographic characteristics

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: The 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	13.02	24.159	11.834	0.009
Gender (male)	1,001	0.519	0.5	0.515	0.5	0.523	0.5	1
Race (black)	1,001	0.619	0.486	0.642	0.48	0.601	0.49	0.869
Mother is single	1,001	0.276	0.447	0.271	0.445	0.279	0.449	1
Mother works	1,001	0.238	0.426	0.23	0.421	0.244	0.43	0.997
Mother secondary complete +	1,001	0.369	0.483	0.369	0.483	0.368	0.483	1
Childcare use at baseline	1,001	0.17	0.376	0.203	0.402	0.144	0.352	0.169
Father lives at home	1,000	0.69	0.463	0.686	0.465	0.693	0.461	1
Father secondary complete +	940	0.406	0.491	0.42	0.494	0.396	0.489	0.99
Wealth index	1,001	0.063	4.657	-0.2	4.17	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	1.9	5.377	2.09	0.978
No. of children <= 5 yrs	1,001	2.676	0.796	2.772	0.84	2.601	0.751	0.012
CCT beneficiary	1,001	0.353	0.478	0.351	0.478	0.354	0.479	1
Panel B: Child development								
<u>Cognitive development</u>								
Bayley Cognitive	743	48.573	14.741	49.36	15.18	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	19.24	8.86	0.466
Bayley Fine Motor	739	32.591	9.462	33.17	9.78	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	6.76	6.429	5.22	0.134
Woodcock-Munoz	229	89.1	9.814	88.236	9.256	90.104	10.378	0.466
ELSA	229	2.485	3.13	2.686	3.157	2.259	3.098	0.509
<u>Health</u>								
Weigh-for-age Z-score	977	-0.362	1.036	-0.438	0.999	-0.302	1.061	0.1
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
<u>Socio-emotional development</u>								
ASQ Total Score	988	46.335	29.502	49.391	30.628	43.932	28.383	0.01

Notes: The table reports baseline child and household characteristics by treatment group for non-attritors. The last column reports two-sided Romano-Wolf p-values for the difference between treatment and control

Appendix Table G3: Heterogeneity by age

Study year:	Y2		Y3		Y4		Y5	
	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value
Cognitive: younger	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	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	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	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	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	-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: The table reports heterogenous ITT effects on child development by child age comparing younger (0-2) and older (2-4) children. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for 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 G4: Heterogeneity by gender

Study year:	Y2		Y3		Y4		Y5	
	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value
Cognitive: Girls	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: Boys	-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: Girls	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	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: Girls	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	-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: The table reports heterogenous ITT effects on child development by child gender. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for 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 G5: Heterogeneity by household wealth

Study year:	Y2		Y3		Y4		Y5	
	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value
Cognitive: Low wealth	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: High wealth	-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: Low wealth	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: High wealth	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: Low wealth	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: High wealth	-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: The table reports heterogenous ITT effects on child development by household wealth. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for 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 G6: Heterogeneity by baseline outcome

Study year:	Y2		Y3		Y4		Y5	
	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value
Cognitive: Low	0.32	0.072	0.542	0.006	0.323	0.083	-0.164	0.726
	-0.219		-0.218		-0.233		-0.273	
Cognitive: High	0.078	0.361	0.227	0.151	0.231	0.165	0.104	0.358
	-0.219		-0.22		-0.237		-0.287	
Difference	-0.242	0.205	-0.315	0.329	-0.092	0.812	0.268	0.563
Health: Low	0.12	0.046	0.212	0.001	0.21	0.003	0.102	0.122
	-0.071		-0.07		-0.075		-0.088	
Health: High	-0.014	0.564	0.02	0.407	0.069	0.231	0.197	0.04
	-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	-0.114	0.035	0.09	0.925	-0.013	0.422	0.007	0.536
	-0.063		-0.063		-0.068		-0.081	
Socioemotional: High	0.116	0.958	0.045	0.748	0.078	0.86	-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: The table reports heterogenous ITT effects on child development by baseline level of development. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered at the individual level are reported in parenthesis below the point estimate. One-tailed p-values and Romano-Wolf (RW) step-down p-values (in parentheses) are reported in the same column. All regressions include randomization strata and tester fixed effects. Covariates include child baseline score, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for 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 G7: Nutritional differences at baseline by child age

	Cohort 0-2		Cohort 2-4		
	Mean	SD	Mean	SD	P-value
Weigh-for-age	0.141	1.06	-0.568	0.962	0.000
BMI-for-age	0.612	1.011	0.337	0.923	0.000
Weigh-for-length	0.503	1.015	0.193	0.916	0.000
Height-for-age	0.978	1.168	-1.252	0.978	0.000
Arm circumference	14.93	1.062	15.652	1.017	0.000
Stunting	0.186	0.39	0.233	0.423	0.065

Notes: The table describe the nutritional status of the sample at baseline by child age. 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 child gender

	Girls		Boys		
	Mean	SD	Mean	SD	P-value
Weigh-for-age	-0.369	1.021	-0.362	1.042	0.917
BMI-for-age	0.425	0.934	0.504	1.009	0.193
Weigh-for-length	0.308	0.932	0.367	1.013	0.336
Height-for-age	-1.083	1.096	-1.159	1.065	0.253
Arm circumference	15.259	1.11	15.353	1.088	0.167
Stunting	0.203	0.403	0.218	0.413	0.564

Notes: The table describe the nutritional status of the sample at baseline by child age.

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

	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, 2016			-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	0.58			
Control mean, 2016			0.718	
Controls	N	Y	N	Y
Observations	4292	4200	1073	1050

Notes: The table presents the results for a model of enrollment in primary education school. The outcome variable in columns (1) and (2) is primary 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.

Appendix Table G10: Comparing estimated and predicted effects

Study year:	Y2	Y3	Y4	Y5
Estimated effects	0.193	0.366	0.27	-0.053
Predicted effects	0.159	0.261	0.086	-0.057

Notes: Thr table reports estimate ITT effect on children's cognitive development (from Table 4) and predicted effects based on our back of the envelope calculation described in the text. To compute the predicted effects, we multiply the 2SLS estimated in Table 10 with the corresponding shares reported in Appendix Figure G5.

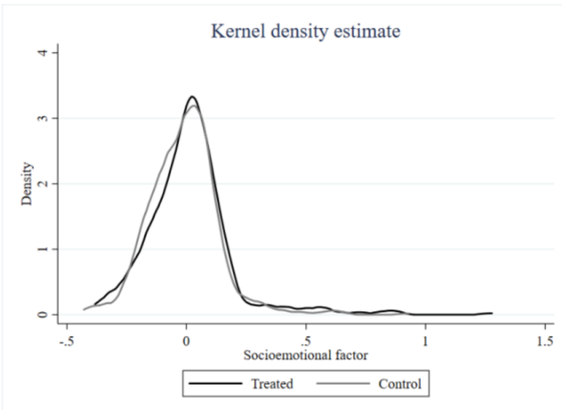
Appendix Table G11: ITT estimates on child outcomes excluding children enrolled in school

Study year:	Y2		Y3		Y4		Y5		
	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Effect (SE)	P-Value	Observations
Cognitive skills	0.172	0.035	0.328	0.021	0.322	0.06	0.374	0.079	2,878
	(0.094)		(0.162)		(0.207)		(0.265)		
Health	0.075	0.02	0.118	0.01	0.15	0.027	0.168	0.119	2,939
	(0.037)		(0.051)		(0.078)		(0.142)		
Socioemotional sills	0.006	0.55	0.06	0.851	0.006	0.543	-0.066	0.094	2,946
	(0.051)		(0.058)		(0.051)		(0.05)		

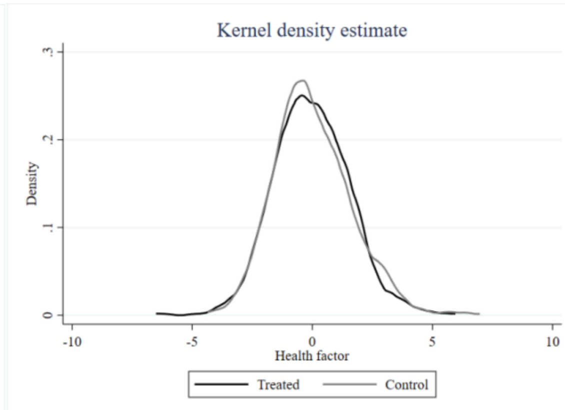
Notes: The table reports ITT effects on child development for children who are not enrolled in school. Each row corresponds to a separate regression for the outcome reported in the row header. Effects are expressed in terms of SD in the control group at baseline. Robust standard errors (SE) clustered 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, a second order polynomial in age, race, maternal education, household wealth, the number of children younger than five in household, and an indicator for whether the child had attended childcare prior to baseline. The sample is restricted to age eligible children as defined in the main text.

.

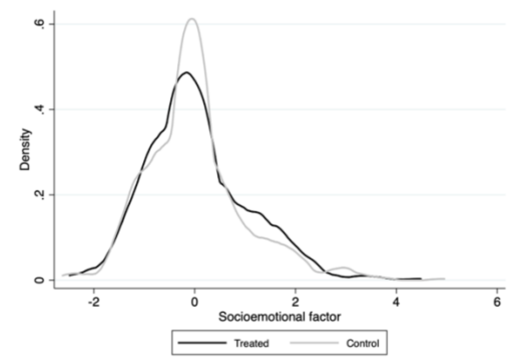
Appendix Figure G1: Distributions of factors scores at baseline by treatment group



Cognitive skills



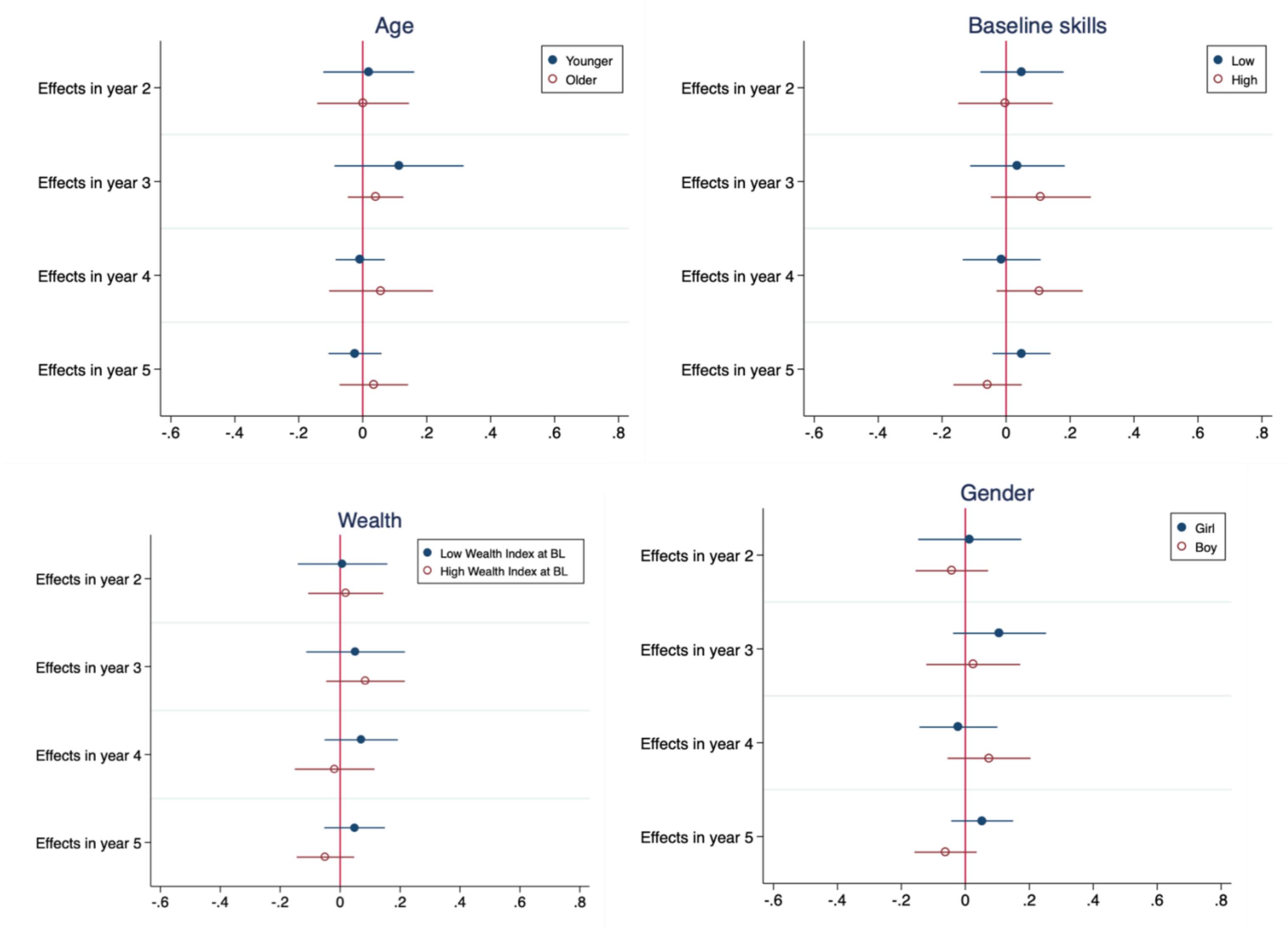
Health



Socio-emotional skills

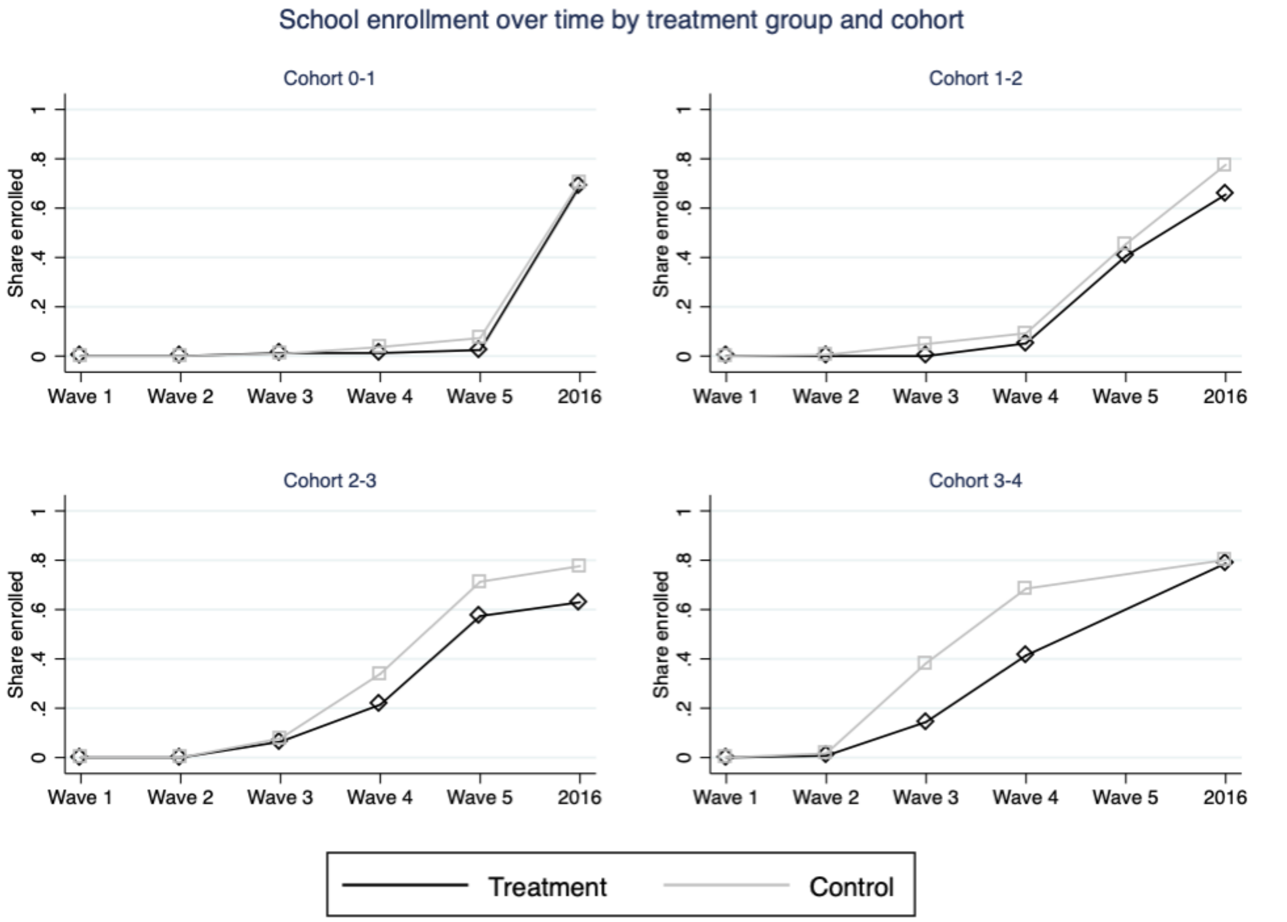
Notes: The figure plots the baseline distributions of factors scores by treatment group.

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



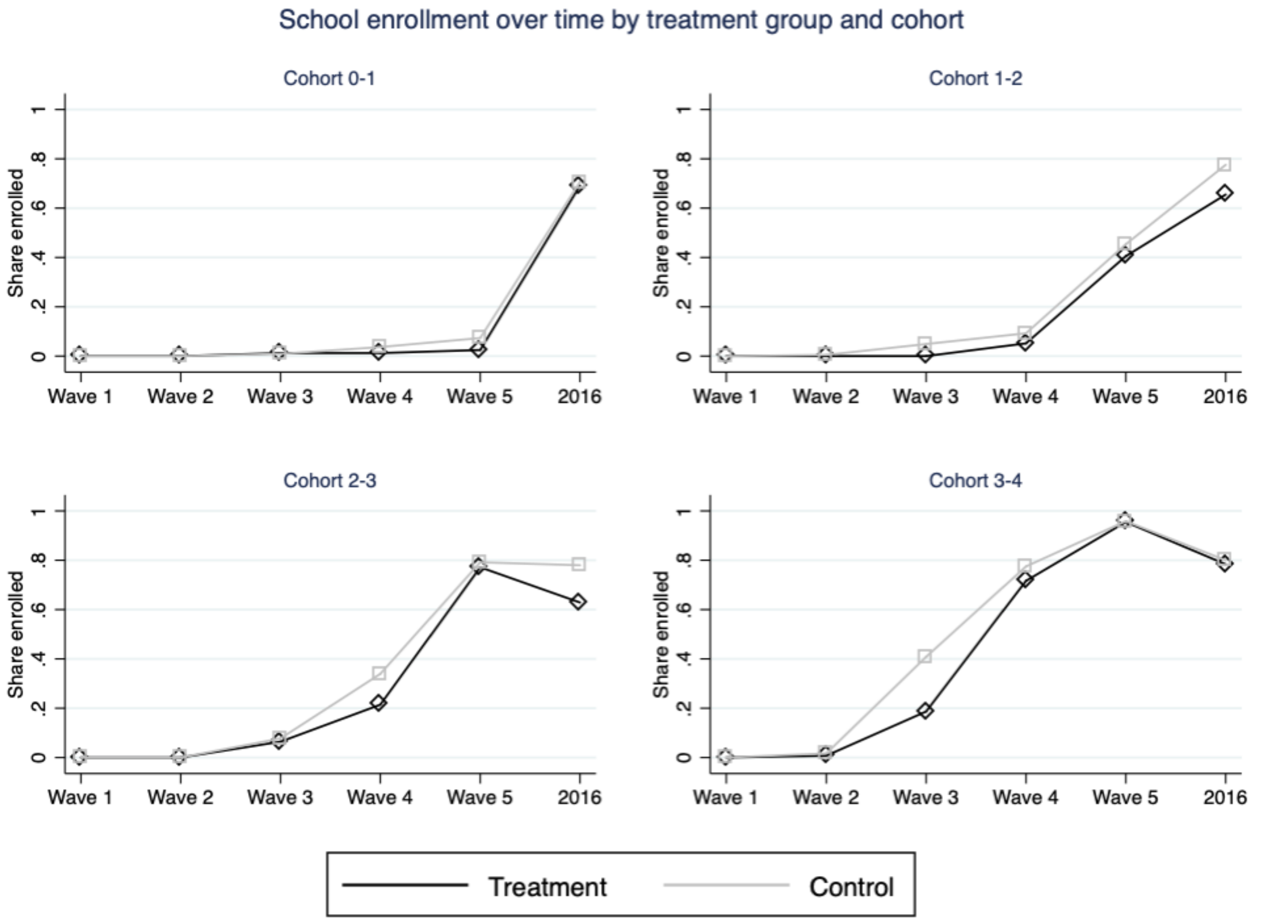
Notes: The figure plots treatment effects by subgroup. The top left panel compares younger (0-2) and older children (2-4). The bottom left panel compares higher and lower SES children (based on whether household wealth is above or below the median in the sample). The bottom right panel compares children 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 G3: Primary school enrollment over time by treatment group and cohort



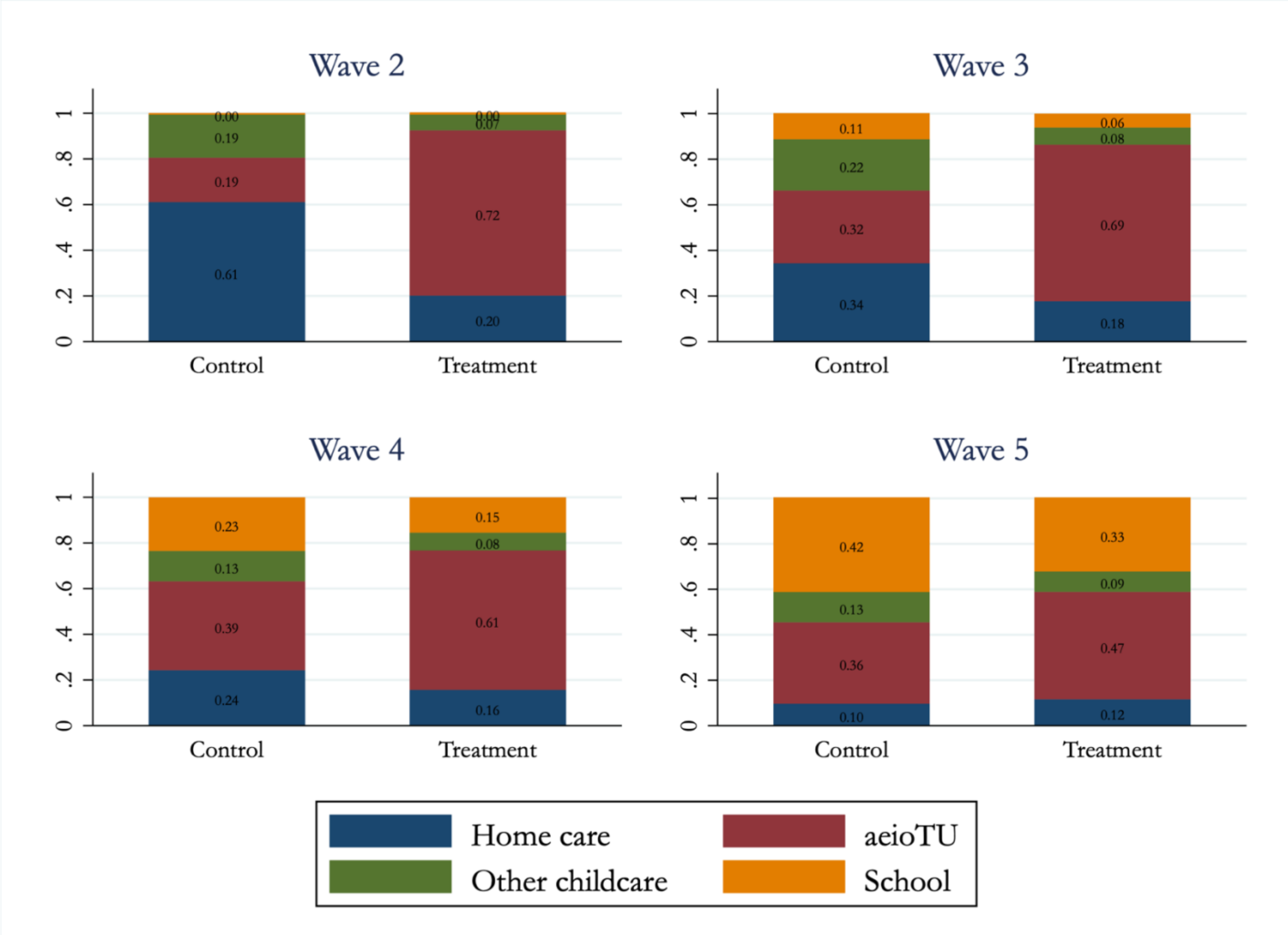
Notes: The figure plots the share of children enrolled in primary school by treatment group and child cohort of birth. The data for 2016 come from the administrative school records of the Integrated Enrollment System (Sistema Integrado de Matricula, SIMAT) of the Ministry of Education. The sample is 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 G4: Primary school enrollment over time by treatment group and cohort (full sample)



Notes: The figure plots the share of children enrolled in primary school by treatment group and child cohort of birth. 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 Figure G5: Counterfactual arrangements by treatment group and study wave



Notes: The figure plots the share of children enrolled in alterative childcare arrangements and primary school by treatment group and study wave. 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.