

CENTER-BASED CARE FOR INFANTS AND TODDLERS: THE AEIOTU RANDOMIZED TRIAL

Working Paper
October 2018

*Milagros Nores,
Ph.D., Raquel
Bernal, Ph.D, & W.
Steven Barnett,
Ph.D.*

*The National
Institute for Early
Education Research*

&

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

About the Authors

Milagros Nores, Ph.D. Dr. Nores is an Associate Director of Research at The National Institute for Early Education Research (NIEER) at Rutgers University. Dr. Nores conducts research at NIEER on issues related to early childhood policy, programs, and evaluation, both nationally and internationally. She has a Ph.D. in Education and Economics from Columbia University and an Ed.M. in Educational Administration and Social Policy from Harvard University. Her expertise and interests are in early childhood development, data-driven policy development, evaluation design, economics, cultural diversity, and English language learning.

Raquel Bernal, Ph.D. Dr. Bernal is a professor at the Economics Department at Universidad de los Andes in Bogotá, Colombia. Her research focuses on social policy, education, human capital, household decisions and labor economics. Bernal holds a PhD in Economics from New York University and MA and BA degrees from Universidad de Los Andes. Her research focuses on the determinants of the process of human capital accumulation and in particular, the determinants of ability at early stages of life.

W. Steven Barnett, Ph.D. Dr. Barnett is Director of the National Institute for Early Education Research (NIEER) and Board of Governors Professor at Rutgers University. His research includes studies of the economics of early care and education including costs and benefits, the long-term effects of preschool programs on children's learning and development, the economics of human development, practical policies for translating research findings into effective public investments and the distribution of educational opportunities. Dr. Barnett earned his Ph.D. in economics at the University of Michigan. He has authored or co-authored over 180 publications. Research interests include the economics of human development and practical policies for translating research findings into effective public investments.

Grateful acknowledgment is made to aeioTU for their commitment to early childhood and opening the doors to our evaluation team; iQuartil for their excellent work managing data collection on site; and our data collectors who worked under very difficult conditions. We also gratefully acknowledge the valuable research assistance provided by Cynthia van der Werf, Roman Zárate, María de la Paz Ferro and Santiago Lacouture. This research was supported by the Jacobs Foundation, the UBS Optimus Foundation and the Inter-American Development Bank. Any views expressed are those of the authors and do not necessarily represent those of the funders.

Correspondence regarding this report should be addressed to Milagros Nores at the National Institute for Early Education Research. Email: mnores@nieer.org.

Permission is granted to cite this material if you acknowledge NIEER and the authors of the item. Permission is not granted to print this material and this working paper should be considered a pre-print. For more information, call the Communications contact at (848) 932-4350, or visit NIEER at nieer.org.

Suggested citation: Nores, M., Bernal, R., & Barnett, W.S. (2018). Center-Based Care for Infants and Toddlers: The aeioTU Randomized Trial. New Brunswick, NJ: National Institute for Early Education Research.

Table of Contents

Table of Contents	5
I. Introduction	6
II. Background	8
III. Methods.....	11
<i>A. Study design</i>	11
<i>B. Sampling, randomization and masking</i>	11
<i>C. Theory of Change</i>	12
<i>D. Measures</i>	13
IV. Results.....	16
<i>A. Baseline characteristics</i>	16
<i>B. Attrition and Compliance</i>	17
<i>C. Estimations of Program Impact</i>	18
V. Discussion and Conclusions.....	20
References.....	24

I. Introduction

Global interest in public investments to improve the development of disadvantaged young children has exponentially risen in recent decades (Black et al. 2016; Nores and Barnett 2010; Berlinski and Schady 2016). Poverty compromises the development of hundreds of millions of children in the developing world at great cost to individuals and their countries (Black et al. 2016). Some studies find early intervention can alter such developmental trajectories (Berlinski and Schady 2016; Cunha et al. 2006; Engle et al. 2007). Today's global interest in the need for strong, high-quality, comprehensive programming for the youngest children is unprecedented, as evidenced with the inclusion of early childhood development as a target in the education goal (goal 4) under the sustainable development goals for 2030 (United Nations 2016). However, questions remain about whether and how best this might be done with public programs that could be scaled up.

In developing and low-income countries, there have been few comprehensive educational interventions and more focus on less costly nutrition, parenting and stimulation, or cash transfer interventions (Nores and Barnett 2010; Engle et al. 2007), especially for children under the age of three (Black et al. 2016). In addition, public policies have tended to favor increasing access over quality, resulting in relatively weak educational interventions (Araujo, López Bóo, and Puyana 2013). As a result, little empirical information exists regarding the effects of comprehensive (integrating education, care, nutrition and health) interventions providing high-quality early education and care to infants and toddlers in the developing world (Britto et al., 2017; Black et al., 2017; Behrman and Urzúa 2013). Infants and toddlers (less than three years old) have been previously underrepresented in studies of large-scale child care in the developing world.

Research in the United States suggests that intensive, high-quality early educational interventions affect cognitive and socioemotional development (Cunha et al. 2006; Barnett 2008; Camilli et al. 2010). Randomized trials find effects on intelligence, subject matter knowledge and skills, pro-social and anti-social behaviors, executive function, delinquency and depression that sometimes, but not always, persist. An important limitation of the U.S. studies is that control groups may access somewhat comparable services. In addition, even the U.S. literature lacks rigorous research on center-based education and child care programs for children birth to five on a large scale (Camilli et al. 2010).

More broadly, a meta-analysis of interventions outside the U.S. and Canada found that interventions with a mix of education and nutrition had larger effects on cognition than cash transfers or nutritional interventions alone (Nores and Barnett 2010).¹ This suggests that inadequate nutrition and inadequate cognitive stimulation both contribute to poor cognitive development, possibly synergistically (Kagitcibasi, Bekman, and Goksel 1995; Chang et al. 2002).

Engle et al. (2011) reviewed research on center-based early childcare and education in low-income countries and found significant impacts on children's cognitive and socio-

¹ Nores and Barnett (2010) include interventions varying from prenatal to age 7: 17 contrasts only included children under 36 months, 14 contrasts included children ages 3 and 7, and seven contrasts included children in both age groups.

emotional development across programs (Engle et al. 2011). The authors concluded that center-based early interventions improve children's cognitive functioning, school readiness and performance, with greater effects for disadvantaged children and in higher-quality programs. Quality might refer to structural features such as class size, child–adult ratio, teacher qualifications, and the physical environment, or other process features such as teacher–child interaction and the environment in which children learn (Yoshikawa et al. 2015). Among the center-based education and child care program evaluations reviewed, only two included toddlers and none included infants.

In Latin America and the Caribbean, public investment in early childhood services has also been increasing recently. However, there is vast heterogeneity in coverage, content, funding, quality, and staff qualifications for programs serving children under age three. Although there has been growing research in the region on the impacts of home-based child care including infants and toddlers (Bernal and Fernandez, 2013; Behrman et al., 2014; Attanasio, Di Maro and Vera-Hernandez, 2004), as well as some research on the effects of center-based early care for children older than age three (Berlinski and Galiani, 2007; Berlinski, Galiani and Gertler, 2009), the evidence on the impacts of center-based child care for infants and toddlers is scarce and inconclusive. An evaluation of a government subsidized center-based day-care program for poor children in Ecuador found negative effects on various dimensions of child development for children older than age three and null effects for the sample of children between zero and six years of age. The authors did not analyze the sample of infants and toddlers separately and propose poor quality of services as the explanation for the results (Rosero and Oosterbeek, 2011).

Bernal et al. (2018) looked into the effects of the transition of children from home-based child care to center-based child care in Colombia for children 0-5 years of age. The study found positive effects on children's health but negative impacts on children's cognitive outcomes possibly associated with the low process quality during the transition. The authors did not study separately the subsample of toddlers and infants.

Finally, Araujo, Dormal and Schady (2017) study the effects of caregiver-child interactions on children younger than two in center-based care in Peru. They find that children with caregivers who exhibit higher-quality interactions showed better development outcomes relative to children in the same centers whose caregivers exhibited low-quality interactions.

In this paper, we investigate the effects of aeioTU, an intensive and comprehensive center-based early education intervention, on the development of disadvantaged infants and toddlers in Colombia using a randomized control trial with a sample of 848 children under the age of three in two communities in northern Colombia. To our knowledge, this is the only study in a developing country that would allow the comparison of outcomes of infants and toddlers randomly assigned to receive high-quality center-based care to a randomly assigned control group. The key features of quality embedded in the aeioTU model and the growth of the program to over 13,000 children throughout the country via public-private partnerships, make of it an interesting case study. Particularly so given the financial and human capital constraints in Colombia. The region has been increasing early childhood development (ECD) investment over the last years. However, mostly parenting, rather than center-based interventions have been at the core of the such growth in services for children

under age three (Black et al. 2016). This study contributes to the understanding the feasibility of growth in center-based care, while maintaining an emphasis on quality.

aeioTu is similar in design to the Abecedarian program, a randomized trial found to produce persistent cognitive gains for disadvantaged children in the U.S. The Abecedarian was an intensive center-based early care intervention starting at 8 weeks and through age 5. It provided 40-45 hours of service per week during 50 weeks per year. Researchers at the University of North Carolina in Chapel Hill closely monitored the intervention. The Abecedarian followed 121 at-risk infants born between 1972 and 1977 in North Carolina. Program participants showed significant effects in high school graduation, higher education enrollment, skilled job employment, and various health outcomes (Barnett and Masse 2007; Muennig et al. 2011; García 2016; García et al.; García, Heckman and Ziff, 2018). However, the sample in our aeioTu study is roughly eight times as large that in the Abecedarian study, with aeioTu having expanded to serving 13,300 children annually in 25 cities across Colombia. Although far from being a national scale program, it operates at scale, in multiple cities, with multiple communities and facing barriers and opportunities beyond small trials.

We measure anthropometry, language, cognitive, motor, and socio-emotional development, and parenting. The results presented in this study contribute to the existing literature and are policy-relevant for the following reasons: (a) the evidence of positive impacts for children from less than a year in the program is highly relevant to governments that must consider what can be assured in practice at scale and in a context where most interventions for this group age are as short term as 8 months or shorter (Black et al. 2016; Nores and Barnett 2010), (b) the scarcity in the literature on center-based care for children under the age of 3 and how little we know on what are the effects of center-based care on children in this age-group, makes the focus on infants and toddlers particularly relevant, something we can do within the scope of this study and given its sample size, and (c) the study speaks to the importance of quality of early education given that the aeioTU model features many relevant factors related to child care quality and operates at a sizable scale.

We find positive effects on language (0.20 standard deviations-SD), cognitive development (0.14 SD), motor development (0.09 SD) and overall development (0.12 SD) eight months into the program. Effects are observed only for girls and not for boys. The effects for girls reach up to 0.33 SD in language. No intervention effects were observed for nutritional outcomes, socio-emotional development or the home environment.

The next section will briefly describe the early childhood policy context in Colombia and how the aeioTU programs fits in this framework. In Section III we describe the study design, the sample, the data and the empirical strategy and in Section IV we present the results of the study and Section V concludes.

II. Background

Colombia has grown consistently over the last decade with growth slowing down only recently, and income inequality only now starting to decline (Solt, 2011). Early childhood and family support policies in Colombia aim to ameliorate inequality and have been led by the Colombian Institute of Family Welfare (*Instituto Colombiano de Bienestar Familiar*, ICBF). Public early childhood care and education coverage for socioeconomically

vulnerable children² under the age of 6 in Colombia remains low (40%), while not low for the Latin American context, and is gradually expanding (World Bank 2013; Bernal and Camacho 2011), in a context where children from low-income households show developmental gaps as early as 12 months of age and these are about one standard deviation by age 5, relative to high-income children in language and cognition (Bernal, Martínez, and Quintero 2015).

Of the 4.3 million children in Colombia younger than six years of age (2.8 million younger than three), about 65% are socioeconomically disadvantaged and 40% of these are served by government programs (Bernal and Camacho 2014). Enrollment rates in public child care programs is close to 30% for children aged three or less, and about 60% for children four to five years of age. In particular, 800,000 vulnerable children younger than six are served through *Hogares Comunitarios de Bienestar* (HCB), a public home-based child care program that provides home-based care services and supplemental nutrition to low-income children under the age of six in the provider's own home. Recent evaluations of the HCB program (Bernal and Fernández, 2013; Attanasio, Di Maro, and Vera-Hernández 2013) have found positive impacts on children's developmental outcomes of about 0.15 standard deviations (SD), despite it being originally conceived as a program to promote female labor supply, and the low quality of the service provided (Bernal and Fernández, 2013).³ In this context, in 2011 the government launched a national early childhood strategy, "*De Cero a Siempre*" (From Zero to Forever), aimed at improving the quality of existing early childhood services as well as increasing enrollment in comprehensive services⁴ for 1.2 million children (Comisión Inter-Sectorial para la Primera Infancia, 2013; Bernal and Ramírez, 2018). As a result, center-based care enrollments grew from 125,000 children in 2011 to about 370,000 children in 2014. The aeioTU program is part of this strategy.

aeioTU is an NGO operating 28 centers by 2016 providing comprehensive early childhood education to about 13,300 low-income children aged 0–5 throughout urban Colombia. The program, described in detail below, features characteristics relevant for quality in early care and education. Since it started operations in 2009, aeioTU has grown through public-private partnerships with the national government and has recently been internationally highlighted as a successful innovative approach in a 2017 white paper from the World Economic Forum and its CEO was awarded the 2018 Klaus J. Jacobs Awards for Social Innovation and Engagement. The government provided a stipend equivalent to USD

² Socioeconomic disadvantage or vulnerability is measured in Colombia using SISBEN scores (a proxy means-indicator based on a household socio-demographic survey). Approximately 65% of children under the age of 6 are considered socioeconomically disadvantaged under this indicator, and are, thus, eligible for welfare services. We refer exclusively to this population throughout this study.

³ This seemingly contradictory result might be due to the fact that in low- and middle-income countries even very basic public pre-school especially for children from deprived backgrounds could have positive impacts, given the very low quality or unavailability of alternative options.

⁴ We define comprehensive child care services here refers to programs that embed pedagogical contents aimed at stimulating cognitive and socio-emotional development and do not simply provide a safe environment to care for the child while the mother works. Specifically, comprehensive services would offer concurrently nutrition, health, care and early education.

1,500 per child per month,⁵ which aeioTU supplemented with an additional 20–30% of own resources at the time of the study.⁶ The aeioTU program provides full-day (9 hours per day) educational care during 11 months of the year, with relatively low child-to-teacher ratios (8:2 for infants, 12:2 for toddlers at the time of this study), high teacher qualification requirements (32% had a BA and the rest had a vocational degree in early childhood education when we started this evaluation), and concerted pre- and in-service training (120 hours pre-service, over 130 hours in-service). Bernal et al. (2018) report that teacher training and coaching strategies were not common and varied significantly across service providers in most comparable public center-based programs at the time this study began, they also report lower educational levels of teachers and a child-to-teacher ratio of 25:1 for toddlers.

The aeioTU program also provides 70% of children’s daily nutritional requirements through breakfast, two snacks and lunch, which is mandatory in all public child care public programs in Colombia,⁷ and also provides regular nutritional monitoring.⁸ aeioTU’s educational program, inspired by the Reggio Emilia approach,⁹ features project-based learning and a balance of teacher-directed and child-initiated activities. Daily activities are guided and structured through specific pedagogical guidelines and group planning sessions.¹⁰ It is important to note that there is no specific curricular guideline for early education in Colombia. The “*De Cero a Siempre*” strategy has emphasized the principle of curricular freedom, and national standards are intentionally broad. Bernal et al. (2018) report that most comparable public center-based programs did not use a structured curriculum or had clear pedagogical guidelines for teachers’ daily activities at the time of this study.

⁵ Using the average COP/USD exchange rate in 2010, at the time this study began.

⁶ The additional funding provided by aeioTU was used for teacher training and the nutritional component of the program, which was underfunded by the governmental stipend, and provided nutritional supplementation over the holidays.

⁷ However, according to aeioTU, this component was underfunded by the government’s stipend in about 20-25%, which was covered with aeioTU’s own resources, and the program also provided nutritional supplementation (micronutrients) during holidays. We did not collect any data to monitor how the nutritional component was actually implemented in centers.

⁸ An in-site nutritionist periodically monitors children’s nutritional status. Children found to be at risk are referred to public health services and the center would adjust the nutritional supplement as recommended by the nutritionist. This was not mandatory for all public child care services at the time of this study.

⁹ The Reggio Emilia Approach is an education philosophy for pre-school and primary education. It is based on the notion that children are capable of constructing their own learning process through their innate curiosity to understand the world. The basic principle is that children learn about themselves and their context through interactions with others and their environment. Thus, adults are mentors and guides of this process rather than mere caregivers or providers of knowledge, in the sense of providing opportunities for children to explore their own interests. The approach recognizes many ways to understand the world and express thoughts, and aims at promoting these communication channels within the educational experience, including art, music, dance, movement, pretend play and exploration.

¹⁰ In line with the Reggio Emilia philosophy, aeioTU emphasizes on the collegial work of center personnel, the presence of the atelier (artist), and a pedagogical coordinating team.

Finally, aeioTU staffed centers with a team of experts including the atelier (in-site artist) and a pedagogical coordinator (director), who played a critical role in the planning of pedagogical guidelines in centers.¹¹ In line with the Reggio Emilia approach, aeioTU emphasizes on the participation of families. In particular, aeioTU holds regular workshops for parents, keeps close contact with families to inform them about the activities their children experience and their progress (through weekly reports and photos), and keeps an open-door policy, which also includes the use of the centers' recreational areas by families during weekends.

In sum, aeioTU's features such as pre- and in-service training, the use of a structured developmentally-oriented curriculum, and high qualification requirements for staff were uncommon among service providers at the time of this study. At the same time, these features are often thought as critical for early care and education quality and linked to better child developmental outcomes (Yoshikawa, Weiland and Brooks-Gunn, 2016; Barnett and Boocock, 1998; Bowman, Donovan and Burns, 2001; Bernal, 2015).¹²

III. Methods

A. Study design

We conducted a randomized controlled trial with families of young children assigned to treatment or control in two early care centers in one city in northern-coastal Colombia. In Figure 1 we show the study's flow chart for sample selection. Children were first assessed in late 2010, prior to random assignment and the beginning of the intervention, and assessed again about eight months post-treatment. Site 1 received baseline in July-September 2010, started intervention in November 2010, and received post-test between June and September 2011. Site 2 received baseline in October-December 2010, started intervention in March 2011, and received post-test between October and December 2011.

B. Sampling, randomization and masking

The sites were selected from the centers being opened by aeioTU around the time the study was planned and funded (2009-2010). Sites had to fulfill two criteria for inclusion: size (no small centers, so that we could power the study for children at different ages) and oversubscription (so that a lottery could be drawn). Two early childhood centers were opening in 2010 in northern Colombia, in two different communities that were deemed by aeioTU as suitable due to their socio-economic vulnerability.¹³ We identified all children under the age of 5 living in these two communities at the time the centers were being built through a census run through door to door visits and through community leaders, for a total

¹¹ At the time of this study, the hiring of these experts was not a national requirement.

¹² We do not have data to confirm that aeioTU had better quality than comparable center-based child care programs targeting the same population. We can only specify that the inputs often linked to center-based quality were higher, on average.

¹³ Other considerations for aeioTU's location choices included the political will of local mayors who often provided the infrastructure as well as approval of ICBF which prioritized underserved areas.

of 1,288 children (see Figure 1). All families were income-eligible for aeioTU by SISBEN scores, and all families expressed interest in enrolling in aeioTU centers if offered a slot.

The evaluation study was designed as a randomized controlled trial based on an oversubscription model. Not all children identified in the census participated in the randomization. Seventy children were excluded from the sample for reasons detailed in Figure 1. In particular, n=66 children were directly offered a center slot for various reasons including sons and daughters of teachers, n=2 children who moved out of the city prior to the opening of centers, and n=2 children outgrew the program between the census and center opening. This rendered a final sample of N=1,218 children, out of which 819 were toddlers. All 1,218 families agreed to participate in the study through written active informed consent.¹⁴

Each aeioTU center in the study had capacity for about 320 children aged 0–5, with just over half of that for children up to age 3. Families were randomized to treatment and stratified by age (five groups), gender and neighborhood within site (three groups) right after baseline data was collected. Slots were randomized for 819 infants and toddlers; 337 were allocated to enroll in the center (the treatment group) and 482 to the control group. We used computer generated random lists to assign children to treatment and control. We did this in a public event, within the community, with all the families present. Power analyses indicated a power of 0.85 for sample sizes of 700 with $\alpha=0.05$ and an expected effect size of 0.25 SD following Nores and Barnett (2010), allowing for an attrition rate of 17%.

Lottery winners were offered a slot to enroll in aeioTU centers. aeioTU followed-up with lottery winners' parents for effective registration of the children. However, some lottery winners decided not to enroll their children in centers even after several calls and visits. Given that our sample includes the universe of children 0-5 years of age residing in these communities, centers had to resort to children in the control group to complete enrollment.¹⁵ More details on compliance and cross-over are discussed later in section IV.

Participants were not blind to their participation in the program. This was not feasible. Child assessors and parent interviewers were blind to treatment status of participants. Realistically, parents could have communicated their status at post-testing so there is the possibility that they learnt this information as they were being assessed.

C. Theory of Change

We hypothesize that the exposure to the aeioTU early education program, particularly in contrast to the existing supply of early childhood services and home learning environments in deprived communities, would have an impact on children's health and education

¹⁴ Ethics Committees at participating institutions approved the study's protocol in 2009.

¹⁵ For this reason, children assigned to the control group were further randomized into ordered waiting lists (by cohorts) so that they were offered program participation if necessary in such order if children assigned to the treatment group declined an aeioTU slot. This makes cross-overs from control to treatment that follow the randomized list, random. However, although centers reported following this list, we did not effectively monitor this process. In principle, this procedure would imply that children high up in the waiting list were also more likely to be assigned to treatment. In fact, the correlation between list order and enrollment in centers was -0.42.

outcomes.¹⁶ In particular, components of quality in the program such as comprehensive pre- and in-service teacher training, the use of a structured developmentally-oriented curriculum, high qualification requirements for staff, low child-to-teacher ratios, the presence of a team of experts including the pedagogical coordinator and the atelier, and strong infrastructure and supports should provide enhanced learning opportunities for children, Thus improving children's language, motor, cognitive and socio-emotional development.

The program offered the same amount of daily calorie intake as other child care services,¹⁷ provided nutritional supplementation during holidays, had on-site nutritional monitoring to detect nutritional risks, and adjust nutritional supplement, if necessary. As a result, we would also expect improvements in children's nutritional status.

Finally, the learning environment in the homes might have improved as a result of aeioTU's emphasis on active participation and education of families. The centers provided workshops in early childhood development and parenting on topics such as positive discipline. This component, continuous exposure to teachers, and the school's communication efforts may have had effects on parental knowledge and feasibly modified parenting behaviors. In addition, the marginal return of parental investments might increase with improvements in the quality of early care and education to the extent that complementarities exist between these sets of inputs in the early human capital production function. Therefore, we assessed whether change occurred in the home learning environment.

It is reasonable to assume though that parents might also switch resources and/or attention to other children (or themselves) as a response to the intervention, to equalize the allocation of resources across children (or household members). This may mitigate the program's impact.

D. Measures

The assessment instruments chosen have been used extensively in evaluations of early care and education including studies in developing countries (Fernald et al. 2017). They have adequate psychometrics and have detected program effects in other studies. Child developmental tests were collected by five graduates in psychology and three students in their senior year in psychology, who were trained to reliability standards (100% agreement with the trainer) by experienced research staff in a two-week training which included live reliability with infants and toddlers. Data collection for all children was conducted in spaces rented and adapted for that purpose, under identical conditions, with parental informed

¹⁶ At baseline, only 12.5% of infants and toddlers had used child care services during the previous year (14% in the treatment group and 11% in the control group). Of these, 90% attended a public child care program such as *Hogares Comunitarios*, 7% attended a private or NGO-sponsored child care program, and 3% were cared for by caregivers in their own home. The rest of the children (87.5%) were being cared at home by parents or unpaid relatives.

¹⁷ However, aeioTU added 20-25% of own resources to the nutritional component in order to be able to comply with the national nutritional guidelines. This implies that, presumably, other providers could not in practice fully comply with the calorie intake requirements given the government's stipend.

consent. Parental interviews were carried out alongside the child's assessment, in a separate room. Families and children were provided small incentives for participation and a snack.

Nutrition: As is standard practice in early intervention studies in developing countries (Fernald, Gertler, and Neufeld 2008), we measured height, weight, and arm circumference following World Health Organization (WHO) standards (WHO 2006; WHO 2007) at baseline and follow-up.

Cognitive Development: We used Cognitive, Motor, and Language scales from the Bayley Scales of Infant Development III (BSID), the most commonly used assessment of infant development (Bayley, 2005), for 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 done for another study on a similar population in Colombia which reports a test-retest reliability of this translation of 0.95–0.98 (Attanasio et al. 2014). The BSID is predictive of later measures of cognitive ability (Blaga et al. 2009; Feinstein 2003). We measured the BSID at baseline and follow-up (if still applicable). Infants and toddlers who outgrew the BSID at post-test were administered a commonly used measure of receptive vocabulary, the Peabody Picture Vocabulary test in Spanish (Test de Vocabulario en Imágenes Peabody, TVIP) (Padilla, Lugo, and Dunn 1986). An overall development score is drawn from the sum of the cognitive, motor, and language scales of the BSID. Raw scores are used in estimations as there are no norms for the Spanish translation and the English version is normed with a sample of children in the United States.

Socio-emotional Development: The Ages and Stages Questionnaire for the Socio-Emotional domain (ASQ:SE) is a parent-completed assessment system for children 6–60 months old. The ASQ:SE measures self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interactions with others (Squires, Bricker, and Twombly 2009a). The ASQ:SE has high levels of reliability and validity (Squires, Bricker, and Twombly 2009b). It was collected at baseline and follow-up through parent interviews. Higher scores represent higher levels of socio-emotional risk or negative behaviors. We used raw scores in all analyses.¹⁸

Home Environment: The Home Observation for Measurement of the Environment (HOME) measures the quality and extent of supports for child development in the home (Caldwell and Bradley 1984). The infant and toddler inventory includes six subscales: (1) responsiveness to parent, (2) avoidance of restriction and punishment, (3) organization of the environment, (4) appropriate play materials, (5) parental involvement, and (6) variety in daily stimulation. We used raw scores in all analyses. The HOME was administered at follow-up only, by six psychologists trained to reliability standards by experienced research staff in a 1.5 weeks training which included live reliability. The inter-rater reliability was

¹⁸ We do not use the risk of socio-emotional development calculated as the fraction of children above a threshold of behavioral problems defined by the test developers given that the ASQ:SE has not been locally validated and it would be inaccurate to use these thresholds (see Frongillo et al., 2014).

above 0.9 on the full scale. The instrument was collected during visits of 1-2 hours to the child's household.¹⁹

Socio-economic characteristics: In addition to the outcome measures described, we surveyed primarily the mother, or the head of the household in each home. We collected socio-economic information on families on schooling attainment, maternal age at birth of the child, race, income and expenditures, employment, assets, health insurance, number of children in the household, and childcare experiences.

E. Statistical Strategy

We hypothesize that the intervention would impact the six sets of outcomes assessed: nutritional, cognitive, language and motor development, socio-emotional development, and home environments. We estimate intention to treat (ITT) effects on the outcome measures using the following ordinary-least-squares (OLS) specification:

$$A_i^t = \lambda_1 + \lambda_2 ITT_i + \lambda_3 A_i^{baseline} + \lambda_4 X_i + \varepsilon_i \quad (1)$$

where ITT_i equals 1 if child i was randomly assigned to treatment and 0 otherwise. A_i^t is an outcome variable for child i in period t (in this case, follow-up), $A_i^{baseline}$ is the same outcome for child i or an available measure of the same developmental domain at baseline, and X_i is a vector of baseline controls including the child's race, age and age squared, maternal education and marital status, and the household wealth index, as well as those sociodemographic characteristics that were unbalanced at baseline such as the number of children younger than five years old in the household and whether the child had attended child care before randomization. It also includes neighborhood, cohort and gender (randomization strata) fixed effects, as well as tester or interviewer fixed effects.²⁰ We adjust two-tailed tests P-values for multiple hypotheses testing using the Romano and Wolf (2005) step-down procedure, where appropriate.²¹ We report both adjusted and unadjusted P-values.

We excluded from the analysis children with developmental outcomes with internally age-standardized values lower than three standard deviations below the mean of the relevant standardized distribution at baseline, since we consider this to be an indication of potential disability.²²

¹⁹ Assessors previously agreed upon appointments with families by phone, and were trained to be unobtrusive, to the extent possible, in the home while observing and interviewing primary caregivers.

²⁰ We also present the main results excluding the tester fixed effects.

²¹ Romano and Wolf (2005) step-down procedures for multiple testing were run within developmental domains (receptive and expressive language, fine and gross motor, all the socio-emotional outcomes and all the subscales of the HOME) extracting t-statistics of effect sizes from Stata and using the Matlab algorithm written by D Wunderli (University of Zurich).

²² The exclusion of outliers is shown in the "Analysis" panel the flow chart in Figure 1. The number of final observations varies by instrument and is between 459 for BSID III to a maximum of 748 for all other outcome variables.

As *ITT* was randomly assigned, we can be confident that the exclusion restriction in Equation 1 holds. Thus, $\hat{\lambda}_2$ estimated by OLS captures the causal impact of random assignment to treatment on the outcome. Variations of this model inquire into heterogeneous effects by gender, initial developmental levels and maternal education.²³

Given crossover between treatment and control groups as the program rolled out, we also estimated the impact of effective enrollment in centers (treatment-on-treated or TOT). In particular, 91 children (27%) assigned to the treatment group did not enroll in aeioTU centers and 80 children (16.5%) in the control group enrolled in the centers (see Figure 1). TOT estimates adjust the ITT (coefficients) by take-up rate. We estimated TOT effects using an instrumental variable approach. That is, we used random assignment (ITT) as an instrument for enrollment in the program. Enrollment is defined as a binary variable that equals 1 if child i was effectively enrolled in the aeioTU center (i.e., registered in center rosters), and 0 otherwise. Enrollment is directly obtained from aeioTU's administrative records.²⁴ This procedure provides an unbiased estimate of the program's effects among those who actually enrolled in the centers. The ITT indicator is a valid instrumental variable as it was randomly assigned, and it significantly explains actual enrollment. A regression of enrollment in centers on random assignment yields a statistically significant positive coefficient with an F-test of 50.5.

IV. Results

A. Baseline characteristics

Table 1 provides summary statistics at baseline by random assignment (ITT) for the subsample of children re-interviewed at follow-up (93%). We include various socio-demographic characteristics of children and their families, as well as baseline developmental outcomes. At baseline children were, on average, 20 months of age. Households had an average of 2.6 children under the age of 5, and 26% were headed by single parents. Mothers averaged 8.5 years of education, and only 36% had high school degrees.

Children were quite nutritionally vulnerable with height-for-age one standard deviation below average according to WHO standards (WHO 2006; WHO 2007). This translates into 21.6% of children stunted at baseline. In 2010, 15.2% children in rural areas were stunted (with comparable socioeconomic conditions as children in our sample) and

²³ We did not find statistically significant differences by child's age when comparing impacts on children 0-1 years of age vs. children 1-3 years of age (results available upon request). Another interesting analysis would be to look into differential program impacts by the child's HOME environment. However, we were only able to collect HOME at follow-up.

²⁴ Centers did not keep digital daily attendance records, so it was not possible to construct an indicator of average daily attendance. aeioTU indicates that by 2011 daily attendance rates (for all children and not exclusively infants and toddlers) varied between 60% and 78% in centers across the country. Attendance rates for other center-based care providers in Colombia or the region was not found publicly available by the authors, to compare rates accordingly.

12%, in urban areas, according to data from the Colombian Longitudinal Household Survey (ELCA, 2010).

Standardized BSID III scores²⁵ at baseline were 90.4 (SD=13.3) for cognition, 88.9 (SD=13.2) for language and 93.6 (SD=13.6) for motor development. This implies that children in the sample were about 0.7 standard deviations below average relative to the published norms (Feinstein 2003), and slightly below scores reported on a recent study including a similar sample of low-income children aged 12–24 months in rural areas in Colombia (Attanasio et al., 2018). The latter reported scores of 92.0 for cognition, and 91.6 for language. In contrast, children between 18 and 36 months of age in Bogota (Colombia’s capital) from average income households scored only within 0.1 below or above the norming sample in the three areas (Rubio-Codina et al., 2015). These results indicate that the children in our sample are significantly worse when compared to peers from better socio-economic households.

Average socio-emotional (ASQ:SE) scores were slightly elevated compared to the validation sample and quite comparable to children from low socio-economic urban households in the Colombian Longitudinal Household Survey in 2013 (ELCA, 2013).

Overall, there are few statistically and non-systematic significant differences between ITT groups at baseline, with treatment families having on average 0.18 more children under the age of five and being more likely to have attended childcare during the year prior to baseline (14% vs. 11%). We also report differences in the cohort of birth. In particular, there were fewer babies than toddlers in the treatment by design, as class sizes for younger children were lower. Finally, we report differences in developmental outcomes at baseline. Children in the treatment group scored higher in BSID receptive vocabulary and total language than children in the control group. At the same time, parents of children in the treatment group reported higher prevalence of problematic behaviors in the areas of compliance, autonomy and total ASQ:SE scores than parents of children in the control.²⁶ Some, but not all, of these differences remain significant after step-down adjustment of P-values. All analyses provided hereinafter control for these baseline measures to increase precision.²⁷

B. Attrition and Compliance

A total of 93%, that is, 763 of the 819 children in the sample was assessed at post-test (see Figure 1). In particular, 22 (6.5%) children in the treatment group and 34 (7%) in the control group were not re-interviewed due to migration to a municipality located too far from our

²⁵ Bayley III composites computed based on published norms provided by test developers. Standardized scores have mean 100 and standard deviation 15.

²⁶ When we re-create the same table for the complete sample interviewed at baseline, these imbalances remain but no others emerge.

²⁷ Tables A1a-i in the online Appendix display demographics and outcomes at baseline by gender and for various subgroups of children relevant for the heterogeneous impacts results presented later.

sites.²⁸ Attrition rates were not statistically significantly different between treatment and control groups, overall, and for subpopulation groups.²⁹

Compliance was high for the control and treatment groups: 73% of lottery winners enrolled in the centers and 83% of lottery losers did not enroll in the centers (see Figure 1).³⁰ Compliance was strongly correlated with ITT, the child's age, and varied significantly by site. In terms of enrollment in aeioTU centers at follow-up, 73% of children in the treatment group surveyed at follow-up were enrolled, as well as 17% of children in the control group. Enrollment was higher for older children, for boys, and for children of more educated working mothers.³¹

At post-testing, 82.2% of children in the treatment group were enrolled in early education services, of these, 91.5% were enrolled in aeioTU, 6.9% in other publicly provided child care service, and 10% in NGO-provided services or private center-based child care. A 16.8% of children in the treatment group did not attend any program, and an additional 1% did not respond. Of those children not attending any program, 92% were being cared by the mother, 2% by the father, and the remainder were cared for by other relatives or non-relatives at the child's home. As per the control group, 37% children were enrolled in early education programs at follow-up, out of which 48% were enrolled in aeioTU centers, 41.2% in other publicly provided alternatives, and the rest in NGO-provided services or private childcare. About 60% of children in the control group were not attending any child care program at follow-up; of these, 86% were being cared by the mother, 2% by the father and the rest, by other relatives or non-relatives.³²

C. Estimations of Program Impact

Table 2 reports the means for the treatment and control group at baseline for the sample of infants and toddlers assessed at follow-up, together with the coefficient of the effect of the intervention, the 95% confidence intervals for the treatment effect, standard two-tailed P-values, and step-down P-values.³³ We first present ITT estimates controlling for baseline outcomes (Equation 1). Next, we present TOT estimates by two-stage least squares using random assignment as an instrumental variable for enrollment in the centers. We then present these by subgroups of interest. Program effects are reported as fractions of a standard deviation relative to the control group. To put estimated effects in perspective, 0.75 SD is equivalent to the ability gap on the vocabulary between low and high-income 3-year-olds in Colombia (Colombian Longitudinal Household Survey ELCA, wave 2013).

The second panel reports effects on language, cognitive and motor development. Most of the results on cognition are shown for the subsample of children who did not outgrow the BSID III between baseline and follow-up. That is, the sample of approximately

²⁸ Table A2 shows attrition rates for the sample and for subgroups of interest.

²⁹ Tables A3a and A3b in the online Appendix assess the determinants of attrition for the complete sample and relevant subgroups.

³⁰ Table A4 in the online Appendix reports compliance rates by subpopulation groups.

³¹ Tables A5a-d in the online Appendix reports the determinants of enrollment in the program in the complete sample and subgroups of interest.

³² See Table A6 in the online appendix.

³³ In Tables A7-A12 in the online appendix, we show these results by subgroups.

480 children (out of 763 infants and toddlers surveyed at both baseline and follow-up) who were under 36 months of age at post-test.³⁴ The ITT effect on receptive language by BSID III was of 0.111 SD (P-value=0.035),³⁵ on expressive vocabulary was of 0.114 SD (P-value=0.035), with the combined language effect being 0.112 SD (P-value=0.006). The TOT effect on receptive language was of 0.204 SD (P-value=0.033), on expressive vocabulary was of 0.208 SD (P-value=0.033), with the combined language effect being of 0.205 SD (P-value=0.005).

To address the fact that BSID III is only observed at follow-up for a subsample of children, we internally age-standardized both, BSID III receptive vocabulary and TVIP scores in the complete sample and then pooled both scores in a single estimation, controlling for the change of instrument in the regression (last row in the second panel in Table 2). This allows us to estimate impacts on 710 children. We find no statistically significant impact on this combined measure of receptive vocabulary for the full sample. In addition, we estimate effects on receptive language for the subsample of children that outgrew the BSID III between baseline and follow-up using TVIP scores (row before the last in the second panel in Table 2). We do not observe significant effects on this subsample of 230 children older than 36 months of age at follow-up either. This might suggest that effects are stronger for younger children or alternatively, that TVIP is a noisy measure of language for children too close to the 36-month threshold. However, we show later that program impacts on TVIP scores for the subsample of older children are significant once we split the sample by gender, which would provide evidence that TVIP can effectively capture significant changes in receptive language at these ages.

The effect on BSID III cognition was 0.074 SD (P-value=0.035) by ITT and 0.138 SD (P-value=0.033) by TOT. The effect on fine motor development was 0.063 (P-value=0.048) by ITT and 0.119 (P-value =0.048) by TOT, and, the effect on gross motor development was 0.047 (P-value=0.076) by ITT and 0.085 (P-value=0.074) by TOT. The effect on total motor development (aggregate of fine and gross motor) was 0.049 (P-value=0.035) by ITT and 0.092 (P-value=0.033) by TOT. The program effects on overall development, as measured by the aggregate of language, cognition and motor development, were 0.064 SD (P-value=0.016) by ITT and 0.117 SD (P-value= 0.014) by TOT.

No evidence was found of statistically significant effects on nutritional outcomes, socio-emotional developmental or the home environment for the full sample.³⁶ We study these further in analysis of heterogeneous effects and explore the reasons behind the main results in the last section.³⁷

³⁴ For this reason, we show in corresponding tables on the online appendix, all analyses including attrition, compliance, enrollment, and baseline equivalence for this specific subsample of children as well.

³⁵ We report here step-down adjusted P-values although both, unadjusted and adjusted, are presented in Table 2.

³⁶ In Table A13 in the online appendix, we show results that exclude tester fixed effects from equation (1). The main results remain unchanged.

³⁷ Given that program effects for the BSID were estimated for the subsample of children who did not outgrow the measure between baseline and follow-up, we also estimated program impacts on nutrition, socio-emotional development and the home environment on this subsample, i.e., children

We replicated these analyses separately for girls and boys, children of mothers without secondary education completion and with it, and stunted and non-stunted children at baseline. Previous studies have suggested results may vary by gender, maternal education, and nutritional status (Blaga et al. 2009; Hoddinott et al. 2013). A summary of these subgroup estimations is shown in Table 3 (ITT estimations) and Table 4 (TOT estimations).³⁸ ITT and TOT estimates for subgroup analyses are virtually identical in terms of which effects are identified, and we cite only the TOT estimates in the text for brevity. As expected, TOT estimated effects are larger.

We still find no evidence of effects on nutrition for different subgroups. We find significant and large positive effects on language, motor and cognitive development for girls, but not boys. Specifically, effects for girls were of 0.33 SD (P-value=0.073) on expressive vocabulary by BSID III, 0.301 SD (P-value=0.014) for the language composite of BSID III, 0.48 SD (P-value=0.02) for the complete sample using age-standardized pooled TVIP and receptive language scores, and also 0.61 SD (P-value=0.051) for TVIP in the subsample of girls who outgrew the BSID III between baseline and follow-up. We also observed an effect of 0.228 SD (P-value=0.014) on fine motor development, 0.183 for overall development by BSID III.

In terms of maternal education, we observe most effects on language, motor and cognitive development for children of more educated mothers and not for children of less educated mothers. For example, the TOT effect on BSID receptive language was 0.379 SD (P-value =0.009) for children of mother with a high school degree or higher, 0.269 SD (P-value=0.090) for BSID expressive vocabulary and 0.320 SD (P-value=0.008) for BSID language total. Similarly, the effect on BSID cognition is 0.209 SD (P-value=0.041). The effect on receptive vocabulary for the full sample is 0.48 SD (P-value=0.057) for children of more educated mothers. There are no differences on motor development by maternal education.

Lastly, there are no differences in impacts on the home environment or socio-emotional development by child's age, gender, or initial developmental status, nor maternal education.

V. Discussion and Conclusions

Infants and toddlers were randomly assigned to treatment in two economically disadvantaged sites in northern Colombia to estimate the impact of a high-quality center-based care intervention. At baseline, only 12.5% of infants and toddlers had previously used child care services, mostly poor quality home-care. The rest of the children (87.5%) were cared at home by parents or unpaid relatives. Children came from very deprived backgrounds as indicated by their socioeconomic conditions, home learning environments, and initial developmental levels. As a result, the counterfactual to the high-quality center-

that were still younger than 36 months of age at follow-up. The results are virtually unchanged (see Table A14 in the online Appendix).

³⁸ Full results for each subgroup are reported in Tables A7-A12 in the online Appendix.

based care intervention studied was predominantly parental care or public home-based care with poor learning environments.

From only 130–150 days of intervention within the 8-month period between program roll-out and post-test, we find positive effects of 0.20 SD for BSID III total language and 0.14 SD for BSID III cognitive, 0.09 SD for BSID III motor development and 0.12 SD for BSID III overall development, for the subsample of children younger than 36 months of age at follow-up. These effects can be thought as percentages of the development gap with respect to more advantaged children this age as poorer children began about 0.75 SD behind. The particularly strong findings for language effects given the short duration of the program, suggest that the program has been most effective in exposing children to richer language environments than they would have experienced otherwise and in encouraging the use of expressive language. The importance of introducing infants and toddlers to a rich language environment has been well documented in the literature (Weisleder and Fernald 2013).

We report significant gender differences in impacts on language and cognitive development in favor of girls. Magnuson et al. (2016) report that the literature on the effects of early childhood interventions presents mixed evidence regarding which gender benefits the most. In particular, gender differences seem to vary by context, type and quality of the intervention, children's ages and the specific developmental domains under study. Garcia, Heckman and Ziff (2018), for example, report that girls benefitted more from the Abecedarian Program in the U.S. than boys did. According to the authors, this is due to the fact that girls came from more deprived households in which the learning environment was worse than that of boys. On the other hand, Muschkin et al. (2018) report that boys benefitted more than girls from the North Carolina's Early Care and Education Initiatives especially in math and reading skills. Both of these programs being relatively comparable to the aeioTU program under study.

In our case, we do not observe systematic baseline differences between boys and girls in socioeconomic characteristics. We do not observe differences in the learning environment by the HOME instrument at follow-up either. On the other hand, there may be program-specific characteristics or gender developmental differences that may explain the gender outcomes. For example, there is a predominant belief that boys lag behind girls in language development particular at very young ages and in fact recent studies reported a consistent advantage of girls especially during the first 30 months of life in early communicative gestures, early vocabulary growth, and vocabulary size and complexity (see Barbu (2015) for a review). We cannot discern whether specific features of the program contributed to these differences.

We also report that cognitive gains are largely associated with children of mothers with higher education. This result seems at odds with the notion that program impacts of early interventions are typically higher for more disadvantaged children. It is important to note that all households in the sample were disadvantaged to start with, and these estimations capture differences within low-income vulnerable households. Having said this, these findings support the theory of complementarities between early care and education investments and parental investments. In addition, it supports the theory that more educated mothers may have stronger behavioral responses to parenting components in interventions such as this one. HOME scores at follow-up are significantly correlated with

maternal education, but we do not observe significant program effects on the HOME environment for children of more educated mothers, therefore, complementariness would be occurring through areas unmeasured by the HOME.

As reported earlier, enrollment in aeioTU centers was higher for children of more educated mothers. By the same token, it is possible that children of more educated mothers were also less likely to be absent, and thus, effectively exposed to more program days. We do not have average daily attendance data to verify if this drives results.

We do not report statistically significant effects on socio-emotional development. The lack of effects might be due to several reasons. The ASQ:SE is parental-reported. A potential limitation of the use of parent-reported measures poses the usual concerns: lack of sensitivity, biases, and or the accurate measurement of underlying constructs. There are also documented differences in child socio-emotional reports across informants (Achenbach, McConaughy and Howell, 1987; Renk and Phares, 2004) which could be related to differences in how children behave in centers versus at home. Negative and neutral behavioral effects have been observed for children this young in other studies (Loeb et al., 2007; Deater-Deckard, Pinkerton and Scarr, 1996). Finally, it is possible that the aeioTU curriculum emphasizes on language and cognition more than on socio-emotional behavior and that changes in behaviors might require more specific targeting.

We do not find evidence of nutritional outcomes. We propose two feasible explanations for the lack of nutritional effects despite aeioTU investing additional resources in the nutritional supplement component of the program. First, children in the sample seem to be quite lagged in terms of height-for-age with stunting reaching almost 22% and an additional 30% of children being at risk of stunting given that their height-for-age z-scores are between -2 and -1 SD. On the other hand, the weight indicators appear to be above the population's mean. In particular, only 3% of children are underweight and less than 1% exhibit wasting. Height is a long-term indicator that depends not only on diet but also on health and sanitation conditions, and therefore it is more difficult to alter in the short term than weight. This means that a change in the children's diet for 8 months might not have been enough to affect this particular nutritional indicator. Second, parents might crowd out the nutritional component at home. As children received breakfast, lunch and snacks, parents might reallocate resources at home to other children in the household. However, we actually find a positive association between the program and parental-reported food insecurity³⁹ of about five percentage points that is statistically significant at 5% confidence level. So, crowding out by parents does not appear supported by the data. It is also important to consider that the evaluation started when the aeioTU program was a very young program with only two years of experience. These two centers started operating with these children right after construction. New programs might take time to achieve optimal procedures. For example, Andrew et al (2016) report that children younger than 3 years of age in center-based care eat only about 40% of lunch portions without an adult's help. Factors such as this one might require time to adjust.

³⁹ The specific question is whether the child skipped at least one meal during the previous week due to monetary constraints.

Our initial positive results on language and cognitive development, only eight months into the program, are quite important for those concerned with the development of disadvantaged infants and toddlers in developing countries. Alternative center-based early education programs in Colombia with several common features (Bernal et al., 2018), or innovations introduced to center-based care (Andrew et al. 2016) have demonstrated small or null effects. Similarly, other early years interventions have shown similar results as those reported in this study after longer periods of exposure reaching up to twice as long as the period we analyzed (Bernal and Fernández 2013; Attanasio et al. 2014). Attanasio et al. (2014) report effects of 0.22 SD on language for children 12–24 months of age at baseline of a home visitation program with weekly visits offered for 18 months. The program was designed and implemented directly by the research team and was carefully controlled and monitored; it did not make part of a governmental program or strategy.

Bernal et al. (2018) report negative and null effects of offering children in home-based childcare to transit to center-based childcare in urban Colombia after 18 months of exposure. Centers in that study broadly complied with comprehensive operational and technical national guidelines (as did aeioTU centers) that relate mostly to administrative guidelines and certain structural service parameters such as the number of children per square meter, characteristics of physical areas, teachers' qualifications, food handling, bookkeeping, etc. However, they differed significantly with respect to the aeioTU program in their pre- and in-service training strategies, the use of specific curricular guidelines for children's learning, professional developmental initiatives for teachers, and children's monitoring, among other features. This is important because both aeioTU and the public centers studied in Bernal et al. (2018) cost the same per child. The findings reported here along with the findings in Bernal et al. (2018) regarding quality issues during a similar initial phase of the center-based child care program studied, reveal that these quality features might be quite crucial to attain impacts on children. In the absence of critical components of child care quality, programs may have null or negative effects which would imply an inefficient use of limited resources.

Above all, our results highlight the importance of evaluating a wider range of potentially scalable early childhood programs in Colombia and other low-income countries. The findings suggest that policy makers should remain open to higher quality interventions in centers and not just to less expensive care or home-based interventions for infants and toddlers. Further studies of early childhood interventions more generally in the region are necessary to understand the differences in quality, dosage and delivery platforms. Future reports from this study hope to provide additional information on the effects of continued exposure to the program, effects at older ages, and several years later, and the persistence of effects as children progress in the aeioTu program and move on to primary schools. However, this first analysis is important as many interventions in developing regions do not surpass even the length of the pre to post-testing period of this study, and comparability requires understanding short-term impacts. This study also suggests that quality and curriculum deserve attention from both researchers and policy makers. The case of aeioTU also highlights the potential of private-public partnerships that can make it feasible to increase access and quality simultaneously on larger scales by combining public and private sector resources.

In interpreting the results of this study, it is important to keep in mind that at the time of the evaluation, the aeioTU program was still a very young program with only a couple of years of experience, while at the same time having a strong focus on growth in access. Since then, the program has put forth a continuous improvement cycle based on monitoring of classroom quality measures and detailed data on children's developmental trajectories (see Nores et al., 2018). With this improvement process in place, which is very much aligned with what is known in terms of quality programs (Frede 2005), we would certainly expect the program to be able to produce even higher impacts on children.

In terms of external validity of our results, it is important to mention that while the program has been scaled across the country, reaching a significant number of children urban, peri-urban and semi-rural communities, the extent to which we can generalize the results to these communities and other countries depends on how similar families may be. What is plausibly common across the rest of the country, and most of Latin America, is that the counter-factual to such an intervention for this age-group is no care, or in a few instances, home-based care or stimulation interventions, which would support generalizability of our results.

References

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological bulletin*, 101(2), 213.
- Andrew, A., Attanasio, O., Bernal, R., Cardona, L., Krutikova, S., Martínez, D., Medina, C., Peña, X., Rubio-Codina, M., & Vera-Hernández, M. (2016). Evaluation of centers of infant development: an early years intervention in Colombia. Unpublished manuscript, Institute for Fiscal Studies.
- Araujo, M.C., López Bóo, F., & Puyana, J.M. (2013). "Overview of Early Childhood Development Services in Latin America and the Caribbean." Washington, DC: Inter-American Development Bank.
- Attanasio, O., Baker-Henningham, H., Bernal, R., Meghir, C., Pineda, D., & Rubio-Codina, C. (2018). "Early Stimulation and Nutrition: The Impacts of a Scalable Intervention". CEDE Working Paper No. 42.
- Attanasio, O., Fernández, C., Fitzsimons, E., Grantham-McGregor, S., 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:g6126.
- Attanasio, O., Di Maro, V., & Vera-Hernández, M. (2013). "Community Nurseries and the Nutritional Status of Poor Children." Evidence from Colombia. *Econ J* 123(571):1025-58.
- Baker-Henningham, H., & López Bóo, F. (2010). "Early Childhood Stimulation Interventions in Developing Countries: A Comprehensive Literature Review." Washington, DC: Inter-American Development Bank.

- Baker-Henningham, H., & López Bóo, F. (2013). “Intervenciones de Estimulación Infantil Temprana en los Países en Vías de Desarrollo.” Washington, DC: Inter-American Development Bank.
- Barbu, S., Nardy, A., Chevrot, J., Guellai, B., Glas, L., Juhel, J., & Lemasson, A. (2015). “Sex Differences in language Across Early Childhood: Family Socioeconomic Status does not Impact Boys and girls Equally.” *Frontiers in Psychology*, Vol. 6.
- Barnett, W.S., & Boocock, S. (Eds.) (1998). *Early Care and Education for Children in Poverty: Promises, Programs and Long-Term Results*, pp.11–44. Albany, NY: SUNY Press.
- Barnett, W.S., and Masse, L. (2007). “Comparative Benefit–Cost Analysis of the Abecedarian Program and its Policy Implications.” *Econ of Ed Rev* 26(1):113-25.
- Barnett, W.S. (2008). “Preschool Education and its Lasting Effects: Research and Policy Implications.” Education Policy Research Unit. Boulder and Tempe: Arizona State University. <http://nepc.colorado.edu/publication/preschool-education>
- Bayley, N. (2005). “Bayley Scales of Infant Development.” San Antonio, TX: The Psychological Corporation, Harcourt Brace & Company.
- Behrman, J., & Urzúa, S. (2013). “Economic Perspectives on Some Important Dimensions of Early Childhood Development in Developing Countries.” In *Handbook of Early Childhood Development: Translating Research to Global Policy*, ed. Pia Rebello Britto, Patrice Engle and Charles Super, 123-41. New York: Oxford University Press.
- Behrman, J., Cheng, Y., & Todd, P. (2004). Evaluating preschool programs when length of exposure to the program varies: A nonparametric approach. *Review of Economics and Statistics*, 86(1), 108-132.
- Berlinski, S., & Schady, N. eds. (2016). *The Early Years: Child Well-Being and the Role of Public Policy*. New York: Springer.
- Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, 93(1-2), 219-234.
- Berlinski, S., & Galiani, S. (2007). The effect of a large expansion on pre-primary school facilities on preschool attendance and maternal employment. *Labour Economics*, 14:665-680.
- Bernal, R., & Camacho, A (2011). “La Importancia de los Programas de Primera Infancia en Colombia.” Bogotá, Colombia: Imprenta Nacional de Colombia.
- Bernal, R., & Camacho, A. (2014). Early Childhood Policy in the Context of Equity and Social Mobility in Colombia. In A. Montenegro and M. Meléndez, eds. *Equidad y Movilidad Social: Diagnósticos y Propuestas para la Transformación de la Sociedad Colombiana*. Editorial Uniandes.
- Bernal, R., & Fernández, C. (2013). “Subsidized Childcare and Child Development in Colombia: Effects of Hogares Comunitarios de Bienestar as a Function of Timing and Length of Exposure.” *Soc Sci Med* 97(November):241–49.
- Bernal, R., Martínez, M.A., & Quintero, C. (2015). *Situación de Niñas y Niños Colombianos Menores de Cinco Años entre 2010 y 2013*. Bogotá, Colombia: Editorial Kimpress.
- 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., Attanasio, O., Peña, X., & Vera-Hernández, M. (2018). "The Effects of the Transition from Home-Based Childcare to Center-Based Childcare in Colombia." *Forthcoming Early Childhood Research Quarterly*.
- Bernal, R., & Ramírez, C. (2018). "Improving child care quality at scale: the effects of From Zero to Forever". CEDE Working Paper No. 40.
- Black, M., Walker, S., Fernald, L., Andersen, C.T., DiGirolamo, A., Lu, ... Grantham-McGregor, S. (2017). Early child development coming of age: science through the life-course. *The Lancet*, 389, 77–90.
- Black, 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." *Lancet* 389(10064):77-90.
- Blaga, O., Shaddy, J., Anderson, C., Kannass, K., Little, T., & Colombo, J. (2009). "Structure and Continuity of Intellectual Development in Early Childhood." *Intell* 37(1):106-13.
- Bowman, B., Donovan, M., & Burns, M. (2001). *Eager to Learn: Educating our Preschoolers*. Washington, DC: National Academy Press.
- Britto, P.R., Lye, S., Proulx, K., Yousafzai, A., Matthews, S., Vaivada, T., ... & the Early Childhood Development Interventions Review Group (2017). Nurturing care: promoting early childhood development. *The Lancet*, 389, 91-102.
- Caldwell, B., and Bradley, R. (1984). "Home Observation for Measurement of the Environment: Administration Manual." Tempe, AZ: Family & Human Dynamics Research Institute, Arizona State University.
- Camilli, G., Vargas, V., Ryan, S., & Barnett, W.S. (2010). "Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development." *Teach Coll Rec* 112(3):579-620.
- Chang, S., Walker, S., Grantham-McGregor, S., & Powell, C. (2002) "Early Childhood Stunting and Later Behaviour and School Achievement." *J Child Psychol Psychiatry* 43(6):775-83.
- 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.
- Cunha, F., Heckman, J., Lochner, L., & Masterov, D. (2006). "Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education*, ed. Eric Hanushek and Finis Welch, 1:697-812. Elsevier.
- Deater-Deckard, K., Pinkerton, R., & Scarr, S. (1996). Child care quality and children's behavioral adjustment: A four-year longitudinal study. *Journal of Child Psychology and Psychiatry*, 37(8), 937-948.
- Encuesta Longitudinal Colombiana. (2013). "Encuesta Longitudinal Colombiana (ELCA)." Bogotá, Colombia: Universidad de los Andes. Available at <https://encuestalongitudinal.uniandes.edu.co/en/>.
- Engle, P., Fernald, L., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., Cabral de Mello, M., Hidrobo, M., Ulkuer, N., Ertem, I., & Iltus, S. (2011) "Strategies for

- Reducing Inequalities and Improving Developmental Outcomes for Young Children in Low-Income and Middle-Income Countries.” *Lancet* 37:1339-53.
- Engle, P., Black, M., Behrman, J., Cabral de Mello, M., Gertler, P., Kapiriri, L., Martorell, R., & Young, M. (2007). “Strategies to Avoid the Loss of Developmental Potential in More Than 200 Million Children in the Developing World.” *Lancet* 369(9557):229-42.
- Feinstein, L. (2003). “Inequality in the Early Cognitive Development of British Children in the 1970 Cohort.” *Economica* 70(277):73-97.
- Fernald, L., Prado, E., Kariger, P., & Raikes, A. (2009). “A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries.” Washington, DC: The World Bank.
- Fernald, L., Gertler, P., & Neufeld, L. (2008). “Role of Cash in Conditional Cash Transfer Programmes for Child Health, Growth, and Development: An Analysis of Mexico's Oportunidades.” *Lancet* 371(9615):828-37.
- Frede, E. (2005). "Assessment in a Continuous Improvement Cycle: New Jersey's Abbott Preschool Program." Invited paper for the National Early Childhood Accountability Task Force with support from The Pew Charitable Trusts, the Foundation for Child Development, and the Joyce Foundation. Available online at: <http://nieer.org>.
- Frongillo, E. A., Tofail, F., Hamadani, J. D., Warren, A. M., & Mehrin, S. F. (2014). Measures and indicators for assessing impact of interventions integrating nutrition, health, and early childhood development. *Annals of the New York academy of sciences*, 1308(1), 68-88.
- García, J., Heckman, J., and Ziff, A. (2018) “Gender Differences in the Benefits of an Influential Early Childhood Program”, Forthcoming *European Economic Review*.
- García, J., Heckman, J., Leaf, D., & Prados, M.J. (2016). “The Life-cycle Benefits of an Influential Early Childhood Program.” NBER Working Paper 22993.
- Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S., & Grantham-McGregor, S. (2014). “Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica.” *Science* 344(6187): 998–1001.
- Hoddinott, J., Alderman, H., Behrman, J., Haddad, L., & Horton, S. (2013). “The Economic Rationale for Investing in Stunting Reduction.” *Matern Child Nutr* 9(S2):69-82.
- Kagiticbasi, C., Bekman, S., & Goksel, A. (1995). “A Multipurpose Model of Nonformal Education: The Mother-Child Education Programme.” In *Coordinators' Notebook* 17. Canada: The Consultative Group on ECCD.
- Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education review*, 26(1), 52-66.
- Macours, K., Barham, T., & Maluccio, J. (2013). “More Schooling and More Learning? Effects of a 3-Year Conditional Cash Transfer Program in Nicaragua after 10 Years.” NEUDC Conference, Dartmouth, New Hampshire.
- Magnuson, K., Kelchen, R., Duncan, G., Schindler, H., Shager, H., & Yoshikawa, H. (2016). *Early Childhood Research Quarterly*, Vol 36 (Q3): 521-536.
- Muennig, P., Robertson, D., Johnson, G., Campbell, F., Pungello, E., & Neidell, M. (2011). “The Effect of an Early Education Program on Adult Health: the Carolina

- Abecedarian Project Randomized Controlled Trial.” *Am J Public Health* 101(3):512-16.
- Muschkin, C., Ladd, H., Dodge, K., & Bai, Y. (2018). “Gender Differences in the Impact of North Carolina’s Early Care and Education Initiatives on Student Outcomes in Elementary School” *Educational Policy*. <https://doi.org/10.1177/0895904818773901>
- Nores, M., Figueras-Daniel, A., López, M.A., & Bernal, R. (2018). “Implementing aeioTU: Quality improvement alongside an efficacy study. Learning while growing.” *Annals of the New York Academy of Sciences*, Vol. 1419, May.
- Nores, M., & Barnett, W.S. (2010). “Benefits of Early Childhood Interventions Across the World: (Under) Investing in the Very Young.” *Econ Educ Rev* 29(2):271-82.
- Padilla, E., Lugo, D., & Dunn, L. (1986). “Test de Vocabulario en Imágenes Peabody (TVIP).” Circle Pines, MN: American Guidance Service (AGS), Inc.
- Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review*, 24(2), 239-254.
- Romano, J., & Michael W. (2005). “Stepwise Multiple Testing as Formalized Data Snooping.” *Econometrica* 73(4):1237-82.
- Rubio-Codina M, Attanasio O, Meghir C, Varela N, & Grantham-McGregor S. (2015). The socio-economic gradient of child development: cross-sectional evidence from children 6-42 months in Bogota. *J Human Resources* Spring vol. 50 no. 2 464-483.
- Solt, F. (2011). “Standardizing the World Income Inequality Database.” *Soc Sci Q* 90(2):231-42.
- Squires, J., Bricker, D., and Twombly, E. (2009). “Technical Report on ASQ:SE.” Baltimore, Co: Paul H. Brookes Publishing.
- Squires, J., & Bricker, D. (2009). “Ages & Stages Questionnaires: A Parent-Completed Child Monitoring system.” (3rd Ed.). Baltimore, Co: Paul H. Brookes Publishing.
- United Nations. (2016). “Sustainable Development Goals. Goal 4: Ensure Inclusive and Quality Education for All and Promote Lifelong Learning.” Washington, DC: United Nations. <http://www.un.org/sustainabledevelopment/education/>
- Walker, S., Chang, S., Powell, C., & Grantham-McGregor, S. (2004). “Psychosocial Intervention Improves the Development of Term Low-Birth-Weight Infants.” *American Society for Nutritional Sciences* 134(6):1417-1423.
- Weisleder, A., & Fernald, A. (2013). “Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary.” *Psychol Sci* 24(11):2143-52.
- World Health Organization. (2006). “WHO Child Growth Standards: Length/height-for-age, Weight-for-age, Weight-for-length, Weight-for-height and Body Mass Index-for-age: Methods and Development.” WHO Multicentre Growth Reference Study Group. Geneva: World Health Organization.
- _____. (2007) “WHO Child Growth Standards: Head Circumference-for-age, Arm Circumference-for-age, Triceps Skinfold-for-age and Subscapular Skinfold-for-age: Methods and Development.” WHO Multicentre Growth Reference Study Group. Geneva: World Health Organization.
- World Bank. (2013) “Colombia. Early Childhood Development.” SABER Country Report, Washington, DC. <http://hdl.handle.net/10986/16280>.

- World Economic Forum. (2017) “Realizing the Human Potential in the Fourth Industrial Revolution. An Agenda for Leaders to Shape the Future of Education, Gender and Work.” White paper. http://www3.weforum.org/docs/WEF_EGW_Whitepaper.pdf.
- Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *The Future of Children*, 26(2), 21-36.
- Yoshikawa, H., Leyva, D., Snow, C., Treviño, E., Barata, M., Weiland, ... Arbour, M. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Dev Psychol*, 51(3), 309–322.

Figure 1. Study's flow chart for sample selection

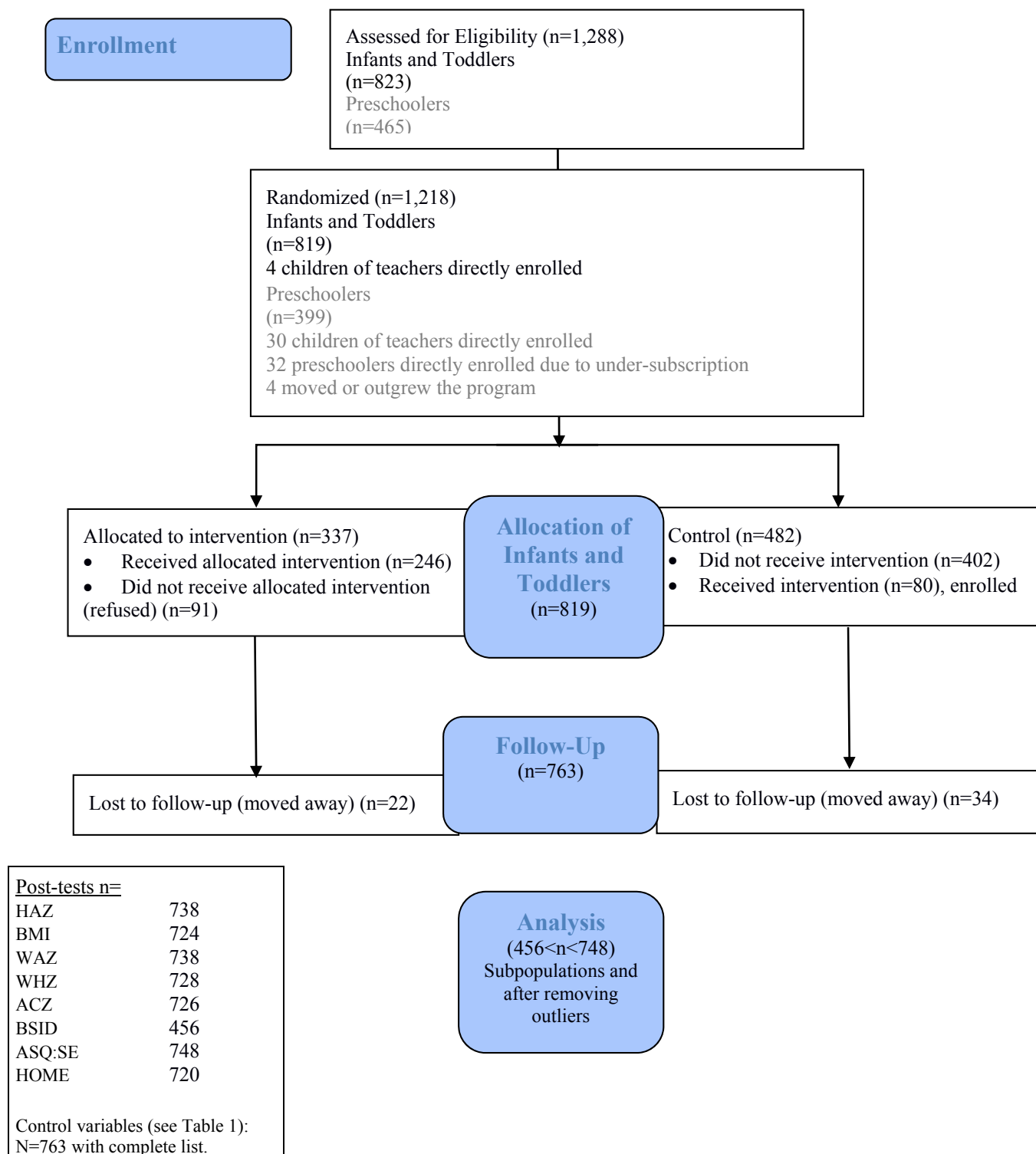


Table 1. Comparison of baseline variables before randomization in 2010, by intent-to-treat for the sample of 789 children followed up in 2011

Socio-Demographics and Outcomes at Baseline	N	Control		Treatment		P-value*	Stepdown P-value†
		Mean	(SD)	Mean	(SD)		
Child's age in months	763	19.79	(8.83)	20.73	(9.69)	0.164	0.868
Child's gender (male)	763	0.53	(0.50)	0.53	(0.50)	0.927	0.997
Child's race (black)	763	0.59	(0.49)	0.63	(0.48)	0.290	0.931
Maternal marital status (single)	763	0.29	(0.45)	0.28	(0.45)	0.723	0.997
Health insurance for child	763	0.78	(0.42)	0.79	(0.41)	0.629	0.996
Mother secondary incomplete	763	0.63	(0.48)	0.64	(0.48)	0.787	0.997
Mother secondary complete and above	763	0.37	(0.48)	0.36	(0.48)	0.787	0.997
Wealth Index‡	763	0.28	(4.90)	-0.10	(3.96)	0.256	0.923
Children books at home	763	1.33	(2.24)	1.40	(3.15)	0.739	0.997
Mother education years	763	8.38	(3.33)	8.37	(3.12)	0.937	0.997
No. of children ≤5 yrs	763	2.60	(0.76)	2.78	(0.86)	0.004	0.051
Childcare by baseline	763	0.11	(0.31)	0.14	(0.34)	0.218	0.914
Neighborhood (La Paz)	763	0.60	(0.49)	0.49	(0.50)	0.003	0.041
Neighborhood (Alpes B)	763	0.04	(0.21)	0.07	(0.26)	0.133	0.818
Neighborhood (Timayui 1)	763	0.18	(0.38)	0.22	(0.41)	0.173	0.868
Neighborhood (Timayui 2)	763	0.18	(0.39)	0.23	(0.42)	0.123	0.813
Cohort 2008	763	0.39	(0.49)	0.37	(0.48)	0.680	0.997
Cohort 2009	763	0.42	(0.49)	0.33	(0.47)	0.008	0.099
Cohort 2010	763	0.12	(0.33)	0.17	(0.37)	0.098	0.749
<i>Nutrition (Z scores)</i>							
Length/Height-for-age	743	-1.06	(1.18)	-1.10	(1.04)	0.625	0.861
BMI-for-age	731	0.57	(0.98)	0.48	(1.00)	0.223	0.511
Weight-for-age	743	-0.22	(1.07)	-0.34	(1.01)	0.129	0.347
Weight-for-length	735	0.44	(0.97)	0.35	(1.00)	0.244	0.537
Arm circumference	734	0.26	(0.85)	0.24	(0.82)	0.733	0.861
<i>Infant development: BSID III Raw scores</i>							
Receptive vocabulary	737	18.79	(7.62)	20.10	(8.03)	0.025	0.057
Expressive vocabulary	739	19.24	(8.86)	20.23	(9.90)	0.157	0.266
Total language	729	38.02	(16.12)	40.50	(17.53)	0.049	0.106
Cognitive	743	48.02	(14.41)	49.36	(15.18)	0.221	0.290
Fine motor	739	32.18	(9.22)	33.17	(9.78)	0.162	0.266
Gross motor	742	46.80	(12.71)	47.49	(13.70)	0.483	0.488
Total motor	735	78.95	(21.54)	80.66	(23.09)	0.303	0.341
BSID III Total	715	164.95	(50.55)	171.27	(53.77)	0.109	0.195
<i>Socio-emotional Development (ASQ: SE)</i>							
Self-regulation	754	16.43	(14.82)	18.32	(16.16)	0.096	0.420
Compliance	564	2.07	(4.06)	2.77	(4.65)	0.058	0.323
Communication	754	1.54	(3.98)	1.90	(4.01)	0.217	0.626
Adaptive functioning	753	6.85	(8.65)	6.92	(8.47)	0.913	0.999
Autonomy	564	4.06	(5.27)	5.15	(5.78)	0.020	0.131
Affect	754	3.82	(4.76)	3.88	(4.92)	0.865	0.999
Interaction	754	5.46	(7.07)	5.42	(6.86)	0.945	0.999
ASQ:SE Total	753	39.79	(26.42)	43.50	(28.23)	0.065	0.323

BSID denotes the raw score from the Bayley Scales of Infant Development 3rd edition (Bayley, 2005), ASQ:SE denotes the raw score from the Ages and Stages Socio-emotional Questionnaire (Squires, Bricker & Tomblly, 2009b). P-values for differences in means ≤ 0.10 between treatment and control children are in bold type.

*Standard P-values.

†Stepdown P-values are for Romano and Wolf (2005) stepdown procedures applied by blocks of baseline variables. 2,000 repetitions.

‡Wealth index calculated through principal component of a set of variables including type and characteristics of dwelling (floors, walls, bathrooms, etc.), availability of public utilities and durable goods.

Table 2. aeioTu intervention. ITT and TOT estimations of program effects by outcome

Variable	N	Post-intervention Means			ITT Estimated Effects				TOT Estimated Effects			
		Control mean*	Treatment mean*	Difference in means	ITT β (95% CI)	D \dagger	P value \ddagger	Stepdown P value \S	TOT β (95% CI)	D \dagger	P value \ddagger	Stepdown P value \S
Nutrition (Z scores)												
Length/Height-for-age	738	-1.15	-1.12	0.03	0.055 (-0.034 to 0.144)	0.047	0.224	0.576	0.097 (-0.058 to 0.252)	0.083	0.219	0.562
BMI-for-age	724	0.35	0.26	-0.09	-0.037 (-0.134 to 0.060)	-0.038	0.456	0.814	-0.066 (-0.236 to 0.104)	-0.067	0.448	0.804
Weight-for-age	738	-0.46	-0.51	-0.05	0.026 (-0.060 to 0.113)	0.025	0.548	0.862	0.046 (-0.103 to 0.196)	0.043	0.543	0.857
Weight-for-length	728	0.19	0.15	-0.04	0.002 (-0.090 to 0.094)	0.002	0.967	0.965	0.003 (-0.158 to 0.165)	0.003	0.967	0.964
Arm circumference	726	-0.37	-0.4	-0.03	-0.026 (-0.117 to 0.065)	-0.031	0.575	0.862	-0.044 (-0.198 to 0.109)	-0.053	0.57	0.857
Language, Cognitive and Motor Development: Bayley Scales of Infant Development III (BSID III)												
BSID receptive vocabulary	480	23.85	24.55	0.70	0.845 (0.182 to 1.507)	0.111	0.013	0.035	1.554 (0.351 to 2.758)	0.204	0.011	0.033
BSID expressive vocabulary	482	24.30	24.88	0.58	1.001 (0.191 to 1.812)	0.114	0.016	0.035	1.834 (0.384 to 3.283)	0.208	0.013	0.033
BSID language total	473	48.14	49.39	1.25	1.807 (0.516 to 3.098)	0.112	0.006	n/a	3.289 (0.986 to 5.593)	0.205	0.005	n/a
BSID cognitive	487	57.77	58.54	0.77	1.082 (0.225 to 1.939)	0.074	0.013	0.035	2.017 (0.453 to 3.581)	0.138	0.012	0.033
BSID fine motor	483	38.37	38.86	0.49	0.589 (0.075 to 1.103)	0.063	0.025	0.048	1.109 (0.142 to 2.076)	0.119	0.025	0.048
BSID gross motor	482	54.47	54.98	0.51	0.602 (-0.077 to 1.281)	0.047	0.082	0.076	1.105 (-0.119 to 2.329)	0.085	0.077	0.074
BSID motor total	478	92.87	93.75	0.88	1.083 (0.075 to 2.091)	0.049	0.035	n/a	2.008 (0.158 to 3.858)	0.092	0.033	n/a
BSID III Total	456	198.65	201.36	2.71	3.262 (0.603 to 5.921)	0.064	0.016	n/a	5.962 (1.204 to 10.720)	0.117	0.014	n/a
TVIP	230	11.94	12.08	0.15	0.196 (-1.563 to 1.955)	0.027	0.826	n/a	0.307 (-2.327 to 2.940)	0.042	0.819	n/a
Receptive vocabulary (Std) #	710	-0.03	0.07	0.10	0.100 (-0.047 to 0.246)	0.1000	0.184	n/a	0.172 (-0.078 to 0.422)	0.173	0.177	n/a

Socio-Emotional Development: Ages and Stages Socio-emotional (ASQ:SE)												
Self-regulation	748	18.91	21.18	2.27	1.083 (-1.065 to 3.232)	0.073	0.323	0.877	1.945 (-1.846 to 5.735)	0.131	0.314	0.864
Compliance	559	2.89	3.75	0.86	0.424 (-0.331 to 1.179)	0.103	0.270	0.868	0.703 (-0.518 to 1.924)	0.171	0.258	0.851
Communication	748	2.09	2.31	0.21	0.038 (-0.694 to 0.769)	0.010	0.919	0.991	0.068 (-1.219 to 1.355)	0.017	0.918	0.990
Adaptive functioning	747	5.68	6.28	0.59	0.258 (-0.911 to 1.428)	0.03	0.665	0.978	0.463 (-1.593 to 2.519)	0.054	0.659	0.976
Autonomy	559	8.12	8.73	0.61	0.050 (-0.714 to 0.813)	0.009	0.899	0.991	0.082 (-1.151 to 1.315)	0.016	0.896	0.990
Affect	748	2.38	2.72	0.34	0.188 (-0.534 to 0.911)	0.04	0.609	0.978	0.337 (-0.934 to 1.607)	0.071	0.603	0.976
Interaction	747	6.98	6.67	-0.31	-0.547 (-1.822 to 0.727)	-0.077	0.400	0.907	-0.979 (-3.219 to 1.262)	-0.138	0.392	0.897
ASQ:SE Total	746	45.33	50.26	4.94	2.037 (-2.447 to 6.522)	0.078	0.373	n/a	3.666 (-4.264 to 11.596)	0.14	0.364	n/a
Home Observation and Measurement of the Environment (HOME)												
Responsivity	720	7.76	7.9	0.15	0.133 (-0.216 to 0.482)	0.057	0.455	0.831	0.232 (-0.367 to 0.831)	0.099	0.447	0.811
Acceptance	720	6.39	6.41	0.02	0.073 (-0.068 to 0.214)	0.079	0.31	0.826	0.128 (-0.115 to 0.370)	0.137	0.303	0.806
Organization	702	4.56	4.63	0.07	0.078 (-0.076 to 0.231)	0.073	0.319	0.826	0.135 (-0.127 to 0.397)	0.126	0.311	0.806
Learning Materials	702	3.61	3.53	-0.08	-0.001 (-0.261 to 0.259)	-0.001	0.993	0.993	-0.002 (-0.446 to 0.442)	-0.001	0.993	0.994
Involvement	702	3.22	3.05	-0.17	-0.059 (-0.282 to 0.163)	-0.04	0.6	0.831	-0.103 (-0.481 to 0.274)	-0.07	0.592	0.817
Variety	720	2.20	2.12	-0.08	-0.106 (-0.269 to 0.057)	-0.093	0.203	0.717	-0.185 (-0.465 to 0.095)	-0.162	0.195	0.696
HOME Total	720	27.90	27.78	-0.11	0.011 (-0.720 to 0.742)	0.002	0.977	n/a	0.019 (-1.234 to 1.272)	0.003	0.976	n/a

Individual lines present the results of separate regressions. Sample sizes vary by instrument due to the collection method (with less households assessed with HOME than children nutritionally) and as children at post testing grow out of the BSID III. In particular, we show BSID III results for children still eligible for BSID at follow-up (456<N<487), as well as TVIP language results for children who outgrew the BSID III at follow-up (N=230). For each estimation we present: the means of the outcome by group (treatment and control) and the difference in means, and the results of the estimations for intent-to-treat and treatment-on-treated. β s are estimated controlling for the corresponding pre-test (except for HOME for which there was no pretest), age, age squared, male, black, mother marital status (single), maternal years of education, household wealth index, household kids, childcare experience before randomization, neighborhood, cohort of birth and Tester FE. P-values ≤ 0.10 are in bold type.

*Means are age adjusted using ANOVAs.

†Effects sizes, also known as Cohen's D, are β s interpreted as fraction of SD in control group at baseline (except for HOME for which there was no baseline testing).

‡Standard P-values.

§Stepdown P-values are for Romano and Wolf (2005) Stepdown procedures applied to blocks of outcomes per type of developmental dimension measured. For Bayley, motor scales are one block and

cognitive and language are another block. Combined receptive vocabulary excluded from the block as it combines two measures of receptive vocabulary and is based on a different sample than other outcomes within that group. Not calculated for the Total's aggregate scores as these are aggregate measures across various dimensions or for one developmental domain.
We internally age-standardized both, BSID III receptive vocabulary and TVIP scores in the complete sample and then pooled both in a single regression controlling for measure type.

Table 3. aeioTu intervention. Intent-to-treat estimations of program effects by outcome and for selected groups

Outcome variables	Females		Males		Less than High school		High School or higher		Stunted		Non-Stunted	
	β (95% CI)	D* (P-value†)	β (95% CI)	D* (P-value†)	β (95% CI)	D* (P-value†)	β (95% CI)	D* (P-value†)	β (95% CI)	D* (P-value†)	β (95% CI)	D* (P-value†)
Nutrition (Z scores)												
Height-for-age	0.013 (-0.101 to 0.127)	0.011 (0.999)	0.090 (-0.046 to 0.226)	0.079 (0.770)	0.064 (-0.053 to 0.181)	0.055 (0.855)	0.028 (-0.111 to 0.167)	0.024 (0.976)	-0.090 (-0.318 to 0.139)	-0.172 (0.926)	0.088 (-0.008 to 0.184)	0.094 (0.409)
BMI-for-age	-0.031 (-0.168 to 0.105)	-0.035 (0.993)	-0.038 (-0.180 to 0.104)	-0.036 (0.993)	-0.094 (-0.215 to 0.027)	-0.095 (0.597)	0.086 (-0.084 to 0.256)	0.087 (0.861)	-0.091 (-0.348 to 0.165)	-0.109 (0.930)	-0.016 (-0.121 to 0.089)	-0.016 (0.965)
Weight-for-age	0.007 (-0.100 to 0.114)	0.007 (0.999)	0.044 (-0.090 to 0.177)	0.038 (0.989)	0.002 (-0.112 to 0.116)	0.002 (0.994)	0.083 (-0.052 to 0.218)	0.074 (0.801)	-0.067 (-0.296 to 0.162)	-0.082 (0.949)	0.056 (-0.037 to 0.148)	0.057 (0.781)
Weight-for-length	-0.002 (-0.135 to 0.132)	-0.002 (0.999)	0.012 (-0.118 to 0.142)	0.011 (0.999)	-0.041 (-0.155 to 0.074)	-0.041 (0.946)	0.107 (-0.053 to 0.268)	0.107 (0.744)	0.049 (-0.190 to 0.288)	0.057 (0.965)	-0.008 (-0.108 to 0.091)	-0.008 (0.979)
Arm circumference	-0.037 (-0.171 to 0.097)	-0.047 (0.993)	-0.009 (-0.138 to 0.119)	-0.010 (0.999)	-0.028 (-0.142 to 0.085)	-0.035 (0.976)	-0.008 (-0.162 to 0.146)	-0.009 (0.994)	-0.130 (-0.369 to 0.109)	-0.171 (0.827)	0.004 (-0.094 to 0.103)	0.006 (0.979)
Cognitive and Language Development: Bayley Scales of Infant Development III (BSID III)												
BSID receptive vocabulary	1.129 (0.099 to 2.159)	0.149 (0.130)	0.495 (-0.390 to 1.381)	0.065 (0.440)	0.276 (-0.565 to 1.117)	0.036 (0.524)	1.808 (0.666 to 2.950)	0.239 (0.010)	0.690 (-1.337 to 2.716)	0.093 (0.698)	0.729 (0.013 to 1.445)	0.095 (0.185)
BSID expressive vocabulary	1.566 (0.302 to 2.830)	0.179 (0.078)	0.217 (-0.856 to 1.290)	0.025 (0.680)	0.608 (-0.424 to 1.639)	0.069 (0.418)	1.561 (0.143 to 2.979)	0.176 (0.110)	-0.569 (-2.725 to 1.587)	-0.064 (0.698)	1.022 (0.130 to 1.914)	0.116 (0.127)
BSID language total	2.723 (0.703 to 4.744)	0.171 (0.013)	0.512 (-1.179 to 2.203)	0.032 (0.545)	0.895 (-0.732 to 2.523)	0.056 (0.272)	3.311 (1.011 to 5.610)	0.205 (0.010)	-0.257 (-4.063 to 3.549)	-0.016 (0.896)	1.785 (0.387 to 3.184)	0.110 (0.027)
BSID cognitive	1.025 (-0.250 to 2.299)	0.075 (0.292)	1.098 (-0.085 to 2.281)	0.072 (0.225)	0.732 (-0.368 to 1.832)	0.049 (0.418)	1.888 (0.421 to 3.356)	0.134 (0.056)	1.899 (-0.937 to 4.736)	0.129 (0.416)	0.672 (-0.226 to 1.570)	0.046 (0.416)
BSID fine motor	1.119 (0.375 to 1.863)	0.125 (0.012)	0.150 (-0.573 to 0.873)	0.016 (0.718)	0.436 (-0.243 to 1.115)	0.045 (0.469)	0.779 (-0.024 to 1.582)	0.087 (0.190)	-0.203 (-1.786 to 1.380)	-0.023 (0.796)	0.522 (-0.019 to 1.064)	0.055 (0.200)
BSID gross motor	0.950 (-0.111 to 2.011)	0.080 (0.213)	0.321 (-0.600 to 1.242)	0.023 (0.718)	0.577 (-0.301 to 1.456)	0.043 (0.469)	0.615 (-0.544 to 1.773)	0.050 (0.469)	0.966 (-1.077 to 3.009)	0.077 (0.675)	0.370 (-0.366 to 1.106)	0.028 (0.675)
BSID motor total	2.059 (0.543 to 3.575)	0.100 (0.014)	0.298 (-1.090 to 1.686)	0.013 (0.663)	0.873 (-0.456 to 2.201)	0.039 (0.199)	1.406 (-0.189 to 3.002)	0.067 (0.154)	-0.419 (-3.526 to 2.688)	-0.020 (0.786)	0.886 (-0.172 to 1.944)	0.040 (0.183)
BSID III Total	4.906 (0.984 to 8.829)	0.101 (0.025)	1.532 (-2.049 to 5.113)	0.029 (0.388)	1.681 (-1.723 to 5.085)	0.032 (0.325)	6.416 (1.945 to 10.886)	0.129 (0.008)	0.146 (-8.997 to 9.289)	0.003 (0.977)	2.812 (0.066 to 5.558)	0.055 (0.082)
TVIP	2.554 (0.248 to 4.859)	0.445 (0.060)	-2.379 (-5.194 to 0.437)	-0.288 (0.091)	-0.032 (-2.288 to 2.223)	-0.004 (0.977)	0.297 (-2.839 to 3.434)	0.042 (0.977)	0.964 (-3.228 to 5.156)	0.140 (0.852)	-0.145 (-2.254 to 1.963)	-0.020 (0.876)
Receptive vocabulary‡	0.273 (0.058 to 0.488)	0.295 (0.026)	-0.086 (-0.293 to 0.120)	-0.083 (0.418)	-0.005 (-0.187 to 0.177)	-0.005 (0.959)	0.279 (0.024 to 0.535)	0.279 (0.064)	0.200 (-0.120 to 0.521)	0.203 (0.386)	0.054 (-0.113 to 0.220)	0.054 (0.527)
Socio-Emotional Development: Ages and Stages Socio-emotional (ASQ:SE)												
Self-regulation	2.700 (-0.541 to 5.942)	0.196 (0.731)	-0.244 (-3.202 to 2.714)	-0.016 (0.999)	1.580 (-1.198 to 4.358)	0.106 (0.971)	-0.211 (-3.798 to 3.375)	-0.014 (0.999)	0.434 (-4.919 to 5.788)	0.028 (0.999)	1.598 (-0.774 to 3.971)	0.109 (0.914)
Compliance	0.401 (-0.729 to 1.531)	0.105 (0.997)	0.136 (-0.908 to 1.180)	0.031 (0.999)	0.492 (-0.487 to 1.471)	0.123 (0.983)	0.474 (-0.785 to 1.733)	0.111 (0.999)	0.851 (-0.835 to 2.537)	0.192 (0.986)	-0.014 (-0.886 to 0.858)	-0.004 (0.999)
Communication	-0.249 (-1.474 to 0.977)	-0.071 (0.999)	0.141 (-0.736 to 1.018)	0.033 (0.999)	0.248 (-0.620 to 1.116)	0.060 (0.999)	-0.644 (-2.024 to 0.736)	-0.183 (0.988)	0.753 (-0.928 to 2.435)	0.203 (0.990)	-0.096 (-0.922 to 0.729)	-0.024 (0.996)

Adaptive functioning	1.471 (-0.268 to 3.210)	0.167 (0.731)	-0.806 (-2.399 to 0.788)	-0.097 (0.980)	0.149 (-1.213 to 1.510)	0.017 (0.999)	0.442 (-1.826 to 2.711)	0.053 (0.999)	0.308 (-2.441 to 3.057)	0.035 (0.999)	0.217 (-1.096 to 1.530)	0.026 (0.996)
Autonomy	-0.073 (-1.110 to 0.963)	-0.014 (0.999)	0.299 (-0.869 to 1.466)	0.056 (0.999)	0.247 (-0.716 to 1.210)	0.049 (0.999)	-0.300 (-1.649 to 1.049)	-0.054 (0.999)	0.677 (-1.103 to 2.458)	0.138 (0.996)	-0.114 (-0.964 to 0.737)	-0.021 (0.996)
Affect	0.142 (-1.019 to 1.304)	0.031 (0.999)	-0.102 (-1.013 to 0.810)	-0.021 (0.999)	0.167 (-0.766 to 1.100)	0.035 (0.999)	0.334 (-0.871 to 1.539)	0.070 (0.999)	-0.007 (-1.798 to 1.785)	-0.001 (0.999)	0.176 (-0.617 to 0.970)	0.037 (0.996)
Interaction	0.026 (-1.896 to 1.949)	0.004 (0.999)	-1.288 (-3.050 to 0.475)	-0.179 (0.831)	-0.099 (-1.727 to 1.529)	-0.013 (0.999)	-1.849 (-4.019 to 0.320)	-0.285 (0.705)	-0.590 (-3.877 to 2.696)	-0.083 (0.996)	-0.672 (-2.058 to 0.714)	-0.095 (0.990)
ASQ:SE Total	5.839 (-1.243 to 12.922)	0.245 (0.192)	-1.883 (-7.666 to 3.901)	-0.067 (0.513)	3.316 (-2.201 to 8.832)	0.125 (0.425)	-0.875 (-8.998 to 7.247)	-0.034 (0.835)	4.018 (-7.182 to 15.218)	0.152 (0.660)	2.025 (-2.918 to 6.969)	0.077 (0.660)
Home Observation and Measurement of the Environment (HOME)												
Responsivity	0.104 (-0.421 to 0.628)	0.046 (0.996)	0.193 (-0.288 to 0.674)	0.079 (0.996)	-0.057 (-0.501 to 0.387)	-0.024 (0.990)	0.261 (-0.317 to 0.839)	0.116 (0.990)	-0.334 (-1.152 to 0.485)	-0.149 (0.966)	0.258 (-0.131 to 0.647)	0.108 (0.833)
Acceptance	0.084 (-0.126 to 0.293)	0.090 (0.996)	0.080 (-0.117 to 0.278)	0.086 (0.996)	0.069 (-0.109 to 0.246)	0.074 (0.990)	0.092 (-0.155 to 0.339)	0.098 (0.990)	-0.218 (-0.525 to 0.089)	-0.243 (0.805)	0.116 (-0.044 to 0.276)	0.124 (0.805)
Organization	0.006 (-0.224 to 0.235)	0.005 (0.996)	0.159 (-0.058 to 0.376)	0.148 (0.814)	0.092 (-0.113 to 0.298)	0.086 (0.990)	-0.013 (-0.246 to 0.220)	-0.013 (0.990)	0.132 (-0.234 to 0.499)	0.123 (0.966)	0.081 (-0.091 to 0.253)	0.077 (0.966)
Learning Materials	-0.146 (-0.551 to 0.258)	-0.072 (0.996)	0.109 (-0.246 to 0.463)	0.058 (0.996)	0.185 (-0.133 to 0.502)	0.106 (0.948)	-0.338 (-0.807 to 0.131)	-0.181 (0.853)	0.085 (-0.528 to 0.697)	0.044 (0.968)	-0.021 (-0.317 to 0.274)	-0.011 (0.968)
Involvement	-0.090 (-0.436 to 0.255)	-0.060 (0.996)	-0.000 (-0.304 to 0.304)	-0.000 (0.999)	-0.099 (-0.370 to 0.172)	-0.071 (0.990)	-0.111 (-0.503 to 0.282)	-0.071 (0.990)	-0.135 (-0.632 to 0.363)	-0.091 (0.968)	-0.054 (-0.308 to 0.201)	-0.036 (0.968)
Variety	-0.004 (-0.251 to 0.244)	-0.003 (0.999)	-0.229 (-0.452 to -0.006)	-0.201 (0.410)	-0.083 (-0.290 to 0.124)	-0.073 (0.990)	-0.129 (-0.405 to 0.147)	-0.113 (0.990)	-0.282 (-0.608 to 0.044)	-0.284 (0.621)	-0.084 (-0.274 to 0.105)	-0.071 (0.966)
HOME Total	-0.075 (-1.166 to 1.016)	-0.015 (0.966)	0.122 (-0.917 to 1.162)	0.021 (0.966)	-0.056 (-0.976 to 0.864)	-0.011 (0.944)	-0.196 (-1.459 to 1.068)	-0.037 (0.944)	-0.716 (-2.422 to 0.989)	-0.140 (0.629)	0.171 (-0.655 to 0.997)	0.031 (0.680)

Individual lines present the results of separate regressions for each subpopulation group. Sample sizes vary by instrument due to the collection method (with less households assessed with HOME than children nutritionally) and as children at post testing grow out of the BSID III. For each estimation we present: the means of the outcome by group (treatment and control) and the difference in means, and the results of the estimations for intent-to-treat and treatment-on-treated. β s are estimated controlling for the corresponding pre-test (except for HOME for which there was no pretest), age, age squared, male, black, mother marital status (single), maternal years of education, household wealth index, household kids, childcare experience before randomization, neighborhood, cohort of birth and Tester FE. P-values ≤ 0.10 are in bold type.

*Effects sizes, also known as Cohen's D, are β s interpreted as fraction of SD in control group at baseline (except for HOME for which there was no baseline testing).

†Stepdown P-values are for Romano and Wolf (2005) Stepdown procedures applied to blocks of outcomes per type of developmental dimension measured. Combined receptive vocabulary excluded from the block as it combines two measures of receptive vocabulary and is based on a different sample than other outcomes within that group. In these set of estimations, step-downs are also calculated within subgroup categories in pairs of columns, that is: females and males, less than high school and high school or higher, and student and non-stunted. Therefore, totals do have estimated step-down P-values within the pair.

‡ We internally age-standardized both, BSID III receptive vocabulary and TVIP scores in the complete sample and then pooled both in a single regression controlling for measure type.

Table 4. aeioTu intervention. Two-stage least squares estimator with enrollment instrumented with ITT by outcomes and for selected groups

Outcome variables	Females		Males		Less than High school		High School or higher		Stunted		Non-Stunted	
	β (95% CI)	D* (P-value†)	β (95% CI)	D* (P-value†)	β (95% CI)	D* (P-value†)	β (95% CI)	D* (P-value†)	β (95% CI)	D* (P-value†)	β (95% CI)	D* (P-value†)
Nutrition (Z scores)												
Height-for-age	0.023 (-0.172 to 0.219)	0.019 (0.998)	0.161 (-0.078 to 0.401)	0.141 (0.755)	0.116 (-0.093 to 0.325)	0.100 (0.843)	0.048 (-0.178 to 0.274)	0.041 (0.974)	-0.129 (-0.436 to 0.178)	-0.247 (0.902)	0.163 (-0.015 to 0.341)	0.174 (0.397)
BMI-for-age	-0.056 (-0.293 to 0.181)	-0.063 (0.990)	-0.069 (-0.317 to 0.180)	-0.065 (0.990)	-0.172 (-0.387 to 0.043)	-0.174 (0.561)	0.150 (-0.137 to 0.436)	0.152 (0.848)	-0.133 (-0.477 to 0.211)	-0.159 (0.902)	-0.031 (-0.223 to 0.162)	-0.030 (0.958)
Weight-for-age	0.012 (-0.171 to 0.196)	0.012 (0.998)	0.078 (-0.156 to 0.313)	0.069 (0.990)	0.004 (-0.197 to 0.205)	0.004 (0.994)	0.142 (-0.081 to 0.365)	0.126 (0.773)	-0.096 (-0.402 to 0.210)	-0.119 (0.929)	0.103 (-0.066 to 0.272)	0.105 (0.765)
Weight-for-length	-0.003 (-0.234 to 0.228)	-0.004 (0.998)	0.022 (-0.210 to 0.254)	0.020 (0.998)	-0.075 (-0.280 to 0.131)	-0.076 (0.936)	0.188 (-0.085 to 0.461)	0.187 (0.700)	0.071 (-0.251 to 0.393)	0.083 (0.958)	-0.016 (-0.199 to 0.168)	-0.015 (0.981)
Arm circumference	-0.064 (-0.292 to 0.164)	-0.083 (0.990)	-0.016 (-0.233 to 0.201)	-0.018 (0.998)	-0.049 (-0.242 to 0.143)	-0.060 (0.974)	-0.014 (-0.267 to 0.240)	-0.016 (0.994)	-0.185 (-0.498 to 0.127)	-0.244 (0.765)	0.008 (-0.165 to 0.181)	0.010 (0.981)
Cognitive and Language Development: Bayley Scales of Infant Development III (BSID III)												
BSID receptive vocabulary	1.979 (0.215 to 3.743)	0.261 (0.128)	0.998 (-0.720 to 2.716)	0.131 (0.411)	0.551 (-1.073 to 2.174)	0.072 (0.505)	2.869 (1.104 to 4.634)	0.379 (0.009)	1.200 (-1.817 to 4.217)	0.162 (0.587)	1.364 (0.040 to 2.688)	0.177 (0.176)
BSID expressive vocabulary	2.886 (0.628 to 5.144)	0.330 (0.073)	0.415 (-1.547 to 2.377)	0.047 (0.662)	1.230 (-0.769 to 3.228)	0.140 (0.384)	2.390 (0.298 to 4.483)	0.269 (0.090)	-0.938 (-4.106 to 2.230)	-0.106 (0.587)	1.924 (0.283 to 3.565)	0.218 (0.119)
BSID language total	4.802 (1.320 to 8.283)	0.301 (0.014)	1.001 (-2.160 to 4.163)	0.062 (0.519)	1.768 (-1.321 to 4.856)	0.110 (0.251)	5.177 (1.705 to 8.650)	0.320 (0.008)	-0.435 (-6.077 to 5.207)	-0.028 (0.872)	3.322 (0.759 to 5.885)	0.205 (0.029)
BSID cognitive	1.859 (-0.370 to 4.088)	0.137 (0.253)	2.186 (-0.071 to 4.444)	0.142 (0.198)	1.521 (-0.689 to 3.731)	0.102 (0.384)	2.937 (0.766 to 5.109)	0.209 (0.041)	3.204 (-0.796 to 7.204)	0.218 (0.317)	1.281 (-0.393 to 2.956)	0.088 (0.317)
BSID fine motor	2.039 (0.678 to 3.401)	0.228 (0.014)	0.304 (-1.107 to 1.715)	0.032 (0.695)	0.917 (-0.488 to 2.321)	0.096 (0.432)	1.203 (0.014 to 2.391)	0.134 (0.168)	-0.352 (-2.769 to 2.064)	-0.039 (0.755)	1.004 (-0.036 to 2.043)	0.106 (0.193)
BSID gross motor	1.720 (-0.129 to 3.570)	0.144 (0.175)	0.624 (-1.101 to 2.349)	0.046 (0.695)	1.180 (-0.564 to 2.925)	0.089 (0.432)	0.925 (-0.713 to 2.562)	0.075 (0.432)	1.566 (-1.280 to 4.411)	0.124 (0.581)	0.694 (-0.653 to 2.041)	0.053 (0.581)
BSID motor total	3.756 (1.041 to 6.471)	0.183 (0.018)	0.583 (-2.036 to 3.203)	0.026 (0.649)	1.804 (-0.877 to 4.485)	0.080 (0.186)	2.129 (-0.168 to 4.425)	0.102 (0.125)	-0.693 (-5.215 to 3.829)	-0.033 (0.737)	1.683 (-0.303 to 3.669)	0.076 (0.176)
BSID III Total	8.666 (1.938 to 15.395)	0.178 (0.018)	2.954 (-3.653 to 9.562)	0.056 (0.359)	3.350 (-3.177 to 9.877)	0.064 (0.316)	9.769 (3.209 to 16.329)	0.197 (0.007)	0.247 (-13.115 to 13.608)	0.005 (0.967)	5.250 (0.214 to 10.287)	0.102 (0.082)
TVIP	3.546 (0.502 to 6.590)	0.618 (0.051)	-3.939 (-8.223 to 0.345)	-0.477 (0.064)	-0.046 (-3.015 to 2.923)	-0.006 (0.967)	0.590 (-4.996 to 6.175)	0.084 (0.965)	1.145 (-2.829 to 5.120)	0.166 (0.735)	-0.250 (-3.652 to 3.153)	-0.034 (0.884)
Receptive vocabulary‡	0.448 (0.102 to 0.794)	0.486 (0.020)	-0.159 (-0.527 to 0.210)	-0.152 (0.406)	-0.008 (-0.319 to 0.303)	-0.008 (0.959)	0.480 (0.052 to 0.909)	0.480 (0.057)	0.274 (-0.126 to 0.673)	0.277 (0.313)	0.099 (-0.202 to 0.399)	0.100 (0.515)
Socio-Emotional Development: Ages and Stages Socio-emotional (ASQ:SE)												
Self-regulation	4.867 (-0.794 to 10.529)	0.354 (0.676)	-0.447 (-5.684 to 4.791)	-0.029 (0.999)	2.940 (-2.093 to 7.972)	0.197 (0.961)	-0.361 (-6.199 to 5.477)	-0.025 (0.999)	0.633 (-6.482 to 7.749)	0.041 (0.998)	3.046 (-1.382 to 7.473)	0.208 (0.894)
Compliance	0.662 (-1.126 to 2.451)	0.174 (0.996)	0.228 (-1.446 to 1.902)	0.052 (0.999)	0.823 (-0.760 to 2.406)	0.205 (0.975)	0.800 (-1.184 to 2.784)	0.187 (0.996)	1.190 (-0.908 to 3.289)	0.269 (0.952)	-0.025 (-1.512 to 1.463)	-0.006 (0.999)
Communication	-0.448 (-2.578 to 1.682)	-0.129 (0.999)	0.259 (-1.296 to 1.813)	0.061 (0.999)	0.461 (-1.115 to 2.038)	0.112 (0.999)	-1.091 (-3.328 to 1.145)	-0.310 (0.980)	1.100 (-1.182 to 3.383)	0.297 (0.984)	-0.182 (-1.708 to 1.344)	-0.046 (0.998)

Adaptive functioning	2.635 (-0.354 to 5.624)	0.298 (0.672)	-1.477 (-4.317 to 1.363)	-0.177 (0.974)	0.276 (-2.187 to 2.740)	0.032 (0.999)	0.749 (-2.909 to 4.407)	0.090 (0.999)	0.449 (-3.199 to 4.097)	0.051 (0.991)	0.410 (-2.016 to 2.837)	0.048 (0.991)
Autonomy	-0.121 (-1.752 to 1.510)	-0.023 (0.999)	0.503 (-1.385 to 2.392)	0.095 (0.999)	0.413 (-1.146 to 1.971)	0.081 (0.999)	-0.509 (-2.653 to 1.634)	-0.091 (0.999)	0.957 (-1.303 to 3.217)	0.196 (0.991)	-0.198 (-1.636 to 1.239)	-0.037 (0.991)
Affect	0.255 (-1.755 to 2.265)	0.055 (0.999)	-0.186 (-1.801 to 1.429)	-0.038 (0.999)	0.311 (-1.381 to 2.003)	0.066 (0.999)	0.564 (-1.384 to 2.512)	0.118 (0.999)	-0.010 (-2.363 to 2.344)	-0.002 (0.999)	0.333 (-1.131 to 1.796)	0.071 (0.991)
Interaction	0.047 (-3.284 to 3.378)	0.007 (0.999)	-2.351 (-5.491 to 0.790)	-0.327 (0.801)	-0.184 (-3.119 to 2.752)	-0.025 (0.999)	-3.143 (-6.682 to 0.396)	-0.485 (0.633)	-0.844 (-5.117 to 3.429)	-0.119 (0.991)	-1.271 (-3.844 to 1.302)	-0.180 (0.984)
ASQ:SE Total	10.553 (-1.869 to 22.975)	0.442 (0.180)	-3.443 (-13.693 to 6.806)	-0.122 (0.494)	6.173 (-3.845 to 16.191)	0.232 (0.402)	-1.505 (-14.815 to 11.805)	-0.059 (0.820)	5.803 (-9.023 to 20.628)	0.219 (0.639)	3.871 (-5.366 to 13.107)	0.148 (0.639)
Home Observation and Measurement of the Environment (HOME)												
Responsivity	0.183 (-0.705 to 1.070)	0.081 (0.993)	0.344 (-0.487 to 1.175)	0.141 (0.993)	-0.103 (-0.882 to 0.675)	-0.043 (0.977)	0.432 (-0.482 to 1.347)	0.193 (0.977)	-0.490 (-1.599 to 0.618)	-0.219 (0.961)	0.476 (-0.229 to 1.180)	0.199 (0.804)
Acceptance	0.147 (-0.208 to 0.502)	0.158 (0.993)	0.143 (-0.199 to 0.486)	0.154 (0.993)	0.123 (-0.189 to 0.435)	0.134 (0.977)	0.153 (-0.239 to 0.544)	0.162 (0.977)	-0.321 (-0.738 to 0.097)	-0.357 (0.730)	0.214 (-0.076 to 0.505)	0.228 (0.754)
Organization	0.010 (-0.377 to 0.397)	0.009 (0.993)	0.285 (-0.090 to 0.659)	0.265 (0.774)	0.166 (-0.193 to 0.524)	0.155 (0.977)	-0.022 (-0.385 to 0.342)	-0.022 (0.977)	0.188 (-0.285 to 0.661)	0.174 (0.961)	0.150 (-0.161 to 0.461)	0.141 (0.961)
Learning Materials	-0.256 (-0.942 to 0.429)	-0.127 (0.993)	0.194 (-0.418 to 0.806)	0.104 (0.993)	0.331 (-0.220 to 0.882)	0.190 (0.926)	-0.553 (-1.291 to 0.184)	-0.297 (0.801)	0.120 (-0.667 to 0.908)	0.062 (0.961)	-0.040 (-0.572 to 0.493)	-0.021 (0.961)
Involvement	-0.158 (-0.738 to 0.422)	-0.106 (0.993)	-0.000 (-0.525 to 0.524)	-0.000 (0.999)	-0.177 (-0.647 to 0.293)	-0.128 (0.977)	-0.181 (-0.792 to 0.429)	-0.116 (0.977)	-0.191 (-0.837 to 0.454)	-0.130 (0.961)	-0.099 (-0.556 to 0.358)	-0.067 (0.961)
Variety	-0.006 (-0.426 to 0.413)	-0.006 (0.999)	-0.408 (-0.793 to -0.022)	-0.359 (0.363)	-0.150 (-0.513 to 0.213)	-0.132 (0.977)	-0.214 (-0.649 to 0.221)	-0.187 (0.977)	-0.415 (-0.858 to 0.029)	-0.418 (0.499)	-0.156 (-0.498 to 0.187)	-0.132 (0.961)
HOME Total	-0.131 (-1.980 to 1.717)	-0.026 (0.965)	0.218 (-1.577 to 2.013)	0.038 (0.965)	-0.101 (-1.714 to 1.512)	-0.020 (0.936)	-0.324 (-2.319 to 1.671)	-0.061 (0.936)	-1.053 (-3.376 to 1.271)	-0.206 (0.580)	0.314 (-1.177 to 1.806)	0.057 (0.673)

Individual lines present the results of separate regressions for each subpopulation group. Sample sizes vary by instrument due to the collection method (with less households assessed with HOME than children nutritionally) and as children at post testing grow out of the BSID III. For each estimation we present: the means of the outcome by group (treatment and control) and the difference in means, and the results of the estimations for intent-to-treat and treatment-on-treated. β s are estimated controlling for the corresponding pre-test (except for HOME for which there was no pretest), age, age squared, male, black, mother marital status (single), maternal years of education, household wealth index, household kids, childcare experience before randomization, neighborhood, cohort of birth and Tester FE. P-values ≤ 0.10 are in bold type.

*Effects sizes, also known as Cohen's D, are β s interpreted as fraction of SD in control group at baseline (except for HOME for which there was no baseline testing).

†Stepdown P-values are for Romano and Wolf (2005) Stepdown procedures applied to blocks of outcomes per type of developmental dimension measured. Combined receptive vocabulary excluded from the block as it combines two measures of receptive vocabulary and is based on a different sample than other outcomes within that group. In these set of estimations, step-downs are also calculated within subgroup categories in pairs of columns, that is: females and males, less than high school and high school or higher, and student and non-stunted. Therefore, totals do have estimated step-down P-values within the pair.

‡ We internally age-standardized both, BSID III receptive vocabulary and TVIP scores in the complete sample and then pooled both in a single regression controlling for measure type.

Appendix

Milagros Nores
Rutgers, The State University
of New Jersey

Raquel Bernal
Universidad de los Andes
Bogotá, Colombia

Steve Barnett
Rutgers, The State University
of New Jersey

Online Appendix, Table of Contents:

Table A1a. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, females.

Table A1b. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, males.

Table A1c. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for children of mothers with less than high school attainment.

Table A1d. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for children of mothers with high school attainment or above.

Table A1e. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for initially stunted children.

Table A1f. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for initially non-stunted children.

Table A1g. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for younger children.

Table A1h. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for older children.

Table A1i. Comparison of baseline variables before randomization in 2010, by intent-to-treat for the sample of children who did not outgrow BSID III

Table A2. Attrition by random assignment and selected subgroups.

Table A3a. Testing determinants of sample attrition using baseline socio-demographic indicators.

Table A3b. Testing determinants of sample attrition using baseline socio-demographic indicators for selected subgroups.

Table A4. Compliance with random assignment by selected subgroups.

Table A5a. Testing determinants of enrollment using randomized assignment to treatment and baseline socio-demographic indicators.

Table A5b. Testing determinants of enrollment using randomized assignment to treatment and baseline socio-demographic indicators for selected subgroups.

Table A5c. Testing determinants of enrollment using randomized assignment to treatment and baseline socio-demographic indicators for the sample of children who did not outgrow BSIDIII.

Table A5d. Testing determinants of enrollment using randomized assignment to treatment and baseline socio-demographic indicators or the sample of children who did not outgrow BSIDIII for selected subgroups.

Table A6. The treatment in context: type of childcare used at follow up by study group.

Table A7a. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for females.

Table A8. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for males.

Table A9. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for children of mothers with less than high school attainment.

Table A10. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for children of mothers with high school attainment or above.

Table A11. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for children stunted at baseline.

Table A12. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for initially non-stunted children.

Table A13. ITT and TOT estimations of program effects by outcome without Tester Fixed Effects

Table A14. ITT and TOT estimations of program effects on nutrition, ASQ:SE and HOME on the subsample of children who did not outgrow BSID III.

Table A1a. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, females.

Socio-Demographics and Outcomes at Baseline	N	Control		Treatment		P-value
		Mean	SD	Mean	SD	
Child's age in months	383	20.40	8.57	20.81	9.57	0.661
Child's gender (male)	383	0.00	0.00	0.00	0.00	-
Child's race (black)	383	0.59	0.49	0.56	0.50	0.552
Maternal marital status (single)	383	0.28	0.45	0.26	0.44	0.582
Health insurance for child	383	0.78	0.42	0.78	0.41	0.877
Mother secondary incomplete	383	0.59	0.49	0.64	0.48	0.318
Mother secondary complete and above	383	0.41	0.49	0.36	0.48	0.318
Wealth Index‡	383	-0.06	4.53	0.48	5.54	0.300
Children books at home	382	1.41	2.42	1.41	2.53	0.976
Mother education years	383	8.50	3.34	8.29	3.12	0.540
No. of children <=5 yrs	383	2.59	0.75	2.74	0.87	0.076
Childcare by baseline	383	0.11	0.31	0.13	0.34	0.470
Neighborhood (La Paz)	383	0.57	0.50	0.47	0.50	0.058
Neighborhood (Alpes B)	383	0.07	0.25	0.08	0.27	0.685
Neighborhood (Timayui 1)	383	0.19	0.39	0.24	0.43	0.203
Neighborhood (Timayui 2)	383	0.17	0.38	0.21	0.41	0.411
Cohort 2008	383	0.39	0.49	0.40	0.49	0.818
Cohort 2009	383	0.46	0.50	0.30	0.46	0.002
Cohort 2010	383	0.09	0.28	0.17	0.38	0.013
<i>Nutrition (Z scores)</i>						
Length/Height-for-age	372	-1.00	1.20	-1.08	0.97	0.458
BMI-for-age	368	0.54	0.89	0.56	0.96	0.809
Weight-for-age	374	-0.23	1.00	-0.25	0.98	0.820
Weight-for-length	369	0.40	0.89	0.44	0.98	0.649
Arm circumference	367	0.29	0.77	0.38	0.82	0.309
<i>Infant development: BSID III Raw scores</i>						
Receptive vocabulary	373	19.58	7.57	20.55	7.88	0.234
Expressive vocabulary	373	20.29	8.74	20.69	10.11	0.683
Total language	370	39.84	15.93	41.37	17.55	0.385
Cognitive	373	49.31	13.59	50.22	15.25	0.546
Fine motor	372	33.31	8.96	33.46	9.44	0.875
Gross motor	373	47.81	11.93	47.71	13.70	0.941
Total motor	370	81.13	20.57	81.05	22.72	0.971
BSID III Total	363	170.32	48.70	172.84	53.52	0.642
<i>Infant development: BSID III Scaled scores</i>						
Receptive vocabulary	372	8.30	2.42	8.62	2.17	0.201
Expressive vocabulary	372	8.15	2.58	7.86	2.78	0.313
Total language	369	16.45	4.43	16.47	4.27	0.967
Cognitive	372	8.10	2.57	8.27	2.74	0.543
Fine motor	371	9.10	2.56	9.17	2.35	0.787
Gross motor	372	9.08	2.98	8.97	2.89	0.738
Total motor	369	18.19	4.80	18.12	4.24	0.877
BSID III Total	362	42.39	10.39	43.29	9.45	0.400
<i>Socio-emotional Development (ASQ: SE)</i>						
Self-regulation	380	15.30	13.75	16.67	13.68	0.341
Compliance	295	1.89	3.81	2.67	4.58	0.113
Communication	380	1.39	3.49	2.00	4.26	0.132
Adaptive functioning	379	6.91	8.83	6.47	8.27	0.625
Autonomy	295	4.09	5.25	5.13	5.99	0.116
Affect	380	3.56	4.62	3.83	4.91	0.587
Interaction	380	6.00	6.92	5.71	7.17	0.694
ASQ:SE Total	379	38.80	23.87	41.65	26.25	0.274

Table A1b. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, males.

Socio-Demographics and Outcomes at Baseline	N	Control		Treatment		P-value
		Mean	SD	Mean	SD	
Child's age in months	436	19.27	9.18	20.52	9.94	0.178
Child's gender (male)	436	1.00	0.00	1.00	0.00	
Child's race (black)	436	0.60	0.49	0.66	0.47	0.181
Maternal marital status (single)	436	0.29	0.46	0.28	0.45	0.685
Health insurance for child	436	0.77	0.42	0.76	0.43	0.806
Mother secondary incomplete	436	0.67	0.47	0.64	0.48	0.578
Mother secondary complete and above	436	0.33	0.47	0.36	0.48	0.578
Wealth Index‡	436	0.56	5.15	-0.50	3.86	0.020
Children books at home	436	1.34	2.13	1.41	3.48	0.790
Mother education years	436	8.33	3.27	8.38	3.13	0.858
No. of children <=5 yrs	436	2.64	0.78	2.81	0.85	0.034
Childcare by baseline	436	0.10	0.30	0.14	0.35	0.187
Neighborhood (La Paz)	436	0.59	0.49	0.49	0.50	0.028
Neighborhood (Alpes B)	436	0.04	0.20	0.07	0.25	0.287
Neighborhood (Timayui 1)	436	0.19	0.39	0.20	0.40	0.675
Neighborhood (Timayui 2)	436	0.18	0.38	0.24	0.43	0.089
Cohort 2008	436	0.38	0.49	0.35	0.48	0.514
Cohort 2009	436	0.36	0.48	0.34	0.47	0.552
Cohort 2010	436	0.17	0.38	0.17	0.37	0.937
<i>Nutrition (Z scores)</i>						
Length/Height-for-age	426	-1.17	1.15	-1.11	1.09	0.626
BMI-for-age	417	0.61	1.06	0.45	1.02	0.135
Weight-for-age	423	-0.25	1.13	-0.39	1.02	0.189
Weight-for-length	420	0.47	1.07	0.31	1.00	0.109
Arm circumference	419	0.22	0.89	0.15	0.81	0.426
<i>Infant development: BSID III Raw scores</i>						
Receptive vocabulary	417	17.98	7.60	19.64	8.26	0.034
Expressive vocabulary	421	18.22	8.80	19.80	9.89	0.086
Total language	412	36.20	16.05	39.72	17.76	0.036
Cognitive	425	46.74	15.36	48.28	15.79	0.313
Fine motor	421	31.02	9.58	32.45	10.46	0.147
Gross motor	424	45.54	13.72	46.91	14.34	0.320
Total motor	419	76.50	22.87	79.38	24.47	0.219
BSID III Total	404	159.19	52.74	168.71	55.40	0.081
<i>Infant development: BSID III Scaled scores</i>						
Receptive vocabulary	416	7.95	2.73	8.38	2.19	0.085
Expressive vocabulary	420	7.61	2.80	7.92	2.50	0.250
Total language	411	15.54	4.99	16.35	4.10	0.085
Cognitive	424	7.89	2.84	8.15	2.42	0.321
Fine motor	420	8.34	2.52	8.92	2.27	0.015
Gross motor	423	8.79	2.98	9.12	2.51	0.228
Total motor	418	17.10	4.58	18.02	3.96	0.033
BSID III Total	403	40.39	10.33	41.88	9.29	0.138
<i>Socio-emotional Development (ASQ: SE)</i>						
Self-regulation	430	17.68	15.63	19.62	17.87	0.233
Compliance	308	2.32	4.38	2.83	4.71	0.329
Communication	430	1.58	4.27	1.75	3.65	0.651
Adaptive functioning	430	6.77	8.33	6.99	8.36	0.787
Autonomy	308	4.25	5.29	5.19	5.57	0.130
Affect	430	3.96	4.85	3.78	4.88	0.718
Interaction	430	5.06	7.19	5.28	6.79	0.757
ASQ:SE Total	430	41.03	28.25	44.47	30.14	0.225

Table A1c. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for children of mothers with less than high school attainment.

Socio-Demographics and Outcomes at Baseline	N	Control		Treatment		P-value
		Mean	SD	Mean	SD	
Child's age in months	520	19.74	9.08	20.97	9.54	0.139
Child's gender (male)	520	0.56	0.50	0.54	0.50	0.617
Child's race (black)	520	0.62	0.49	0.63	0.48	0.795
Maternal marital status (single)	520	0.29	0.46	0.32	0.47	0.515
Health insurance for child	520	0.74	0.44	0.77	0.42	0.461
Mother secondary incomplete	520	1.00	0.00	1.00	0.00	-
Mother secondary complete and above	520	0.00	0.00	0.00	0.00	-
Wealth Index‡	520	-0.44	4.66	-0.28	5.07	0.720
Children books at home	519	1.13	1.98	1.17	2.22	0.841
Mother education years	520	6.42	2.27	6.56	2.33	0.500
No. of children <=5 yrs	520	2.70	0.81	2.85	0.90	0.058
Childcare by baseline	520	0.08	0.27	0.13	0.34	0.058
Neighborhood (La Paz)	520	0.56	0.50	0.47	0.50	0.033
Neighborhood (Alpes B)	520	0.05	0.21	0.07	0.26	0.178
Neighborhood (Timayui 1)	520	0.19	0.39	0.23	0.42	0.261
Neighborhood (Timayui 2)	520	0.20	0.40	0.23	0.42	0.472
Cohort 2008	520	0.37	0.48	0.38	0.49	0.795
Cohort 2009	520	0.40	0.49	0.34	0.48	0.174
Cohort 2010	520	0.14	0.35	0.14	0.35	0.981
<i>Nutrition (Z scores)</i>						
Length/Height-for-age	503	-1.17	1.16	-1.23	1.05	0.584
BMI-for-age	497	0.61	0.99	0.44	0.98	0.057
Weight-for-age	506	-0.29	1.03	-0.46	1.02	0.077
Weight-for-length	499	0.46	0.98	0.28	0.98	0.043
Arm circumference	500	0.25	0.82	0.18	0.80	0.377
<i>Infant development: BSID III Raw scores</i>						
Receptive vocabulary	502	18.51	7.65	19.96	7.72	0.037
Expressive vocabulary	504	18.94	8.78	19.94	9.61	0.228
Total language	499	37.44	16.03	40.01	16.91	0.084
Cognitive	503	47.69	14.91	49.28	14.67	0.235
Fine motor	500	31.98	9.59	33.30	9.40	0.126
Gross motor	504	46.30	13.32	47.84	13.08	0.197
Total motor	497	78.17	22.56	81.15	22.11	0.144
BSID III Total	488	163.01	51.96	170.82	51.35	0.100
<i>Infant development: BSID III Scaled scores</i>						
Receptive vocabulary	501	8.00	2.56	8.39	2.29	0.081
Expressive vocabulary	503	7.78	2.61	7.67	2.70	0.656
Total language	498	15.79	4.60	16.12	4.34	0.426
Cognitive	502	8.00	2.76	7.99	2.71	0.969
Fine motor	499	8.72	2.64	9.01	2.29	0.200
Gross motor	503	8.86	2.94	8.96	2.65	0.678
Total motor	496	17.55	4.74	17.95	4.05	0.328
BSID III Total	487	41.22	10.67	41.89	9.59	0.477
<i>Socio-emotional Development (ASQ: SE)</i>						
Self-regulation	512	16.98	14.92	18.03	14.94	0.434
Compliance	383	2.08	4.01	2.50	4.57	0.342
Communication	512	1.48	4.13	1.96	4.18	0.200
Adaptive functioning	512	6.91	8.70	6.59	8.37	0.673
Autonomy	383	4.03	5.07	4.69	5.51	0.223
Affect	512	3.72	4.72	4.02	4.96	0.497
Interaction	512	5.81	7.39	5.63	7.12	0.778
ASQ:SE Total	512	40.59	26.60	42.67	27.47	0.388

Table A1d. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for children of mothers with high school attainment or above.

Socio-Demographics and Outcomes at Baseline	N	Control		Treatment		P-value
		Mean	SD	Mean	SD	
Child's age in months	299	19.90	8.62	20.09	10.16	0.864
Child's gender (male)	299	0.48	0.50	0.54	0.50	0.313
Child's race (black)	299	0.56	0.50	0.60	0.49	0.570
Maternal marital status (single)	299	0.28	0.45	0.17	0.38	0.032
Health insurance for child	299	0.83	0.38	0.78	0.42	0.240
Mother secondary incomplete	299	0.00	0.00	0.00	0.00	-
Mother secondary complete and above	299	1.00	0.00	1.00	0.00	-
Wealth Index‡	299	1.47	4.99	0.36	4.04	0.044
Children books at home	299	1.77	2.65	1.83	4.17	0.869
Mother education years	299	11.81	1.53	11.53	1.25	0.097
No. of children <=5 yrs	299	2.47	0.67	2.65	0.77	0.032
Childcare by baseline	299	0.15	0.36	0.16	0.37	0.900
Neighborhood (La Paz)	299	0.62	0.49	0.50	0.50	0.051
Neighborhood (Alpes B)	299	0.07	0.25	0.07	0.25	0.965
Neighborhood (Timayui 1)	299	0.19	0.39	0.21	0.41	0.650
Neighborhood (Timayui 2)	299	0.13	0.34	0.22	0.42	0.033
Cohort 2008	299	0.42	0.50	0.37	0.49	0.394
Cohort 2009	299	0.42	0.50	0.28	0.45	0.013
Cohort 2010	299	0.12	0.32	0.22	0.42	0.015
<i>Nutrition (Z scores)</i>						
Length/Height-for-age	295	-0.94	1.18	-0.87	0.97	0.596
BMI-for-age	288	0.52	0.98	0.62	1.01	0.396
Weight-for-age	291	-0.15	1.12	-0.09	0.93	0.630
Weight-for-length	290	0.41	1.00	0.54	0.99	0.275
Arm circumference	286	0.26	0.87	0.39	0.84	0.218
<i>Infant development: BSID III Raw scores</i>						
Receptive vocabulary	288	19.13	7.57	20.25	8.73	0.250
Expressive vocabulary	290	19.64	8.89	20.71	10.66	0.355
Total language	283	38.79	16.18	41.35	18.98	0.224
Cognitive	295	48.38	14.08	48.99	17.05	0.738
Fine motor	293	32.31	8.97	32.26	10.97	0.969
Gross motor	293	47.13	12.31	46.31	15.58	0.615
Total motor	292	79.52	20.84	78.42	26.13	0.690
BSID III Total	279	167.03	49.66	170.33	60.01	0.617
<i>Infant development: BSID III Scaled scores</i>						
Receptive vocabulary	287	8.32	2.65	8.69	1.97	0.205
Expressive vocabulary	289	8.01	2.87	8.29	2.47	0.389
Total language	282	16.29	4.99	16.93	3.82	0.249
Cognitive	294	7.97	2.66	8.59	2.25	0.038
Fine motor	292	8.66	2.46	9.09	2.34	0.135
Gross motor	292	9.03	3.06	9.21	2.77	0.614
Total motor	291	17.71	4.67	18.27	4.17	0.296
BSID III Total	278	41.54	9.94	43.75	8.90	0.060
<i>Socio-emotional Development (ASQ: SE)</i>						
Self-regulation	298	15.82	14.62	18.68	18.08	0.134
Compliance	220	2.14	4.28	3.22	4.76	0.083
Communication	298	1.50	3.52	1.69	3.49	0.634
Adaptive functioning	297	6.71	8.36	7.04	8.24	0.735
Autonomy	220	4.40	5.57	6.03	6.15	0.042
Affect	298	3.84	4.79	3.43	4.74	0.465
Interaction	298	5.00	6.48	5.21	6.69	0.788
ASQ:SE Total	297	38.93	25.70	44.09	30.13	0.115

Table A1e. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for initially stunted children.

Socio-Demographics and Outcomes at Baseline	N	Control		Treatment		P-value
		Mean	SD	Mean	SD	
Child's age in months	173	21.52	8.84	23.57	8.68	0.134
Child's gender (male)	173	0.57	0.50	0.54	0.50	0.689
Child's race (black)	173	0.65	0.48	0.71	0.46	0.441
Maternal marital status (single)	173	0.30	0.46	0.20	0.41	0.164
Health insurance for child	173	0.78	0.42	0.71	0.46	0.309
Mother secondary incomplete	173	0.72	0.45	0.74	0.44	0.796
Mother secondary complete and above	173	0.28	0.45	0.26	0.44	0.796
Wealth Index‡	173	-0.62	4.68	-0.74	3.38	0.861
Children books at home	173	1.09	2.29	1.22	2.20	0.709
Mother education years	173	7.67	3.36	8.03	2.74	0.464
No. of children <=5 yrs	173	2.80	0.79	3.16	0.98	0.008
Childcare by baseline	173	0.09	0.28	0.19	0.39	0.049
Neighborhood (La Paz)	173	0.58	0.50	0.41	0.49	0.028
Neighborhood (Alpes B)	173	0.07	0.25	0.07	0.26	0.897
Neighborhood (Timayui 1)	173	0.18	0.39	0.20	0.41	0.742
Neighborhood (Timayui 2)	173	0.17	0.38	0.32	0.47	0.026
Cohort 2008	173	0.38	0.49	0.46	0.50	0.304
Cohort 2009	173	0.42	0.50	0.29	0.46	0.076
Cohort 2010	173	0.10	0.30	0.07	0.26	0.590
<i>Nutrition (Z scores)</i>						
Length/Height-for-age	173	-2.58	0.52	-2.52	0.46	0.428
BMI-for-age	171	0.58	0.84	0.43	1.00	0.294
Weight-for-age	172	-1.17	0.81	-1.27	0.74	0.414
Weight-for-length	171	0.26	0.86	0.10	1.01	0.243
Arm circumference	170	-0.21	0.76	-0.17	0.79	0.705
<i>Infant development: BSID III Raw scores</i>						
Receptive vocabulary	167	19.32	7.39	22.17	7.50	0.017
Expressive vocabulary	167	19.62	8.85	22.48	10.05	0.055
Total language	164	38.94	15.81	44.76	17.27	0.028
Cognitive	167	48.94	14.70	54.58	12.80	0.012
Fine motor	167	32.82	8.97	35.92	7.81	0.024
Gross motor	170	47.53	12.57	51.44	9.71	0.033
Total motor	166	80.15	21.03	87.95	16.70	0.013
BSID III Total	159	167.26	49.84	188.72	44.61	0.007
<i>Infant development: BSID III Scaled scores</i>						
Receptive vocabulary	167	7.74	2.60	8.45	1.57	0.049
Expressive vocabulary	167	7.27	2.63	7.84	2.78	0.183
Total language	164	15.03	4.72	16.40	3.77	0.054
Cognitive	167	7.26	2.58	8.48	2.05	0.002
Fine motor	167	8.21	2.64	9.14	1.91	0.016
Gross motor	170	8.23	2.98	8.92	2.21	0.105
Total motor	166	16.35	4.66	18.16	2.94	0.006
BSID III Total	159	37.71	11.97	41.97	6.92	0.013
<i>Socio-emotional Development (ASQ: SE)</i>						
Self-regulation	172	17.77	15.30	16.54	12.19	0.578
Compliance	144	2.95	4.42	1.97	3.33	0.146
Communication	172	1.55	3.71	1.70	3.75	0.797
Adaptive functioning	172	6.07	8.79	5.65	8.61	0.760
Autonomy	144	3.86	4.90	4.84	5.40	0.257
Affect	172	3.69	4.90	2.97	4.14	0.318
Interaction	172	6.21	7.09	4.63	5.38	0.116
ASQ:SE Total	172	42.26	26.49	38.35	22.50	0.316

Table A1f. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for initially non-stunted children.

Socio-Demographics and Outcomes at Baseline	N	Control		Treatment		P-value
		Mean	SD	Mean	SD	
Child's age in months	646	19.33	8.87	19.90	9.90	0.444
Child's gender (male)	646	0.52	0.50	0.54	0.50	0.638
Child's race (black)	646	0.58	0.49	0.59	0.49	0.775
Maternal marital status (single)	646	0.29	0.45	0.28	0.45	0.953
Health insurance for child	646	0.77	0.42	0.79	0.41	0.655
Mother secondary incomplete	646	0.61	0.49	0.62	0.49	0.801
Mother secondary complete and above	646	0.39	0.49	0.38	0.49	0.801
Wealth Index‡	646	0.51	4.90	0.13	5.01	0.332
Children books at home	645	1.45	2.26	1.46	3.27	0.950
Mother education years	646	8.61	3.26	8.42	3.21	0.464
No. of children <=5 yrs	646	2.57	0.75	2.68	0.80	0.074
Childcare by baseline	646	0.11	0.31	0.13	0.33	0.541
Neighborhood (La Paz)	646	0.58	0.49	0.50	0.50	0.033
Neighborhood (Alpes B)	646	0.05	0.22	0.07	0.26	0.273
Neighborhood (Timayui 1)	646	0.19	0.39	0.23	0.42	0.251
Neighborhood (Timayui 2)	646	0.17	0.38	0.20	0.40	0.387
Cohort 2008	646	0.39	0.49	0.35	0.48	0.374
Cohort 2009	646	0.40	0.49	0.33	0.47	0.048
Cohort 2010	646	0.14	0.35	0.19	0.40	0.068
<i>Nutrition (Z scores)</i>						
Length/Height-for-age	625	-0.66	0.94	-0.72	0.79	0.395
BMI-for-age	614	0.57	1.02	0.52	0.99	0.540
Weight-for-age	625	0.02	0.98	-0.08	0.91	0.169
Weight-for-length	618	0.49	1.02	0.44	0.98	0.585
Arm circumference	616	0.38	0.81	0.37	0.79	0.816
<i>Infant development: BSID III Raw scores</i>						
Receptive vocabulary	623	18.57	7.69	19.54	8.16	0.132
Expressive vocabulary	627	19.08	8.82	19.66	9.91	0.442
Total language	618	37.65	16.17	39.45	17.62	0.190
Cognitive	631	47.67	14.57	47.84	15.89	0.888
Fine motor	626	31.90	9.46	32.19	10.34	0.714
Gross motor	627	46.34	13.05	46.25	14.75	0.933
Total motor	623	78.25	22.18	78.28	24.69	0.988
BSID III Total	608	163.73	51.49	166.33	55.80	0.554
<i>Infant development: BSID III Scaled scores</i>						
Receptive vocabulary	621	8.23	2.58	8.51	2.32	0.166
Expressive vocabulary	625	8.03	2.71	7.90	2.60	0.552
Total language	616	16.24	4.73	16.40	4.28	0.657
Cognitive	629	8.19	2.73	8.14	2.68	0.817
Fine motor	624	8.83	2.53	9.02	2.40	0.369
Gross motor	625	9.12	2.96	9.08	2.81	0.872
Total motor	621	17.98	4.67	18.05	4.32	0.847
BSID III Total	606	42.35	9.69	42.68	9.88	0.680
<i>Socio-emotional Development (ASQ: SE)</i>						
Self-regulation	638	16.21	14.66	18.71	16.98	0.047
Compliance	459	1.85	3.98	3.01	4.97	0.006
Communication	638	1.47	3.97	1.91	3.99	0.173
Adaptive functioning	637	7.05	8.50	7.04	8.22	0.986
Autonomy	459	4.26	5.38	5.27	5.89	0.059
Affect	638	3.79	4.70	4.02	5.05	0.551
Interaction	638	5.32	7.06	5.70	7.31	0.506
ASQ:SE Total	637	39.33	26.19	44.44	29.67	0.022

Table A1g. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for younger children.

Socio-Demographics and Outcomes at Baseline	N	Control		Treatment		P-value
		Mean	SD	Mean	SD	
Child's age in months	408	12.38	5.01	12.08	5.55	0.576
Child's gender (male)	408	0.54	0.50	0.56	0.50	0.685
Child's race (black)	408	0.57	0.50	0.62	0.49	0.353
Maternal marital status (single)	408	0.30	0.46	0.29	0.46	0.837
Health insurance for child	408	0.76	0.43	0.73	0.45	0.438
Mother secondary incomplete	408	0.63	0.48	0.64	0.48	0.824
Mother secondary complete and above	408	0.37	0.48	0.36	0.48	0.824
Wealth Index‡	408	0.39	4.72	0.32	5.32	0.877
Children books at home	408	1.38	2.30	1.41	3.52	0.935
Mother education years	408	8.56	3.18	8.35	3.19	0.513
No. of children <=5 yrs	408	2.62	0.80	2.75	0.85	0.116
Childcare by baseline	408	0.05	0.21	0.04	0.20	0.891
Neighborhood (La Paz)	408	0.52	0.50	0.50	0.50	0.612
Neighborhood (Alpes B)	408	0.07	0.25	0.05	0.23	0.641
Neighborhood (Timayui 1)	408	0.23	0.42	0.22	0.41	0.699
Neighborhood (Timayui 2)	408	0.18	0.38	0.23	0.42	0.186
Cohort 2008	408	0.03	0.17	0.04	0.20	0.460
Cohort 2009	408	0.71	0.45	0.61	0.49	0.026
Cohort 2010	408	0.26	0.44	0.35	0.48	0.061
<i>Nutrition (Z scores)</i>						
Length/Height-for-age	395	-0.96	1.22	-0.97	1.04	0.935
BMI-for-age	388	0.60	1.01	0.60	1.09	0.991
Weight-for-age	397	-0.10	1.11	-0.17	1.03	0.553
Weight-for-length	389	0.51	1.03	0.51	1.09	0.971
Arm circumference	383	0.35	0.86	0.42	0.82	0.460
<i>Infant development: BSID III Raw scores</i>						
Receptive vocabulary	237	11.41	3.08	11.22	2.98	0.639
Expressive vocabulary	237	10.62	4.60	9.81	4.54	0.190
Total language	234	21.98	7.26	21.11	7.05	0.371
Cognitive	242	32.73	10.69	31.56	10.59	0.407
Fine motor	240	22.64	7.73	21.69	7.55	0.350
Gross motor	237	33.83	12.81	31.01	12.21	0.094
Total motor	237	56.45	20.25	52.79	19.39	0.170
BSID III Total	231	110.73	36.88	106.24	36.07	0.366
<i>Infant development: BSID III Scaled scores</i>						
Receptive vocabulary	237	8.83	2.89	9.40	2.60	0.129
Expressive vocabulary	237	8.87	3.02	8.91	2.92	0.916
Total language	234	17.69	5.28	18.26	4.60	0.402
Cognitive	242	9.20	3.25	9.93	2.67	0.070
Fine motor	240	8.81	2.77	9.51	2.50	0.050
Gross motor	237	9.81	3.16	9.54	2.89	0.519
Total motor	237	18.64	5.16	19.00	4.45	0.583
BSID III Total	231	44.35	10.20	46.25	9.50	0.159
<i>Socio-emotional Development (ASQ: SE)</i>						
Self-regulation	405	11.84	12.31	11.69	10.91	0.900
Compliance	201	1.20	3.03	0.81	2.44	0.342
Communication	405	1.34	3.79	1.33	3.04	0.973
Adaptive functioning	404	9.51	9.61	9.04	8.80	0.615
Autonomy	201	1.07	2.90	0.81	2.81	0.526
Affect	405	5.06	4.88	4.91	5.18	0.762
Interaction	405	3.72	6.57	4.04	6.42	0.629
ASQ:SE Total	404	33.51	24.74	32.37	22.90	0.641

Table A1h. Comparison of baseline variables, before randomization, 2010, by intent-to-treat for the baseline sample, for older children.

Socio-Demographics and Outcomes at Baseline	N	Control		Treatment		P-value
		Mean	SD	Mean	SD	
Child's age in months	411	27.35	4.61	28.87	4.34	0.001
Child's gender (male)	411	0.51	0.50	0.51	0.50	0.952
Child's race (black)	411	0.62	0.49	0.62	0.49	0.883
Maternal marital status (single)	411	0.28	0.45	0.24	0.43	0.469
Health insurance for child	411	0.79	0.41	0.81	0.39	0.497
Mother secondary incomplete	411	0.64	0.48	0.65	0.48	0.846
Mother secondary complete and above	411	0.36	0.48	0.35	0.48	0.846
Wealth Index‡	411	0.14	5.02	-0.40	4.07	0.248
Children books at home	410	1.36	2.23	1.42	2.59	0.804
Mother education years	411	8.26	3.42	8.34	3.06	0.812
No. of children <=5 yrs	411	2.62	0.74	2.80	0.87	0.019
Childcare by baseline	411	0.17	0.37	0.23	0.42	0.100
Neighborhood (La Paz)	411	0.64	0.48	0.47	0.50	0.000
Neighborhood (Alpes B)	411	0.04	0.20	0.09	0.28	0.058
Neighborhood (Timayui 1)	411	0.14	0.35	0.23	0.42	0.027
Neighborhood (Timayui 2)	411	0.17	0.38	0.22	0.42	0.211
Cohort 2008	411	0.75	0.43	0.70	0.46	0.213
Cohort 2009	411	0.10	0.30	0.05	0.21	0.044
Cohort 2010	411	0.00	0.00	0.00	0.00	-
<i>Nutrition (Z scores)</i>						
Length/Height-for-age	403	-1.22	1.11	-1.22	1.01	0.971
BMI-for-age	397	0.55	0.96	0.41	0.87	0.149
Weight-for-age	400	-0.38	1.01	-0.48	0.95	0.323
Weight-for-length	400	0.37	0.95	0.24	0.87	0.160
Arm circumference	403	0.15	0.80	0.10	0.79	0.561
<i>Infant development: BSID III Raw scores</i>						
Receptive vocabulary	240	18.84	4.71	18.41	3.74	0.461
Expressive vocabulary	243	20.01	4.93	18.56	5.25	0.032
Total language	238	38.87	8.99	36.93	8.25	0.099
Cognitive	245	50.99	6.09	49.90	5.72	0.168
Fine motor	241	34.17	3.49	34.17	2.87	0.999
Gross motor	243	50.44	3.50	50.62	3.60	0.708
Total motor	240	84.61	6.06	84.85	5.69	0.756
BSID III Total	233	174.63	18.69	171.80	16.98	0.251
<i>Infant development: BSID III Scaled scores</i>						
Receptive vocabulary	240	7.26	2.30	7.39	2.03	0.661
Expressive vocabulary	243	7.13	2.06	6.67	2.46	0.122
Total language	238	14.39	3.83	14.07	4.00	0.540
Cognitive	245	7.46	2.28	7.37	2.42	0.785
Fine motor	241	8.60	2.53	8.94	2.20	0.292
Gross motor	243	8.36	2.08	8.73	2.16	0.186
Total motor	240	16.97	3.78	17.68	3.72	0.157
BSID III Total	233	40.12	11.55	40.20	10.93	0.957
<i>Socio-emotional Development (ASQ: SE)</i>						
Self-regulation	405	21.40	15.59	24.57	17.74	0.057
Compliance	402	2.58	4.50	3.67	5.13	0.024
Communication	405	1.65	4.03	2.38	4.59	0.087
Adaptive functioning	405	4.08	6.26	4.57	7.19	0.466
Autonomy	402	5.77	5.50	7.22	5.67	0.011
Affect	405	2.44	4.20	2.75	4.34	0.466
Interaction	405	7.35	7.10	6.86	7.19	0.489
ASQ:SE Total	405	46.62	26.15	53.48	29.38	0.014

Table A1i. Comparison of baseline variables, before randomization in 2010, by intent-to-treat for the sample of children who did not outgrow BSID III.

Socio-Demographics and Outcomes at Baseline	N	Control		Treatment		P-value
		Mean	SD	Mean	SD	
Child's age in months	487	15.35	6.35	14.36	6.62	0.101
Child's gender (male)	487	0.53	0.50	0.55	0.50	0.772
Child's race (black)	487	0.59	0.49	0.63	0.48	0.292
Maternal marital status (single)	487	0.28	0.45	0.29	0.45	0.874
Health insurance for child	487	0.76	0.43	0.78	0.42	0.685
Mother secondary incomplete	487	0.64	0.48	0.65	0.48	0.844
Mother secondary complete and above	487	0.36	0.48	0.35	0.48	0.844
Wealth Index‡	487	0.33	4.72	-0.03	4.23	0.394
Children books at home	487	1.43	2.33	1.42	3.58	0.954
Mother education years	487	8.41	3.24	8.19	3.16	0.460
No. of children <=5 yrs	487	2.62	0.75	2.75	0.85	0.073
Childcare by baseline	487	0.07	0.25	0.06	0.24	0.804
Neighborhood (La Paz)	487	0.57	0.50	0.55	0.50	0.766
Neighborhood (Alpes B)	487	0.05	0.21	0.07	0.25	0.354
Neighborhood (Timayui 1)	487	0.19	0.40	0.16	0.37	0.406
Neighborhood (Timayui 2)	487	0.19	0.40	0.22	0.41	0.516
Cohort 2008	487	0.23	0.42	0.19	0.39	0.247
Cohort 2009	487	0.59	0.49	0.54	0.50	0.270
Cohort 2010	487	0.18	0.38	0.27	0.45	0.013
<i>Nutrition (Z scores)</i>						
Length/Height-for-age	475	-0.99	1.20	-0.95	1.07	0.721
BMI-for-age	467	0.62	0.98	0.62	1.09	0.971
Weight-for-age	478	-0.12	1.08	-0.14	1.07	0.848
Weight-for-length	469	0.49	0.98	0.53	1.09	0.722
Arm circumference	465	0.34	0.86	0.42	0.84	0.352
<i>Infant development: BSID III Raw scores</i>						
Receptive vocabulary	477	15.20	5.46	14.75	4.93	0.370
Expressive vocabulary	480	15.46	6.70	14.11	6.57	0.033
Total language	472	30.60	11.77	28.93	11.02	0.128
Cognitive	487	42.10	12.57	40.48	12.55	0.168
Fine motor	481	28.52	8.29	27.76	8.50	0.333
Gross motor	480	42.38	12.44	40.65	13.35	0.151
Total motor	477	70.86	20.42	68.47	21.56	0.225
BSID III Total	464	143.12	43.23	138.64	43.32	0.281
<i>Socio-emotional Development (ASQ: SE)</i>						
Self-regulation	484	13.37	12.52	12.66	11.40	0.535
Compliance	301	1.04	2.79	1.06	2.81	0.968
Communication	484	1.39	3.73	1.41	3.38	0.953
Adaptive functioning	483	8.21	9.25	8.59	8.94	0.662
Autonomy	301	1.17	3.32	1.01	2.95	0.671
Affect	484	4.82	4.97	5.16	5.13	0.463
Interaction	484	4.43	6.56	4.44	6.52	0.983
ASQ:SE Total	483	34.51	23.73	34.26	23.21	0.908

Table A2. Attrition by random assignment and selected subgroups. We show the number of children that we were not able to follow-up in the second year of the study, by gender, mother’s education and stunting. Percentages in parenthesis. We also test for distribution differences in the attrition indicator by ITT. P-values ≤ 0.10 are in bold type.

Variables	Control	Treatment	P-value
Females	15 (7%)	8 (5%)	0.550
Males	19 (7%)	14 (8%)	0.912
Less than High school	21 (7%)	14 (6%)	0.951
High School or higher	13 (7%)	8 (7%)	0.451
Stunted	10 (10%)	5 (7%)	0.590
Non-stunted	24 (6%)	17 (6%)	0.998
Did not outgrow BSID III	22 (6%)	13 (6%)	0.770
Total	34 (7%)	22 (7%)	0.770

Table A3a. Testing determinants of sample attrition using baseline socio-demographic indicators.

Dep. Variable -> Attrition	(1)	(2)
ITT	-0.005 (0.770)	-0.015 (0.414)
Age		0.001 (0.939)
Age Squared		-0.000 (0.959)
Male		0.017 (0.352)
Black		-0.007 (0.695)
Single Mom		-0.012 (0.530)
Mom above secondary		0.002 (0.914)
Wealth Index		0.000 (0.951)
Children in the household		0.012 (0.267)
Childcare before baseline		0.010 (0.720)
Neighborhood (Alpes B)		0.110 (0.004)
Neighborhood (Timayui 1)		0.068 (0.005)
Neighborhood (Timayui 2)		-0.013 (0.575)
Birth cohort 2008		0.013 (0.776)
Birth cohort 2009		-0.020 (0.753)
Birth cohort 2010		0.051 (0.531)
Constant	0.071 (0.000)	0.009 (0.936)
Observations	819	819
R-squared	0.000	0.033
F-stat	0.086	1.697
Prob > F	0.770	0.0423

The dependent variable is whether the observation was lost at follow-up.

P-values in parenthesis. P-values ≤ 0.10 are in bold type.

Table A3b. Testing determinants of sample attrition using baseline socio-demographic indicators for selected subgroups.

Dep. Variable -> Attrition	Females	Males	Less than HS	HS plus	Stunted	Non-stunted	Younger	Older	Did not outgrow BSID III
ITT	-0.029 (0.256)	-0.008 (0.756)	-0.009 (0.688)	-0.010 (0.756)	-0.021 (0.644)	-0.010 (0.627)	-0.014 (0.573)	-0.014 (0.647)	-0.007 (0.739)
Age	0.019 (0.062)	-0.019 (0.074)	-0.009 (0.355)	0.019 (0.123)	0.038 (0.073)	-0.006 (0.473)	-0.010 (0.639)	0.000 (0.994)	-0.005 (0.742)
Age Squared	0.000 (0.055)	0.000 (0.064)	0.000 (0.430)	-0.000 (0.173)	-0.001 (0.066)	0.000 (0.406)	0.000 (0.585)	-0.000 (0.984)	0.000 (0.606)
Male	- -	- -	0.022 (0.319)	0.012 (0.691)	-0.016 (0.719)	0.023 (0.241)	0.023 (0.310)	0.012 (0.676)	0.022 (0.287)
Black	-0.008 (0.770)	-0.010 (0.716)	-0.016 (0.489)	0.011 (0.709)	0.069 (0.136)	-0.021 (0.304)	0.002 (0.948)	-0.010 (0.712)	-0.002 (0.922)
Single Mom	-0.002 (0.931)	-0.030 (0.289)	-0.030 (0.215)	0.030 (0.390)	0.059 (0.239)	-0.023 (0.276)	-0.025 (0.305)	-0.002 (0.946)	-0.036 (0.067)
Mom above secondary	0.005 (0.839)	-0.004 (0.897)	- -	- -	0.014 (0.779)	-0.004 (0.848)	-0.000 (0.999)	0.007 (0.804)	0.003 (0.910)
Wealth Index	0.004 (0.130)	-0.004 (0.152)	-0.000 (0.886)	0.001 (0.728)	0.01 (0.088)	-0.001 (0.731)	0.003 (0.572)	-0.003 (0.416)	0.002 (0.540)
Children in the household	-0.007 (0.652)	0.034 (0.032)	0.011 (0.383)	0.011 (0.600)	-0.003 (0.913)	0.017 (0.188)	0.032 (0.062)	-0.009 (0.563)	0.033 (0.025)
Childcare before baseline	-0.001 (0.982)	0.022 (0.583)	-0.010 (0.789)	0.047 (0.292)	-0.142 (0.036)	0.045 (0.151)	0.024 (0.683)	0.008 (0.801)	0.005 (0.907)
Neighborhood (Alpes B)	0.156 (0.002)	0.080 (0.165)	0.075 (0.130)	0.179 (0.004)	0.17 (0.054)	0.089 (0.035)	0.119 (0.123)	0.101 (0.186)	0.153 (0.030)
Neighborhood (Timayui 1)	0.026 (0.443)	0.113 (0.001)	0.044 (0.147)	0.122 (0.003)	-0.018 (0.770)	0.084 (0.001)	0.077 (0.047)	0.055 (0.184)	0.088 (0.014)
Neighborhood (Timayui 2)	0.002 (0.943)	-0.039 (0.243)	-0.028 (0.337)	0.011 (0.791)	-0.026 (0.678)	-0.007 (0.790)	-0.018 (0.402)	-0.004 (0.892)	-0.000 (0.998)
Birth cohort 2008	-0.086 (0.135)	0.14 (0.047)	-0.006 (0.919)	0.063 (0.393)	-0.143 (0.119)	0.062 (0.227)	0.058 (0.642)	-0.002 (0.977)	0.074 (0.239)
Birth cohort 2009	-0.119 (0.138)	0.092 (0.362)	-0.046 (0.563)	0.037 (0.732)	-0.226 (0.076)	0.049 (0.509)	-0.010 (0.848)	-0.000 (0.998)	0.068 (0.220)
Birth cohort 2010	0.001 (0.990)	0.121 (0.324)	-0.010 (0.925)	0.182 (0.193)	0.136 (0.506)	0.081 (0.377)	0.043 (0.642)	- -	0.129 (0.149)
Constant	-0.013 (0.931)	0.026 (0.866)	0.150 (0.258)	-0.280 (0.145)	-0.170 (0.566)	-0.011 (0.926)	-0.003 (0.989)	0.090 (0.914)	-0.119 (0.461)
Observations	383	436	520	299	173	646	408	411	542
R-squared	0.049	0.071	0.034	0.069	0.137	0.043	0.071	0.018	0.068
F-stat	1.274	2.137	1.201	1.402	1.544	1.755	.	0.463	1.409
Prob > F	0.216	0.00785	0.267	0.145	0.0908	0.0337	.	0.251	0.132

The dependent variable is whether the observation was lost at follow-up. P-values in parenthesis. P-values ≤ 0.10 are in bold type.

Table A4. Compliance with random assignment by selected subgroups. We show the number of children that complied with the ITT by gender, mother’s education and stunting and according to ITT and enrollment. Percentages in parenthesis. That is, the number of children that enrolled with ITT=1 and that did not enroll with ITT=0. We also test for distribution differences in the compliance indicator by ITT. P-values ≤ 0.10 are in bold type.

Variables	Lottery Losers-Not enrolled	Lottery Winners-Enrolled	P-value
Females	191 (84%)	117 (75%)	0.027
Males	211 (82%)	129 (72%)	0.004
Less than High school	261 (85%)	147 (68%)	0.000
High School or higher	141 (79%)	99 (82%)	0.451
Stunted	95 (91%)	52 (76%)	0.004
Non-stunted	307 (81%)	194 (72%)	0.008
Did not outgrow BSIDIII	248 (81%)	133 (73%)	0.021
Total	402 (83%)	246 (73%)	0.000

Table A5a. Testing determinants of enrollment using randomized assignment to treatment and baseline socio-demographic indicators.

Dep. Variable -> Child enrolled in aeioTU at follow-up	(1)	(2)
ITT	0.564 (0.000)	0.557 (0.000)
Age		0.023 (0.051)
Age Squared		0.000 (0.067)
Male		-0.007 (0.793)
Black		0.009 (0.766)
Single Mom		-0.035 (0.278)
Mom above secondary		0.083 (0.005)
Wealth Index		0.005 (0.130)
Children in the household		-0.001 (0.939)
Childcare before baseline		0.031 (0.502)
Neighborhood (Alpes B)		0.002 (0.975)
Neighborhood (Timayui 1)		-0.030 (0.437)
Neighborhood (Timayui 2)		-0.033 (0.396)
Birth cohort 2008		-0.110 (0.120)
Birth cohort 2009		-0.124 (0.216)
Birth cohort 2010		0.086 (0.507)
Constant	0.166 (0.000)	0.018 (0.918)
Observations	819	819
R-squared	0.321	0.345
F-stat	387.1	26.38
Prob > F	0	0

P-values ≤ 0.10 are in bold type.

Table A5b. Testing determinants of enrollment using randomized assignment to treatment and baseline socio-demographic indicators for selected subgroups.

Dep. Variable -> Child enrolled in aeioTU at follow-up	Females	Males	Less than HS	HS plus	Stunted	Non-stunted	Younger	Older
ITT	0.581 (0.000)	0.539 (0.000)	0.535 (0.000)	0.574 (0.000)	0.654 (0.000)	0.53 (0.000)	0.557 (0.000)	0.554 (0.000)
Age	0.024 (0.162)	0.025 (0.132)	0.020 (0.172)	0.023 (0.232)	0.004 (0.876)	0.025 (0.053)	0.061 (0.011)	-0.009 (0.914)
Age Squared	-0.001 (0.143)	-0.001 (0.204)	-0.000 (0.279)	-0.001 (0.206)	-0.000 (0.675)	-0.000 (0.132)	-0.002 (0.024)	0.000 (0.888)
Male	-	-	-0.028 (0.428)	0.015 (0.745)	0.011 (0.851)	-0.008 (0.797)	-0.028 (0.478)	0.018 (0.652)
Black	0.028 (0.518)	-0.017 (0.676)	0.035 (0.360)	-0.030 (0.535)	-0.073 (0.222)	0.023 (0.509)	-0.034 (0.435)	0.048 (0.261)
Single Mom	-0.059 (0.205)	-0.017 (0.701)	-0.014 (0.721)	-0.087 (0.114)	0.015 (0.817)	-0.048 (0.191)	-0.031 (0.482)	-0.037 (0.424)
Mom above secondary	0.053 (0.212)	0.106 (0.013)	-	-	0.024 (0.704)	0.088 (0.010)	0.041 (0.321)	0.134 (0.002)
Wealth Index	0.003 (0.525)	0.007 (0.115)	0.007 (0.069)	-0.000 (0.961)	0.009 (0.221)	0.003 (0.295)	0.008 (0.039)	0.001 (0.754)
Children in the household	-0.007 (0.774)	0.002 (0.935)	-0.011 (0.603)	0.025 (0.456)	0.001 (0.984)	-0.010 (0.643)	0.019 (0.402)	-0.019 (0.458)
Childcare before baseline	-0.040 (0.552)	0.088 (0.167)	-0.008 (0.901)	0.050 (0.467)	-0.084 (0.329)	0.052 (0.328)	-0.110 (0.319)	0.057 (0.304)
Neighborhood (Alpes B)	0.024 (0.769)	-0.032 (0.725)	0.064 (0.421)	-0.094 (0.329)	0.136 (0.230)	-0.017 (0.810)	-0.023 (0.726)	0.017 (0.813)
Neighborhood (Timayui 1)	-0.067 (0.233)	0.002 (0.965)	0.030 (0.538)	-0.139 (0.029)	0.088 (0.273)	-0.055 (0.213)	-0.084 (0.095)	0.004 (0.941)
Neighborhood (Timayui 2)	0.034 (0.557)	-0.065 (0.209)	-0.006 (0.906)	-0.075 (0.263)	0.087 (0.273)	-0.063 (0.162)	-0.131 (0.009)	0.045 (0.413)
Birth cohort 2008	-0.003 (0.979)	-0.160 (0.144)	0.012 (0.893)	-0.342 (0.003)	-0.183 (0.122)	-0.058 (0.504)	-0.152 (0.269)	-0.033 (0.706)
Birth cohort 2009	-0.048 (0.716)	-0.151 (0.337)	0.022 (0.862)	-0.375 (0.024)	-0.245 (0.134)	-0.038 (0.761)	-0.37 (0.000)	-0.065 (0.587)
Birth cohort 2010	0.206 (0.261)	0.046 (0.810)	0.262 (0.109)	-0.212 (0.329)	-0.383 (0.145)	0.224 (0.151)	-0.124 (0.245)	
Constant	-0.040 (0.874)	0.006 (0.980)	-0.108 (0.612)	0.349 (0.242)	0.338 (0.375)	-0.075 (0.709)	0.092 (0.684)	0.231 (0.825)
Observations	383	436	520	299	173	646	408	411
R-squared	0.380	0.338	0.324	0.407	0.507	0.325	0.371	0.351
F-stat	14.99	14.29	16.13	12.97	10.02	18.91	.	17.67
Prob > F	0	0	0	0	0	0	.	0

P-values in parenthesis. P-values ≤ 0.10 are in bold type.

Table A5c. Testing determinants of enrollment using randomized assignment to treatment and baseline socio-demographic indicators for the sample of children who did not outgrow BSIDIII.

Dep. Variable -> Child enrolled in aeioTU at follow-up	(1)	(2)
ITT	0.543 (0.000)	0.535 (0.000)
Age		0.038 (0.053)
Age Squared		-0.001 (0.093)
Male		-0.047 (0.210)
Black		-0.054 (0.175)
Single Mom		-0.012 (0.783)
Mom above secondary		0.053 (0.187)
Wealth Index		0.009 (0.045)
Children in the household		0.012 (0.613)
Childcare before baseline		-0.031 (0.693)
Neighborhood (Alpes B)		0.018 (0.833)
Neighborhood (Timayui 1)		-0.027 (0.606)
Neighborhood (Timayui 2)		-0.064 (0.199)
Birth cohort 2009		-0.062 (0.402)
Birth cohort 2010		0.207 (0.082)
Constant	0.184 (0.000)	-0.079 (0.680)
Observations	487	487
R-squared	0.291	0.329
F-stat	198.8	15.39
Prob > F	0	0

P-values in parenthesis. P-values ≤ 0.10 are in bold type.

Table A5d. Testing determinants of enrollment using randomized assignment to treatment and baseline socio-demographic indicators on the sample of children who did not outgrow BSIDIII for selected subgroups.

Dep. Variable -> Child enrolled in aeioTU at follow-up	Females	Males	Less than HS	HS plus	Stunted	Non-stunted	Younger	Older
ITT	0.567 (0.000)	0.499 (0.000)	0.479 (0.000)	0.647 (0.000)	0.561 (0.000)	0.523 (0.000)	0.557 (0.000)	0.474 (0.000)
Age	0.026 (0.349)	0.035 (0.248)	0.037 (0.154)	0.030 (0.305)	0.022 (0.625)	0.032 (0.151)	0.061 (0.011)	0.421 (0.043)
Age Squared	-0.001 (0.496)	-0.001 (0.298)	-0.001 (0.188)	-0.001 (0.501)	-0.000 (0.724)	-0.001 (0.323)	-0.002 (0.024)	-0.008 (0.049)
Male	-	-	-0.047 (0.328)	-0.055 (0.359)	-0.044 (0.635)	-0.033 (0.433)	-0.028 (0.478)	-0.156 (0.039)
Black	-0.043 (0.452)	-0.063 (0.260)	-0.030 (0.557)	-0.062 (0.323)	-0.110 (0.246)	-0.050 (0.259)	-0.034 (0.435)	0.007 (0.935)
Single Mom	-0.055 (0.356)	0.027 (0.654)	0.027 (0.608)	-0.093 (0.191)	0.070 (0.487)	-0.030 (0.526)	-0.031 (0.482)	0.067 (0.397)
Mom above secondary	0.044 (0.436)	0.061 (0.290)			-0.143 (0.159)	0.074 (0.096)	0.041 (0.321)	0.135 (0.118)
Wealth Index	0.007 (0.255)	0.012 (0.048)	0.016 (0.005)	-0.003 (0.644)	0.014 (0.254)	0.008 (0.091)	0.008 (0.039)	0.015 (0.074)
Children in the household	-0.057 (0.144)	0.057 (0.076)	0.017 (0.574)	0.019 (0.643)	-0.041 (0.493)	0.015 (0.582)	0.019 (0.402)	-0.075 (0.147)
Childcare before baseline	-0.120 (0.338)	-0.010 (0.926)	-0.213 (0.050)	0.258 (0.025)	-0.323 (0.079)	0.044 (0.627)	-0.110 (0.319)	0.148 (0.233)
Neighborhood (Alpes B)	0.062 (0.592)	-0.018 (0.888)	0.086 (0.471)	-0.050 (0.681)	0.269 (0.225)	0.008 (0.934)	-0.023 (0.726)	0.145 (0.358)
Neighborhood (Timayui 1)	-0.046 (0.540)	-0.027 (0.722)	-0.032 (0.615)	-0.009 (0.923)	0.040 (0.754)	-0.023 (0.692)	-0.084 (0.095)	0.106 (0.374)
Neighborhood (Timayui 2)	0.022 (0.769)	-0.136 (0.048)	-0.070 (0.265)	-0.059 (0.465)	0.225 (0.081)	-0.104 (0.069)	-0.131 (0.009)	0.164 (0.089)
Birth cohort 2008	-	-	-	-	-	-	-0.152 (0.269)	0.133 (0.256)
Birth cohort 2009	-0.107 (0.257)	-0.032 (0.803)	-0.052 (0.556)	-0.039 (0.778)	-0.016 (0.909)	-0.302 (0.001)	-0.37 (0.000)	0.131 (0.385)
Birth cohort 2010	0.215 (0.206)	0.158 (0.387)	0.222 (0.139)	0.195 (0.331)	-0.045 (0.881)	-0.287 (0.033)	-0.124 (0.245)	-
Constant	0.164 (0.597)	-0.176 (0.493)	-0.075 (0.763)	-0.067 (0.827)	0.121 (0.806)	0.159 (0.293)	0.092 (0.684)	0.231 (0.825)
Observations	225	262	314	173	88	399	408	411
R-squared	0.395	0.306	0.299	0.452	0.459	0.331	0.371	0.351
F-stat	9.782	7.795	9.127	9.293	4.077	12.64	.	17.67
Prob > F	0	0	0	0	2.55e-05	0	.	0

P-values in parenthesis. P-values ≤ 0.10 are in bold type.

Table A6. The treatment in context: type of childcare used at follow up by study group.

Child care attendance at follow-up by study group (%)	Control	Treatment
<i>Enrolled in Early Childhood Services</i>	36.8%	82.2%
Publicly provided alternative service (such as home day care)	41.2%	6.9%
Center-based care provided by NGO	5.5%	0.8%
Private center-based care	4.2%	0.8%
AeioTu center	47.9%	91.5%
Public elementary school with preschool	0.6%	0.0%
<i>Non Enrolled: Care by</i>	60.0%	16.8%
Mother	86.2%	92.3%
Father	1.9%	1.9%
Other relative	9.3%	5.7%
Non-relative	0.7%	0.0%
<i>Did not report</i>	3.2%	1.0%

Table A7a. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for females. See notes in Table 2 (paper).

Variable	N	Control mean	Treatment mean	Difference in means	ITT Conditional effect	ITT P value‡	ITT Stepdown P value§	TOT Conditional effect	TOT P value‡	TOT Stepdown P value§
Nutrition (Z scores)										
Length/Height-for-age	346	-1.08	-1.13	-0.05	0.013	0.821	0.999	0.023	0.815	0.998
BMI-for-age	340	0.33	0.35	0.02	-0.031	0.654	0.993	-0.056	0.643	0.990
Weight-for-age	348	-0.43	-0.42	0.00	0.007	0.899	0.999	0.012	0.896	0.998
Weight-for-length	342	0.20	0.26	0.06	-0.002	0.979	0.999	-0.003	0.978	0.998
Arm circumference	341	-0.34	-0.30	0.04	-0.037	0.592	0.993	-0.064	0.581	0.990
Language, Cognitive and Motor Development: Bayley Scales of Infant Development III (BSID III) - Raw Scores										
Receptive vocabulary¶	332	0.02	0.28	0.27	0.273	0.013	0.026	0.448	0.011	0.020
TVIP	108	8.98	11.40	2.43	2.554	0.030	0.060	3.546	0.023	0.051
Receptive vocabulary	225	24.89	26.02	1.13	1.129	0.032	0.130	1.979	0.028	0.128
Expressive vocabulary	226	24.90	26.14	1.23	1.566	0.015	0.078	2.886	0.013	0.073
Language total	223	49.89	52.34	2.45	2.723	0.009	0.013	4.802	0.007	0.014
Cognitive	225	58.60	59.23	0.63	1.025	0.114	0.292	1.859	0.102	0.253
Fine motor	225	38.98	40.10	1.13	1.119	0.003	0.012	2.039	0.004	0.014
Gross motor	223	54.77	55.77	1.00	0.950	0.079	0.213	1.720	0.068	0.175
Motor total	222	93.74	95.81	2.08	2.059	0.008	0.014	3.756	0.007	0.018
BSID III Total Raw	213	202.2	207.09	4.89	4.906	0.015	0.025	8.666	0.012	0.018
Socio-Emotional Development: Ages and Stages Socio-emotional (ASQ:SE)										
Self-regulation	355	17.28	21.42	4.14	2.700	0.102	0.731	4.867	0.092	0.676
Compliance	273	2.62	3.65	1.03	0.401	0.486	0.997	0.662	0.466	0.996
Communication	355	2.25	2.33	0.08	-0.249	0.690	0.999	-0.448	0.679	0.999
Adaptive functioning	354	5.00	6.76	1.76	1.471	0.097	0.731	2.635	0.084	0.672
Autonomy	273	8.33	8.92	0.59	-0.073	0.889	0.999	-0.121	0.884	0.999
Affect	355	2.24	2.69	0.45	0.142	0.810	0.999	0.255	0.803	0.999
Interaction	354	6.82	7.12	0.30	0.026	0.979	0.999	0.047	0.978	0.999
ASQ:SE Total	353	42.76	51.97	9.21	5.839	0.106	0.192	10.553	0.096	0.180
Home Observation and Measurement of the Environment (HOME)										
Responsivity	341	7.69	7.83	0.14	0.104	0.697	0.996	0.183	0.686	0.993
Acceptance	341	6.42	6.45	0.03	0.084	0.432	0.996	0.147	0.416	0.993
Organization	333	4.65	4.61	-0.04	0.006	0.962	0.996	0.010	0.960	0.993
Learning Materials	333	3.64	3.43	-0.21	-0.146	0.477	0.996	-0.256	0.463	0.993
Involvement	333	3.20	3.00	-0.20	-0.090	0.608	0.996	-0.158	0.592	0.993
Variety	341	2.18	2.19	0.01	-0.004	0.977	0.996	-0.006	0.976	0.999
HOME Total	341	27.91	27.67	-0.24	-0.075	0.893	0.996	-0.131	0.889	0.965

Table A8. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for males. See notes in Table 2 (paper).

Variable	N	Control mean	Treatment mean	Difference in means	ITT Conditional effect	ITT P value‡	ITT Stepdown P value§	TOT Conditional effect	TOT P value‡	TOT Stepdown P value§
Nutrition (Z scores)										
Length/Height-for-age	392	-1.21	-1.11	0.10	0.090	0.193	0.770	0.161	0.186	0.755
BMI-for-age	384	0.37	0.18	-0.18	-0.038	0.598	0.993	-0.069	0.586	0.990
Weight-for-age	390	-0.49	-0.59	-0.10	0.044	0.522	0.989	0.078	0.513	0.990
Weight-for-length	386	0.19	0.06	-0.14	0.012	0.858	0.999	0.022	0.854	0.998
Arm circumference	385	-0.40	-0.48	-0.08	-0.009	0.888	0.999	-0.016	0.885	0.998
Language, Cognitive and Motor Development: Bayley Scales of Infant Development III (BSID III) - Raw Scores										
Receptive vocabulary¶	378	-0.06	-0.11	-0.05	-0.086	0.412	0.418	-0.159	0.398	0.406
TVIP	122	15.11	12.8	-2.31	-2.379	0.097	0.091	-3.939	0.071	0.064
Receptive vocabulary	255	22.89	23.32	0.43	0.495	0.271	0.440	0.998	0.254	0.411
Expressive vocabulary	256	23.75	23.82	0.07	0.217	0.691	0.680	0.415	0.678	0.662
Language total	250	46.51	46.91	0.4	0.512	0.551	0.545	1.001	0.533	0.519
Cognitive	262	57.05	57.96	0.91	1.098	0.069	0.225	2.186	0.058	0.198
Fine motor	258	37.84	37.79	-0.05	0.150	0.683	0.718	0.304	0.672	0.695
Gross motor	259	54.23	54.29	0.06	0.321	0.493	0.718	0.624	0.477	0.695
Motor total	256	92.12	91.95	-0.17	0.298	0.673	0.663	0.583	0.661	0.649
BSID III Total	243	195.47	196.51	1.04	1.532	0.400	0.388	2.954	0.379	0.359
Socio-Emotional Development: Ages and Stages Socio-emotional (ASQ:SE)										
Self-regulation	393	20.40	20.96	0.56	-0.244	0.871	0.999	-0.447	0.867	0.999
Compliance	286	3.15	3.84	0.69	0.136	0.798	0.999	0.228	0.789	0.999
Communication	393	1.95	2.29	0.33	0.141	0.751	0.999	0.259	0.744	0.999
Adaptive functioning	393	6.31	5.85	-0.46	-0.806	0.321	0.980	-1.477	0.307	0.974
Autonomy	286	7.91	8.55	0.64	0.299	0.615	0.999	0.503	0.600	0.999
Affect	393	2.51	2.74	0.23	-0.102	0.827	0.999	-0.186	0.821	0.999
Interaction	393	7.12	6.27	-0.85	-1.288	0.152	0.999	-2.351	0.142	0.801
ASQ:SE Total	393	47.65	48.75	1.10	-1.883	0.522	0.999	-3.443	0.509	0.494
Home Observation and Measurement of the Environment (HOME)										
Responsivity	379	7.82	7.96	0.15	0.193	0.430	0.996	0.344	0.416	0.993
Acceptance	379	6.37	6.39	0.02	0.080	0.424	0.996	0.143	0.412	0.993
Organization	369	4.48	4.64	0.16	0.159	0.150	0.814	0.285	0.136	0.774
Learning Materials	369	3.59	3.62	0.03	0.109	0.547	0.996	0.194	0.533	0.993
Involvement	369	3.24	3.10	-0.14	0.000	0.999	0.999	0.000	0.999	0.999
Variety	379	2.22	2.07	-0.16	-0.229	0.044	0.410	-0.408	0.038	0.363
HOME Total	379	27.89	27.88	0.00	0.122	0.817	0.996	0.218	0.811	0.965

Table A9. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for children of mothers with less than high school attainment. See notes in Table 2 (paper).

Variable	N	Control mean	Treatment mean	Difference in means	ITT Conditional effect	ITT P value‡	ITT Stepdown P value§	TOT Conditional effect	TOT P value‡	TOT Stepdown P value§
Nutrition (Z scores)										
Length/Height-for-age	466	-1.26	-1.23	0.03	0.064	0.284	0.855	0.116	0.276	0.843
BMI-for-age	458	0.39	0.18	-0.20	-0.094	0.128	0.597	-0.172	0.118	0.561
Weight-for-age	469	-0.52	-0.63	-0.12	0.002	0.970	0.994	0.004	0.969	0.994
Weight-for-length	460	0.22	0.08	-0.14	-0.041	0.487	0.946	-0.075	0.475	0.936
Arm circumference	465	-0.39	-0.45	-0.06	-0.028	0.622	0.976	-0.049	0.615	0.974
Language, Cognitive and Motor Development: Bayley Scales of Infant Development III (BSID III) - Raw Scores										
Receptive vocabulary¶	450	-0.06	-0.05	0.01	-0.005	0.960	0.959	-0.008	0.959	0.959
TVIP	140	11.55	11.21	-0.34	-0.032	0.977	0.977	-0.046	0.976	0.967
Receptive vocabulary	310	24.14	24.37	0.23	0.276	0.519	0.524	0.551	0.505	0.505
Expressive vocabulary	312	24.3	24.6	0.30	0.608	0.247	0.418	1.230	0.227	0.384
Language total	308	48.39	48.96	0.57	0.895	0.280	0.272	1.768	0.261	0.251
Cognitive	314	58.05	58.6	0.55	0.732	0.191	0.418	1.521	0.177	0.384
Fine motor	311	38.62	39.02	0.40	0.436	0.207	0.469	0.917	0.200	0.432
Gross motor	310	54.69	55.17	0.49	0.577	0.197	0.469	1.180	0.184	0.432
Motor total	307	93.34	94.06	0.72	0.873	0.197	0.199	1.804	0.186	0.186
BSID III Total	297	199.47	201.5	2.03	1.681	0.332	0.325	3.350	0.313	0.316
Socio-Emotional Development: Ages and Stages Socio-emotional (ASQ:SE)										
Self-regulation	473	18.76	21.72	2.96	1.580	0.264	0.971	2.940	0.252	0.961
Compliance	356	2.87	3.82	0.95	0.492	0.324	0.983	0.823	0.307	0.975
Communication	473	1.86	2.19	0.32	0.248	0.575	0.999	0.461	0.565	0.999
Adaptive functioning	473	5.40	5.75	0.35	0.149	0.830	0.999	0.276	0.826	0.999
Autonomy	356	8.04	8.91	0.88	0.247	0.614	0.999	0.413	0.603	0.999
Affect	473	2.53	2.66	0.14	0.167	0.725	0.999	0.311	0.718	0.999
Interaction	473	7.19	7.10	-0.09	-0.099	0.905	0.999	-0.184	0.902	0.999
ASQ:SE Total	473	44.89	50.85	5.96	3.316	0.238	0.425	6.173	0.227	0.402
Home Observation and Measurement of the Environment (HOME)										
Responsivity	456	7.52	7.58	0.06	-0.057	0.800	0.990	-0.103	0.795	0.977
Acceptance	456	6.36	6.38	0.03	0.069	0.448	0.990	0.123	0.437	0.977
Organization	441	4.36	4.49	0.13	0.092	0.378	0.990	0.166	0.364	0.977
Learning Materials	441	3.02	3.18	0.16	0.185	0.254	0.948	0.331	0.238	0.926
Involvement	441	2.97	2.79	-0.18	-0.099	0.474	0.990	-0.177	0.458	0.977
Variety	456	2.14	2.05	-0.09	-0.083	0.428	0.990	-0.150	0.416	0.977
HOME Total	456	26.61	26.61	0.00	-0.056	0.905	0.944	-0.101	0.902	0.936

Table A10. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for children of mothers with high school attainment or above. See notes in Table 2 (paper).

Variable	N	Control mean	Treatment mean	Difference in means	ITT Conditional effect	ITT P value‡	ITT Stepdown P value§	TOT Conditional effect	TOT P value‡	TOT Stepdown P value§
Nutrition (Z scores)										
Length/Height-for-age	272	-0.97	-0.93	0.04	0.028	0.689	0.976	0.048	0.677	0.974
BMI-for-age	266	0.29	0.41	0.12	0.086	0.322	0.861	0.150	0.305	0.848
Weight-for-age	269	-0.36	-0.29	0.07	0.083	0.226	0.801	0.142	0.211	0.773
Weight-for-length	268	0.16	0.29	0.13	0.107	0.190	0.744	0.188	0.176	0.700
Arm circumference	261	-0.32	-0.29	0.04	-0.008	0.920	0.994	-0.014	0.916	0.994
Language, Cognitive and Motor Development: Bayley Scales of Infant Development III (BSID III) - Raw Scores										
Receptive vocabulary¶	260	0.03	0.30	0.27	0.279	0.032	0.064	0.480	0.028	0.057
TVIP	90	13.04	13.77	0.73	0.297	0.851	0.977	0.590	0.834	0.965
Receptive vocabulary	170	23.27	24.93	1.66	1.808	0.002	0.010	2.869	0.002	0.009
Expressive vocabulary	170	24.26	25.46	1.20	1.561	0.031	0.110	2.390	0.025	0.090
Language total	165	47.62	50.31	2.69	3.311	0.005	0.010	5.177	0.004	0.008
Cognitive	173	57.24	58.54	1.30	1.888	0.012	0.056	2.937	0.008	0.041
Fine motor	172	37.92	38.56	0.63	0.779	0.057	0.190	1.203	0.047	0.168
Gross motor	172	54.12	54.60	0.48	0.615	0.296	0.469	0.925	0.266	0.432
Motor total	171	92.05	93.16	1.11	1.406	0.084	0.154	2.129	0.069	0.125
BSID III Total	159	197.02	201.29	4.26	6.416	0.005	0.008	9.769	0.004	0.007
Socio-Emotional Development: Ages and Stages Socio-emotional (ASQ:SE)										
Self-regulation	275	19.16	20.21	1.05	-0.211	0.908	0.999	-0.361	0.903	0.999
Compliance	203	2.93	3.63	0.70	0.474	0.459	0.999	0.800	0.427	0.999
Communication	275	2.48	2.52	0.04	-0.644	0.359	0.999	-1.091	0.338	0.980
Adaptive functioning	274	6.16	7.23	1.07	0.442	0.701	0.999	0.749	0.687	0.999
Autonomy	203	8.25	8.38	0.12	-0.300	0.661	0.999	-0.509	0.64	0.999
Affect	275	2.13	2.82	0.68	0.334	0.586	0.999	0.564	0.569	0.999
Interaction	274	6.63	5.90	-0.72	-1.849	0.094	0.705	-3.143	0.081	0.633
ASQ:SE Total	273	46.05	49.20	3.14	-0.875	0.832	0.835	-1.505	0.824	0.820
Home Observation and Measurement of the Environment (HOME)										
Responsivity	264	8.16	8.45	0.29	0.261	0.375	0.990	0.432	0.353	0.977
Acceptance	264	6.45	6.46	0.02	0.092	0.463	0.990	0.153	0.444	0.977
Organization	261	4.90	4.86	-0.04	-0.013	0.912	0.990	-0.022	0.907	0.977
Learning Materials	261	4.61	4.13	-0.48	-0.338	0.158	0.853	-0.553	0.141	0.801
Involvement	261	3.63	3.50	-0.13	-0.111	0.579	0.990	-0.181	0.559	0.977
Variety	264	2.32	2.25	-0.06	-0.129	0.357	0.990	-0.214	0.333	0.977
HOME Total	264	30.12	29.81	-0.31	-0.196	0.761	0.944	-0.324	0.749	0.936

Table A11. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for children stunted at baseline. See notes in Table 2 (paper).

Variable	N	Control mean	Treatment mean	Difference in means	ITT Conditional effect	ITT P value‡	ITT Stepdown P value§	TOT Conditional effect	TOT P value‡	TOT Stepdown P value§
Nutrition (Z scores)										
Length/Height-for-age	156	-2.16	-2.23	-0.06	-0.090	0.438	0.926	-0.129	0.407	0.902
BMI-for-age	152	0.36	0.11	-0.24	-0.091	0.482	0.930	-0.133	0.445	0.902
Weight-for-age	154	-1.12	-1.30	-0.18	-0.067	0.567	0.949	-0.096	0.536	0.929
Weight-for-length	152	0.06	-0.03	-0.08	0.049	0.685	0.965	0.071	0.662	0.958
Arm circumference	153	-0.69	-0.80	-0.12	-0.130	0.284	0.827	-0.185	0.243	0.765
Language, Cognitive and Motor Development: Bayley Scales of Infant Development III (BSID III) - Raw Scores										
Receptive vocabulary¶	142	-0.29	-0.03	0.26	0.200	0.219	0.386	0.274	0.178	0.313
TVIP	55	10.49	11.92	1.43	0.964	0.644	0.852	1.145	0.562	0.735
Receptive vocabulary	88	23.89	25.2	1.30	0.690	0.499	0.698	1.200	0.430	0.587
Expressive vocabulary	89	24.72	25.08	0.36	-0.569	0.600	0.698	-0.938	0.557	0.587
Language total	87	48.5	50.11	1.61	-0.257	0.893	0.896	-0.435	0.878	0.872
Cognitive	88	57.41	60.07	2.66	1.899	0.186	0.416	3.204	0.115	0.317
Fine motor	89	38.81	39.54	0.73	-0.203	0.799	0.796	-0.352	0.772	0.755
Gross motor	88	54.24	56.13	1.89	0.966	0.349	0.675	1.566	0.276	0.581
Motor total	88	93.2	95.13	1.93	-0.419	0.789	0.786	-0.693	0.761	0.737
BSID III Total	83	198.6	204.82	6.22	0.146	0.975	0.977	0.247	0.971	0.967
Socio-Emotional Development: Ages and Stages Socio-emotional (ASQ:SE)										
Self-regulation	154	20.43	20.4	-0.03	0.434	0.873	0.999	0.633	0.860	0.998
Compliance	130	2.40	4.00	1.60	0.851	0.319	0.986	1.190	0.263	0.952
Communication	154	1.96	2.18	0.22	0.753	0.377	0.990	1.100	0.342	0.984
Adaptive functioning	154	5.49	6.13	0.64	0.308	0.825	0.999	0.449	0.808	0.991
Autonomy	130	8.87	9.09	0.22	0.677	0.452	0.996	0.957	0.403	0.991
Affect	154	2.12	2.02	-0.10	-0.007	0.994	0.999	-0.010	0.994	0.999
Interaction	154	7.50	6.61	-0.89	-0.590	0.723	0.996	-0.844	0.697	0.991
ASQ:SE Total	154	47.95	49.96	2.01	4.018	0.479	0.660	5.803	0.440	0.639
Home Observation and Measurement of the Environment (HOME)										
Responsivity	150	7.98	7.90	-0.08	-0.334	0.421	0.966	-0.490	0.383	0.961
Acceptance	150	6.47	6.36	-0.11	-0.218	0.163	0.805	-0.321	0.131	0.730
Organization	147	4.28	4.39	0.11	0.132	0.476	0.966	0.188	0.433	0.961
Learning Materials	147	3.11	3.07	-0.05	0.085	0.785	0.968	0.120	0.763	0.961
Involvement	147	2.92	2.61	-0.31	-0.135	0.593	0.968	-0.191	0.559	0.961
Variety	150	2.22	2.11	-0.11	-0.282	0.089	0.621	-0.415	0.066	0.499
HOME Total	150	27.1	26.54	-0.56	-0.716	0.407	0.629	-1.053	0.372	0.580

Table A12. Intent-to-treat and Treatment-on-Treated estimations of program effects by outcome for initially non-stunted children. See notes in Table 2 (paper).

Variable	N	Control mean	Treatment mean	Difference in means	ITT Conditional effect	ITT P value‡	ITT Stepdown P value§	TOT Conditional effect	TOT P value‡	TOT Stepdown P value§
Nutrition (Z scores)										
Length/Height-for-age	582	-0.87	-0.83	0.04	0.088	0.074	0.409	0.163	0.072	0.397
BMI-for-age	572	0.35	0.30	-0.05	-0.016	0.760	0.965	-0.031	0.755	0.958
Weight-for-age	584	-0.28	-0.31	-0.03	0.056	0.238	0.781	0.103	0.232	0.765
Weight-for-length	576	0.23	0.20	-0.04	-0.008	0.869	0.979	-0.016	0.866	0.981
Arm circumference	573	-0.28	-0.29	-0.01	0.004	0.929	0.979	0.008	0.928	0.981
Language, Cognitive and Motor Development: Bayley Scales of Infant Development III (BSID III) - Raw Scores										
Receptive vocabulary¶	568	0.04	0.10	0.06	0.054	0.527	0.527	0.099	0.519	0.515
TVIP	175	12.49	12.16	-0.33	-0.145	0.892	0.876	-0.250	0.885	0.884
Receptive vocabulary	392	23.84	24.4	0.57	0.729	0.046	0.185	1.364	0.043	0.176
Expressive vocabulary	393	24.21	24.82	0.61	1.022	0.025	0.127	1.924	0.022	0.119
Language total	386	48.07	49.21	1.15	1.785	0.012	0.027	3.322	0.011	0.029
Cognitive	399	57.88	58.21	0.32	0.672	0.142	0.416	1.281	0.133	0.317
Fine motor	394	38.28	38.7	0.42	0.522	0.059	0.200	1.004	0.058	0.193
Gross motor	394	54.55	54.73	0.18	0.370	0.323	0.675	0.694	0.312	0.581
Motor total	390	92.83	93.42	0.60	0.886	0.101	0.183	1.683	0.097	0.176
BSID III Total	373	198.76	200.55	1.79	2.812	0.045	0.082	5.250	0.041	0.082
Socio-Emotional Development: Ages and Stages Socio-emotional (ASQ:SE)										
Self-regulation	594	18.5	21.37	2.87	1.598	0.186	0.914	3.046	0.177	0.894
Compliance	429	3.04	3.67	0.64	-0.014	0.975	0.999	-0.025	0.974	0.999
Communication	594	2.13	2.34	0.21	-0.096	0.819	0.996	-0.182	0.815	0.998
Adaptive functioning	593	5.74	6.32	0.58	0.217	0.746	0.996	0.410	0.740	0.991
Autonomy	429	7.90	8.62	0.72	-0.114	0.793	0.996	-0.198	0.786	0.991
Affect	594	2.45	2.89	0.44	0.176	0.663	0.996	0.333	0.655	0.991
Interaction	593	6.84	6.68	-0.15	-0.672	0.342	0.990	-1.271	0.332	0.984
ASQ:SE Total	592	44.63	50.34	5.71	2.025	0.421	0.660	3.871	0.411	0.639
Home Observation and Measurement of the Environment (HOME)										
Responsivity	570	7.70	7.90	0.21	0.258	0.194	0.833	0.476	0.185	0.804
Acceptance	570	6.37	6.43	0.06	0.116	0.154	0.805	0.214	0.148	0.754
Organization	555	4.64	4.69	0.05	0.081	0.354	0.966	0.150	0.344	0.961
Learning Materials	555	3.75	3.65	-0.10	-0.021	0.887	0.968	-0.040	0.884	0.961
Involvement	555	3.30	3.16	-0.14	-0.054	0.679	0.968	-0.099	0.671	0.961
Variety	570	2.20	2.13	-0.07	-0.084	0.383	0.966	-0.156	0.372	0.961
HOME Total	570	28.11	28.10	-0.01	0.171	0.685	0.680	0.314	0.679	0.673

Table A13. ITT and TOT estimations of program effects by outcome without Tester Fixed Effects

Variable	N	Post-intervention Means			ITT Estimated Effects				TOT Estimated Effects			
		Control mean*	Treatment mean*	Difference in means	ITT β (95% CI)	D \dagger	P value‡	Stepdown P value§	TOT β (95% CI)	D \dagger	P value‡	Stepdown P value§
Nutrition (Z scores)												
Length/Height-for-age	738	-1.15	-1.12	0.03	0.055 (-0.034 to 0.144)	0.047	0.227	0.593	0.097 (-0.059 to 0.252)	0.082	0.223	0.587
BMI-for-age	724	0.35	0.26	-0.09	-0.036 (-0.134 to 0.061)	-0.037	0.464	0.846	-0.065 (-0.236 to 0.106)	-0.066	0.458	0.837
Weight-for-age	738	-0.46	-0.51	-0.05	0.026 (-0.061 to 0.112)	0.024	0.559	0.891	0.045 (-0.106 to 0.196)	0.042	0.556	0.881
Weight-for-length	728	0.19	0.15	-0.04	0.004 (-0.088 to 0.096)	0.004	0.931	0.937	0.007 (-0.156 to 0.171)	0.007	0.930	0.937
Arm circumference	726	-0.37	-0.4	-0.03	-0.020 (-0.113 to 0.073)	-0.024	0.674	0.891	-0.034 (-0.192 to 0.124)	-0.041	0.671	0.888
Language, Cognitive and Motor Development: Bayley Scales of Infant Development III (BSID III)												
BSID receptive vocabulary	480	23.85	24.55	0.70	0.913 (0.243 to 1.583)	0.12	0.008	0.023	1.677 (0.460 to 2.895)	0.220	0.007	0.023
BSID expressive vocabulary	482	24.30	24.88	0.58	1.054 (0.243 to 1.865)	0.120	0.011	0.023	1.936 (0.475 to 3.397)	0.220	0.010	0.023
BSID language total	473	48.14	49.39	1.25	1.911 (0.613 to 3.209)	0.119	0.004	n/a	3.479 (1.156 to 5.803)	0.216	0.003	n/a
BSID cognitive	487	57.77	58.54	0.77	1.224 (0.336 to 2.112)	0.084	0.007	0.023	2.291 (0.660 to 3.923)	0.157	0.006	0.023
BSID fine motor	483	38.37	38.86	0.49	0.683 (0.122 to 1.244)	0.073	0.017	0.03	1.291 (0.231 to 2.351)	0.138	0.017	0.030
BSID gross motor	482	54.47	54.98	0.51	0.709 (0.010 to 1.407)	0.055	0.047	0.049	1.300 (0.038 to 2.562)	0.100	0.044	0.047
BSID motor total	478	92.87	93.75	0.88	1.266 (0.186 to 2.347)	0.058	0.022	n/a	2.352 (0.361 to 4.344)	0.107	0.021	n/a
BSID III Total	456	198.65	201.36	2.71	3.601 (0.829 to 6.373)	0.07	0.011	n/a	6.584 (1.616 to 11.553)	0.129	0.01	n/a
TVIP	230	11.94	12.08	0.15	0.189 (-1.574 to 1.952)	0.026	0.833	n/a	0.179 (-0.077 to 0.435)	0.180	0.171	n/a
Receptive vocabulary (Std) #	710	-0.03	0.07	0.10	0.103 (-0.047 to 0.253)	0.104	0.176	n/a	0.294 (-2.360 to 2.949)	0.041	0.827	n/a
Socio-Emotional Development: Ages and Stages Socio-emotional (ASQ:SE)												
Self-regulation	748	18.91	21.18	2.27	1.419 (-0.769 to 3.607)	0.096	0.203	0.714	2.528 (-1.326 to 6.383)	0.171	0.198	0.697
Compliance	559	2.89	3.75	0.86	0.566 (-0.189 to 1.320)	0.138	0.141	0.623	0.943 (-0.295 to 2.182)	0.23	0.135	0.600
Communication	748	2.09	2.31	0.21	0.157 (-0.591 to 0.904)	0.040	0.680	0.962	0.279 (-1.035 to 1.593)	0.071	0.677	0.958

Adaptive functioning	747	5.68	6.28	0.59	0.337 (-0.835 to 1.508)	0.039	0.573	0.962	0.599 (-1.457 to 2.654)	0.070	0.568	0.958
Autonomy	559	8.12	8.73	0.61	0.180 (-0.581 to 0.941)	0.034	0.642	0.962	0.300 (-0.947 to 1.547)	0.057	0.637	0.958
Affect	748	2.38	2.72	0.34	0.289 (-0.452 to 1.029)	0.061	0.445	0.925	0.513 (-0.789 to 1.814)	0.108	0.44	0.925
Interaction	747	6.98	6.67	-0.31	-0.372 (-1.658 to 0.915)	-0.053	0.571	0.962	-0.659 (-2.915 to 1.596)	-0.093	0.566	0.958
ASQ:SE Total	746	45.33	50.26	4.94	2.909 (-1.674 to 7.493)	0.111	0.213	n/a	5.195 (-2.900 to 13.290)	0.198	0.208	n/a
Home Observation and Measurement of the Environment (HOME)												
Responsivity	720	7.76	7.9	0.15	0.154 (-0.192 to 0.500)	0.065	0.384	0.850	0.269 (-0.330 to 0.867)	0.114	0.379	0.838
Acceptance	720	6.39	6.41	0.02	0.054 (-0.084 to 0.192)	0.058	0.440	0.850	0.095 (-0.144 to 0.334)	0.102	0.435	0.838
Organization	702	4.56	4.63	0.07	0.079 (-0.080 to 0.238)	0.074	0.329	0.850	0.138 (-0.136 to 0.411)	0.128	0.323	0.838
Learning Materials	702	3.61	3.53	-0.08	0.038 (-0.219 to 0.294)	0.019	0.773	0.909	0.066 (-0.375 to 0.506)	0.034	0.77	0.905
Involvement	702	3.22	3.05	-0.17	-0.041 (-0.262 to 0.180)	-0.027	0.718	0.909	-0.071 (-0.449 to 0.308)	-0.048	0.714	0.905
Variety	720	2.20	2.12	-0.08	-0.092 (-0.253 to 0.070)	-0.080	0.265	0.827	-0.160 (-0.440 to 0.119)	-0.141	0.26	0.815
HOME Total	720	27.90	27.78	-0.11	0.092 (-0.639 to 0.824)	0.017	0.804	n/a	0.162 (-1.102 to 1.425)	0.030	0.802	n/a

Individual lines present the results of separate regressions. Sample sizes vary by instrument due to the collection method (with less households assessed with HOME than children nutritionally) and as children at post testing grow out of the BSID III. In particular, we show BSID III results for children still eligible for BSID at follow-up (456<N<487), as well as TVIP language results for children who outgrew the BSID III at follow-up (N=230). For each estimation we present: the means of the outcome by group (treatment and control) and the difference in means, and the results of the estimations for intent-to-treat and treatment-on-treated. β s are estimated controlling for the corresponding pre-test (except for HOME for which there was no pretest), age, age squared, male, black, mother marital status (single), maternal years of education, household wealth index, household kids, childcare experience before randomization, neighborhood and cohort of birth. P-values ≤ 0.10 are in bold type.

*Means are age adjusted using ANOVAs.

†Effects sizes, also known as Cohen’s D, are β s interpreted as fraction of SD in control group at baseline (except for HOME for which there was no baseline testing).

‡Standard P-values.

§Stepdown P-values are for Romano and Wolf (2005) Stepdown procedures applied to blocks of outcomes per type of developmental dimension measured. For Bayley, motor scales are one block and cognitive and language are another block. Combined receptive vocabulary excluded from the block as it combines two measures of receptive vocabulary and is based on a different sample than other outcomes within that group. Not calculated for the Total’s aggregate scores as these are aggregate measures across various dimensions or for one developmental domain.

We internally age-standardized both, BSID III receptive vocabulary and TVIP scores in the complete sample and then pooled both in a single regression controlling for measure type.

Table A14. ITT and TOT estimations of program effects on nutrition, ASQ:SE and HOME on the subsample of children who did not outgrow BSID III

Variable	N	Post-intervention Means			ITT Estimated Effects				TOT Estimated Effects			
		Control mean*	Treatment mean*	Difference in means	ITT β (95% CI)	D \dagger	P value \ddagger	Stepdown P value \S	TOT β (95% CI)	D \dagger	P value \ddagger	Stepdown P value \S
Nutrition (Z scores)												
Length/Height-for-age	488	-1.15	-1.09	0.06	0.076 (-0.050 to 0.203)	0.062	0.237	0.6	0.143 (-0.090 to 0.376)	0.116	0.229	0.583
BMI-for-age	478	0.42	0.4	-0.02	-0.039 (-0.169 to 0.091)	-0.040	0.554	0.743	-0.075 (-0.317 to 0.167)	-0.077	0.544	0.728
Weight-for-age	491	-0.40	-0.39	0.01	0.04 (-0.078 to 0.158)	0.037	0.505	0.743	0.075 (-0.142 to 0.291)	0.069	0.498	0.728
Weight-for-length	482	0.24	0.25	0.01	-0.007 (-0.133 to 0.119)	-0.007	0.911	0.921	-0.014 (-0.248 to 0.221)	-0.014	0.909	0.919
Arm circumference	477	-0.32	-0.34	-0.02	-0.066 (-0.185 to 0.054)	-0.079	0.281	0.600	-0.118 (-0.330 to 0.093)	-0.142	0.272	0.583
Socio-Emotional Development: Ages and Stages Socio-emotional (ASQ:SE)												
Self-regulation	497	18.25	19.57	1.32	1.825 (-0.775 to 4.425)	0.145	0.168	0.638	3.427 (-1.303 to 8.157)	0.271	0.155	0.602
Compliance	309	2.31	2.23	-0.08	-0.27 (-1.106 to 0.566)	-0.095	0.525	0.903	-0.451 (-1.795 to 0.893)	-0.158	0.509	0.892
Communication	497	1.70	1.52	-0.18	-0.307 (-1.102 to 0.488)	-0.083	0.448	0.903	-0.577 (-2.034 to 0.881)	-0.157	0.437	0.892
Adaptive functioning	496	4.94	5.38	0.44	0.337 (-1.054 to 1.729)	0.037	0.634	0.903	0.634 (-1.906 to 3.175)	0.07	0.624	0.892
Autonomy	309	6.98	8.18	1.20	0.861 (-0.072 to 1.794)	0.257	0.070	0.383	1.446 (-0.083 to 2.976)	0.432	0.064	0.342
Affect	497	2.56	2.82	0.26	0.163 (-0.640 to 0.966)	0.033	0.690	0.903	0.307 (-1.161 to 1.774)	0.062	0.682	0.892
Interaction	496	6.71	6	-0.71	-0.624 (-2.036 to 0.788)	-0.092	0.386	0.903	-1.179 (-3.782 to 1.425)	-0.174	0.374	0.892
ASQ:SE Total	495	41.36	43.13	1.77	2.290 (-2.539 to 7.118)	0.096	0.352	n/a	4.317 (-4.510 to 13.145)	0.181	0.337	n/a
Home Observation and Measurement of the Environment (HOME)												
Responsivity	478	7.71	7.89	0.17	0.181 (-0.248 to 0.610)	0.077	0.407	0.813	0.327 (-0.429 to 1.083)	0.139	0.396	0.803
Acceptance	478	6.46	6.43	-0.03	-0.018 (-0.178 to 0.142)	-0.021	0.826	0.967	-0.032 (-0.314 to 0.249)	-0.039	0.821	0.968
Organization	474	4.57	4.57	0.01	0.008 (-0.178 to 0.194)	0.007	0.933	0.967	0.014 (-0.312 to 0.340)	0.013	0.931	0.968
Learning Materials	474	3.66	3.46	-0.20	-0.148 (-0.464 to 0.168)	-0.077	0.357	0.813	-0.266 (-0.822 to 0.290)	-0.139	0.347	0.803
Involvement	474	3.28	3.1	-0.19	-0.164 (-0.437 to 0.108)	-0.113	0.237	0.713	-0.295 (-0.768 to 0.178)	-0.202	0.221	0.687

Variety	478	2.16	1.98	-0.18	-0.133 (-0.323 to 0.057)	-0.12	0.17	0.655	-0.24 (-0.577 to 0.096)	-0.217	0.161	0.626
HOME Total	478	27.86	27.45	-0.41	-0.251 (-1.136 to 0.635)	-0.047	0.578	n/a	-0.453 (-2.009 to 1.103)	-0.085	0.568	n/a

Individual lines present the results of separate regressions. Sample sizes vary by instrument due to the collection method (with less households assessed with HOME than children nutritionally). For each estimation we present: the means of the outcome by group (treatment and control) and the difference in means, and the results of the estimations for intent-to-treat and treatment-on-treated. β s are estimated controlling for the corresponding pre-test (except for HOME for which there was no pretest), age, age squared, male, black, mother marital status (single), maternal years of education, household wealth index, household kids, childcare experience before randomization, neighborhood, cohort of birth and Tester FE. P-values ≤ 0.10 are in bold type.

*Means are age adjusted using ANOVAs.

†Effects sizes, also known as Cohen’s D, are β s interpreted as fraction of SD in control group at baseline (except for HOME for which there was no baseline testing).

‡Standard P-values.

§Stepdown P-values are for Romano and Wolf (2005) Stepdown procedures applied to blocks of outcomes per type of developmental dimension measured. Not calculated for the Total’s aggregate scores as these are aggregate measures across various dimensions or for one developmental domain.