NBER WORKING PAPER SERIES

EARLY CHILDHOOD INTERVENTION FOR THE POOR: LONG TERM OUTCOMES

Pamela Jervis
Lina Cardona-Sosa
Michele Giannola
Sally Grantham-McGregor
Costas Meghir
Marta Rubio-Codina
Monimalika Day
Orazio Attanasio
Alison Andrew
Britta Augsburg

Working Paper 32165
http://www.nber.org/papers/w32165

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2024, Revised January 2026

The trial was implemented with extreme care and attention by Pratham, under the leadership of Rukmini Banerji. Swarnaprava Pradhan and Puspanjali Parida provided excellent implementation in the field. Karishma Vas and Pankhuri Mishra were instrumental in achieving the follow up. The present follow-up study was funded by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Advanced Grant 695300 - HKADeC - ERC-2015-AdG/ERC-2015- AdG, PI Attanasio) and the Economic and Social Research Council (ESRC) (Grants ES/M010147/1 and ES/K010700/1, PI O. Attanasio). The original trial was funded by a personal donation from Mr. Rushton Turner, obtained following a presentation and a discussion with Costas Meghir, which was turned into a grant (PI's O. Attanasio, B. Augsburg, S. Grantham McGregor, C. Meghir and M. Rubio-Codina). Funding was also provided by the Waterloo Foundation (Grant No 969-1310, PI's O. Attanasio, B. Augsburg and M. Rubio-Codina) and by the National Institutes of Health, USA (Grant R01 HD 72120, PI C. Meghir). Britta Augsburg was one of the initiators of the original experiment. Alison Andrew and Britta Augsburg worked on the design and implementation of the original experiment and we thank them for their contribution. P. Jervis gratefully acknowledges financial support from the Institute for Research in Market Imperfections and Public Policy MIPP (ICS13 002 ANID) and the Center for Research in Inclusive Education, Chile (SCIA ANID CIE160009). ISRCTN89476603, AEARCTR-0000169, AEARCTR-0005444. The study was reviewed and approved by the Research Ethics Committees of University College London (UCL), UK (IRB Approval Number 2168/001) and of the Institute for Financial Management and Research, India (IRB00007107). The follow-up has also obtained ethical approval from the UCL Research Ethics

Committee (IRB Approval Number 16727/001) and the registry trial can be found at https://www.socialscienceregistry.org/trials/5444. The children's caregivers provided written informed consent before study participation. The views here presented do not represent the Inter-American Development Bank, its board of directors, the countries they represent, or the National Bureau of Economic Research.

Early Childhood Intervention for the Poor: Long Term Outcomes
Pamela Jervis, Lina Cardona-Sosa, Michele Giannola, Sally Grantham-McGregor, Costas Meghir, Marta Rubio-Codina, Monimalika Day, and Orazio Attanasio
NBER Working Paper No. 32165
February 2024
JEL No. I25, I30, I38, J13, O15

ABSTRACT

Early childhood interventions aim to improve parenting, children's skills, thus reducing poverty. While short-term effects are well documented, evidence on longer-term impacts in low-income countries remains limited. We present results of a medium-term follow-up to a scalable home-visiting randomized intervention in India that previously improved children's cognitive and language development. After 4.5 years (average child age 7.5 years), significant impacts persisted in early numeracy (0.33 SD, p=0.007) and marginally in literacy (0.27 SD, p=0.064), driven by the most disadvantaged children. IQ differences faded. Parental investment in education and participation in stimulating activities with the child remained increased.

Pamela Jervis
Universidad de Chile
Department of Industrial Engineering
and Institute for Fiscal Studies
pjervisr@uchile.cl

Lina Cardona-Sosa
The World Bank
lcardonasosa@worldbank.org

Michele Giannola
University of Naples Federico II
Department of Economics and Statistics
and Institute for Fiscal Studies and CSEF
michele.giannola@unina.it

Sally Grantham-McGregor
Institute of Child Health
sallymcgregor@yahoo.com

Costas Meghir
Yale University
Department of Economics
and IZA
and also NBER
c.meghir@yale.edu

Marta Rubio-Codina
Inter-American Development Bank
martarubio@iadb.org

Monimalika Day
Dr. B.R Ambedkar University Delhi
monimalika@aud.ac.in

Orazio Attanasio
Yale University
Department of Economics
and CEPR
and also NBER
orazio.attanasio@yale.edu

A randomized controlled trials registry entry is available at https://www.socialscienceregistry.org/trials/5444/his

1 Introduction[1]

In India, as in many other Low and Middle-Income Countries (LMICs), school enrollment has increased steadily (UNESCO Institute for Statistics, 2025), yet learning outcomes have remained stagnant at very low levels (Angrist et al., 2021). "Schooling without learning" remains a pressing concern, as documented by repeated large-scale surveys (ASER, 2023). A central question is why expanded schooling has not translated into commensurate gains in learning.

One strand of literature emphasizes the quality of education provided (Duflo, Hanna, and Ryan, 2012; Banerjee et al., 2017; Akyeampong et al., 2023). A complementary line of work highlights the critical importance of the early years, emphasizes the formative role of the early years, showing that deficits arising in early childhood constrain subsequent skill formation through what has been termed dynamic complementarities (Lancet, 2007; Cunha and Heckman, 2007; Cunha, Heckman, and Schennach, 2010). Evidence from India underscores this point: a longitudinal study following 14,000 preschool children across three states finds that low cognitive and language skills at age five strongly predict poorer performance in primary school (Kaul and Bhattacharjea, 2019).

The presence of dynamic complementarities implies that improvements in school quality are likely to be most effective when they build on strong early foundations. Interventions such as *Teaching at the Right Level* (Banerjee et al., 2017) or policies aimed at reducing absenteeism (Duflo, Hanna, and Ryan, 2012) yield larger gains for children who enter school with higher levels of cognitive and socio-emotional development. The implications are par-

[1] **Author Contribution:** The original trial was designed, directed and analyzed by Alison Andrew, Orazio Attanasio, Britta Augsburg, Monimalika Day, Sally Grantham-McGregor, Costas Meghir and Marta Rubio Codina. The contributions to the Followup study are: PJ and SG-Mc contributed to the choice of tests to measure child development for the follow up. PJ, LC and MG designed the data collection instruments and carried out analysis. LC conducted the data cleaning. PJ, LC, MG, and SG-Mc contributed to the interpretation and writing the paper. LC and MG participated in the piloting of child development's tests, PJ trained the trainers, LC and MG trained and supervised field staff for data collection. SG-Mc, CM, MRC, MD and OA participated in the initial trial. PJ, LC, MG, SG-Mc, CM, MRC, MD and OA critically reviewed the paper for important intellectual content.

ticularly stark in India, where widespread early-life deprivation limits many children's ability to benefit from rapid economic growth. The resulting inequality—generated by the dynamic process of child development, as discussed in Attanasio (2025)—may ultimately become a binding constraint on aggregate growth itself (Aghion, Almås, and Meghir, 2025).

Consistent with this perspective, evidence shows that gaps in cognitive development between children from poor and middle-class families emerge early and widen steadily with age (Rubio-Codina et al., 2015; Rubio-Codina and Grantham-McGregor, 2020). Research in Early Childhood Development (ECD) highlights that these early deficits, associated with growing up in poverty, are difficult to reverse because they become biologically embedded, affecting brain development and the endocrine system (Jensen et al., 2021), with apparent delays even in the first year. These delays increase with age and are related to poor school performance and adult outcomes. Although such issues arise wherever poverty exists, they are particularly salient for LMICs (Black et al., 2016; Hamadani et al., 2014), putting millions of children at risk of not reaching their developmental potential (Bhutta et al., 2022).

In most societies, very young children spend much of their time at home with their mothers, making the quality of home interactions and material resources critical for development (Attanasio et al., 2020; Attanasio, Cattan, and Meghir, 2022). Poverty restricts both resources and parental bandwidth, often generating stress and anxiety that limit the ability of parents to provide stimulating and nurturing care (Engle and Black, 2008; Mani et al., 2013; Kandpal and Baylis, 2019). Unequal gender relations and early motherhood further exacerbate these adverse conditions (Calvi, 2020).

Since the first *Lancet Series on Child Development* (Grantham-McGregor et al., 2007), which documented the large share of children in low- and middle-income countries (LMICs), under 5 years of age, who do not reach their developmental potential, there has been growing interest in designing scalable interventions that can break this cycle (Richter et al., 2017). Many of these build on the intervention by Grantham-McGregor et al. (1991). Originally

developed in Jamaica and now known as *Reach Up and Learn*[2], it emphasizes caregiver–child interactions and stimulation in the home environment. It produced substantial short- and long-term gains, especially among vulnerable children (Gertler et al., 2014).

In the short term, many ECD programs beyond the Jamaica intervention have improved child development and parental investment (see, e.g., Attanasio, Cattan, and Meghir, 2022). However, some short-term impacts may reflect temporary mechanisms, such as transitory parental behavioral changes that fade over time (Heckman et al., 2020; Andrew et al., 2018). Whether early interventions can permanently alter life trajectories depends on the dynamic process of skill formation and later investments, which may complement the human capital accumulated early in life (Heckman, Pinto, and Savelyev, 2013; Garcia et al., 2021; Attanasio, Meghir, and Nix, 2020).

The evidence on whether early gains persist is limited and mixed in LMICs, with the most compelling evidence being that of the Jamaican intervention. Although some studies identify sustained benefits, such as in Uganda, where Kakwangire et al. (2024) report lasting cognitive impacts at age 8 or in Bangladesh, where Hossain et al. (2023) find sustained effects for non-anemic children, others point to a pattern of "fade-out." For instance, in Colombia, initial improvements in child development had vanished after two years (Attanasio et al., 2014; Andrew et al., 2018). Similarly, the *Thula Sana* intervention in South Africa found no cognitive effects at a 13-year follow-up (Tomlinson et al., 2021), and clinic-based programs in Jamaica showed no sustained gains at age six (Smith et al., 2021). This heterogeneity, along with the possibility of impacts that attenuate in the medium term may re-emerge later (Jeong, Pitchik, and Fink, 2021; Bailey et al., 2017; Heckman, Pinto, and Savelyev, 2013), underscores the importance of tracking participants over a substantial part of the life cycle to understand the dynamics of skill formation. The fact that long-lasting effects have been observed more consistently in high-income countries (Campbell et al., 2014; Heckman et al., 2010; Doyle, 2020) may indicate the importance of the specific setting for sustained impacts,

[2] `https://reachupandlearn.com/`

such as institutions, school quality, and initial conditions, although there are too few long term follow ups in LMICs to establish this.

In this paper, we examine the medium term effects of a home-visiting early childhood intervention, in Cuttack, India, targeted at children who were 10-20 months old at the start (Andrew et al., 2019). Like the Jamaican study, our research focuses on peri-urban slum populations, which in our case largely consists of recent migrants from rural areas seeking economic opportunities. This focus on slum-dwelling children is especially relevant for rapidly urbanizing countries such as India. However, unlike the Jamaican study, we did not specifically target stunted children,[3] although stunting rates in our setting remain high—around 22%, compared with 30% in the broader Odisha population (Sah et al., 2024). Our study therefore examines a broader poor population, resembling the target group of a scaled-up intervention.

We implemented the *Reach-Up-and-learn* program, adapted to our specific context. The program was designed to be scalable both in cost and in delivery. It relied on local community workers to improve caregiver–child interactions and the home environment. The participants were extremely poor and socially vulnerable, but the baseline data indicated high aspirations for their children, suggesting both the need and the openness to interventions that promote human development.

(Andrew et al., 2019), showed that at the end of the intervention, children's cognitive development, parental investments, and maternal mental health improved substantially. Here, we assess whether these gains were sustained when children were aged seven to eight, four and a half years later. We now find that the treated children maintained significant gains in overall achievement, comprising numeracy and literacy, dimensions directly related to future skill accumulation and school success, although the IQ did not improve significantly. Importantly, the effects were concentrated among the children who were most vulnerable at

[3]Stunting is defined as height-for-age <-2 standard standard deviations of WHO standard.

baseline, measured by height for age, allowing them to catch up with their less vulnerable peers.

This medium-term follow-up occurred at a critical juncture, around the onset of formal schooling. In a context where hundreds of millions of children in LMICs leave school without basic numeracy and literacy skills (Snilstveit et al., 2017), and where early skills strongly predict later attainment (Heckman, 2006), sustained early gains appear particularly promising. Our findings underscore the need and feasibility of interventions that promote early development and school readiness, which are essential prerequisites for skill accumulation and ultimately poverty reduction.

The remainder of the paper proceeds as follows. Section 2 describes the intervention and the experimental design; Section 3 presents the follow-up study; Section 4 reports the results; and Section 5 concludes with broader implications.

2 The parenting intervention and its short run impacts

The short-term effects of the intervention under study were assessed through a Randomized Controlled Trial (RCT) and are reported in Andrew et al. (2019). We briefly describe the original trial, the main outcomes that were measured and the results that were obtained.

The Intervention. The study was carried out in 2012 in the peri-urban slums of the city of Cuttack in Odisha, India. It implemented an intervention based on the Jamaican home-visiting program Reach Up (Grantham-McGregor and Smith, 2016) adapted to the socio-cultural context of this target population.

The intervention, implemented in collaboration with Pratham[4], consisted of weekly one-hour home visits with the target child and the primary caregiver over a period of 18

[4]Pratham is the major NGO in India working on education. They were our partner in this and other experimental projects in India (https://www.pratham.org/about/).

months. The objective of the visits was to improve caregiver-child interactions, and to support the caregivers' ability to assist their child's development through developmentally appropriate activities that tapped into daily routines. One of the main ideas of the approach is to provide parents with practical demonstrations of activities that are likely to be developmentally productive at the specific age of their child, and provide an opportunity for the parent to practice and receive positive feedback (Grantham-McGregor et al., 1991).[5]

The visits were conducted by local women with no specific qualifications. Twenty-seven home visitors (one per treatment slum) were recruited among local women and trained for four weeks, with two subsequent refresher training lasting two-to-three days each. During the weekly sessions, the visitor (HV) worked through the structured curriculum that emphasized the development of cognitive and language skills through increased interactions between the child and caregivers and introduced play activities, picture books, images to stimulate storytelling and conversation, and homemade toys from surplus household materials, all of which involve almost no financial cost. The HVs left the toys in the home until the next visit, and encouraged caregivers to continue playing and chatting with the child between visits, give positive reinforcement, and eliminate physical punishment. Materials were replaced each week with new ones, according to the curriculum, and none were left with the family at the end of the intervention.

The local women who delivered the intervention were between 18 and 55 years old, and 74.1% had completed high school. Three mentors with prior field experience received training in key principles of child development and were responsible for training, supervising, and supporting the home visitors through weekly meetings. These mentors were employed by our implementing partner, Pratham, which managed all aspects of program delivery. A psychologist provided ongoing technical support throughout the intervention period. Over

[5]Countries in which Reach Up has been adapted include Bangladesh, Brazil, China, Colombia, India and Peru. A review of its evaluations can be found in Jervis et al. (2023a). See also Grantham-McGregor et al. (1991), Aboud and Yousafzai (2015), Attanasio et al. (2020) and Jeong et al. (2021) for the importance of encouraging parenting practices to support the socio-emotional development of infants and young children.

the 18-month period, the cost per treated child was US$251 at a time when GDP per capita in India was US$1,434 (nominal 2012 prices). This cost could be lower if the program were implemented on a scale, due to scale economies.[6]

The intervention tapped into existing local skills, knowledge and connections by recruiting local women. Moreover, the HV were, in most cases, known to and trusted by the families that they were visiting to deliver the intervention. These features are important for the community buy-in, scalability, and sustainability of such interventions.

The Experimental Design. We selected 54 urban slums (clusters). These clusters were well defined geographical areas delimited by waterways and roads. Within each cluster, eligible households with a child aged 10-20 months at baseline, who was not a twin and had no obvious disability, were identified through a door-to-door census. The original RCT identified 775 eligible households, of which between 7-9 children per cluster were randomly selected for inclusion in the study. Consent was obtained before treatment status was revealed. This procedure resulted in 421 children enrolled in the study. In the Appendix we show the relevant power calculations for the original experiment.

Randomization and Blinding. After an initial census, the 54 study clusters were stratified into 3 groups by the number of eligible children identified in the cluster (fewer than nine, 21%, nine or more, 66%, and where adjacent areas were added, 13%). Then within each stratum, clusters were randomized to the treatment or control group (27 clusters in the treatment group and 27 clusters in the control group), using computer generated random numbers. While it was not feasible to blind households to their allocation to the home visiting program, testers and interviewers were blind to the treatment status of participants.

[6]The home-visiting branch of the larger scale experiment described in Grantham-McGregor et al. (2020) cost $135 per child per year, which in real terms is approximately 30% less.

Measured Outcomes and and Short-Run Results. Baseline data was collected in November-December 2013, and outcomes at the end of the intervention were measured in May-June 2015. At baseline, child development was measured using an adapted version of the Ages and Stages Questionnaires, third edition (ASQ-3) (see Squires and Bricker, 2009). At endline, child development was measured with the Bayley Scales of Infant and Toddler Development, third edition (BSID III, see Bayley, 2006). As reported in Andrew et al. (2019), at endline, after 18 months of treatment, children's cognition had benefited significantly by 0.349 SD (Romano-Wolf multiple hypotheses testing (RW)[7] p-value =0.017). Other measured outcomes also improved but not as significantly: receptive language by 0.224 SD (RW p-value=0.18) and expressive language by 0.192 SD (RW p-value=0.18) and fine motor skills by 0.111 (RW p-value 0.36).[8] In heterogeneity analyses children whose mothers had higher levels of education (8th Standard or better) improved more than those with mothers who had lower educational levels (the effect size on the Bayley was 0.353 (standard error 0.148) versus -0.025 (standard error 0.161), p-value of the difference = 0.024) and boys improved more than girls (but the difference was not significant, except for receptive language). Also, children who were moderately stunted at baseline improved significantly more than non-stunted children (the effect size on the Bayley was 0.794 (standard error 0.266) versus 0.158 (standard error 0.140), p-value of the difference = 0.019). Finally, mothers' depressive symptoms decreased by 0.27% SD in the CESD-10 scale (p-value 0.013)[9]

3 Follow up study

The follow up study collected data on children's achievement in literacy, numeracy and intelligence four and half years after the end of the intervention, when the children were

[7](Romano and Wolf, 2005b) Four hypotheses tested.

[8]The results in the text do not adjust for covariates. Adjusting for prespecified baseline controls, the estimated effect sizes and corresponding stepdown p-values for the Bayley-III Z-Scores are: Cognition 0.293 (RW p=0.053), Receptive language 0.180 (RW p=0.319), and Expressive language 0.111 (RW p=0.456). Here and in what follows, impacts are measured in terms of Standard Deviation (SD) of the control group.

[9]Other than for gender, the heterogeneity analysis is reported in the online section for supporting information of Andrew et al. (2019).

around seven to eight years of age.

3.1 Descriptive statistics, balance and attrition

Descriptives and Balance. Table 1 reports some descriptive statistics of the sample that was enrolled at the baseline of the original study. The sample was economically disadvantaged with 49% of the households living below the poverty line, and over 20% of the children being stunted at baseline. Only 60% of the households in the sample had piped water connection. However, almost the whole sample had an electricity connection. At baseline, 51% of the children were boys and 49% were girls, and the mean age was 15 months.

Table 1 shows that the characteristics of the treatment and control groups were balanced at baseline, except for maternal education, which was higher in the treatment group although this difference is not statistically significant once p-values are corrected for multiple hypothesis testing. Nevertheless, we control for this imbalance in the analysis, as pre-specified for the original trial.

Attrition. At the first follow-up (endline of the intervention), we were able to locate 378 children. Between December 2019 and January 2020, four and a half years after the conclusion of the intervention, we attempted to locate all the 421 children who were initially enrolled in the study and were able to track 314 children (74.6% of the original sample).

Attrition was not related to treatment status. The p-value for the effect of treatment on attrition in both follow ups is 0.26 (the coefficient on treatment is 0.025, with an associated standard error of 0.022). Considering the second follow-up only, treated and control children were equally likely to be interviewed (the coefficient on treatment is 0.037, p-value=0.46). Moreover, baseline characteristics of the subjects lost do not differ significantly from those remaining in the study. Of the 12 characteristics we considered the fine motor ASQ score difference between those remaining and lost had a p-value of 0.04 and for the ASQ commu-

Table 1: Baseline Balance

	Treatment (T)		Control (C)		p-value: Treatment = Control
	Mean	S.D.	Mean	S.D.	
Age in months	14.721	3.066	15.112	3.234	0.589
Male %	0.565	0.497	0.476	0.501	0.053
Firstborn %	0.478	0.501	0.467	0.500	0.671
Mother's years of education	8.091	3.359	6.722	3.842	0.005
Asset index Z-score	0.133	0.925	-0.126	0.932	0.057
ASQ-3 problem solving Z-score	0.029	0.986	-0.028	1.004	0.859
ASQ-3 communication Z-score	-0.015	1.041	0.015	0.947	0.499
ASQ-3 fine motor Z-score	-0.035	1.030	0.035	0.959	0.867
Stunted	0.209	0.407	0.233	0.424	0.180
Maternal knowledge of child development Z-score	-0.028	0.793	0.027	0.867	0.928
Quality of home environment Z-score	0.043	0.883	-0.043	0.763	0.172
Maternal depressive symptoms Z-score	-0.026	0.897	0.025	0.837	0.779
Below Urban Poverty Line	0.483	0.501	0.486	0.501	0.345
Income (Rs) per Capita per Day	109.747	218.092	98.446	218.178	0.877
Roof made from metal sheet/thatch/polyethylene	0.388	0.488	0.462	0.500	0.378
House has dirt floor	0.057	0.233	0.052	0.222	0.958
House has piped water connection	0.598	0.491	0.575	0.495	0.661
Household has electricity connection	0.986	0.119	0.986	0.118	0.585
Household owns a fridge	0.402	0.491	0.308	0.463	0.107
Observations	209		212		

Notes: The table reports balance in baseline characteristics between the treatment group and the control group. p-values are for tests of equality of the means across treatment and control groups. Z-scores have a mean 0 and standard deviation of 1 in the control group. [a] The Urban poverty line as defined by the Rangarajan committee is Rs. 47 per household member per day. [b] The average exchange rate during the baseline survey (November/December 2013) was Rs. 62/USD. ASQ-3 is the Ages and Stages Questionnaire (3rd edition), a survey instrument to measure child development based on parent report.

nication score 0.039. The remaining 10 p-values range from 0.088 to 0.913.[10]

3.2 Outcome measures

Children's age at the time of the follow-up ranged between 6.8 and 8 years (mean 7.4 SD 0.26). The primary outcomes for the follow-up were children's literacy and numeracy skills, as well as their general intelligence (IQ, which includes cognition, language, and components of executive functioning). To assess literacy, we combined two instruments, which measure foundational reading skills and early literacy. For foundational reading skills, we used a test developed in India, the Annual Status of Education Report (ASER), which builds on Pratham's experience in evaluating children's literacy skills throughout India and which has been previously used to evaluate pre-school interventions (ASER, 2012; Vagh, 2010; Dillon et al., 2017). For early literacy we selected age-appropriate items from the "Early Literacy" sub-scale of a Screening-for-School-Readiness test adapted for India to evaluate a pre-school intervention (Meghir et al. (2023)), which includes 21 items for naming letters, matching letters and matching objects with letters, and rearranging letters into words questions. To assess numeracy we combined two instruments that measure math facts fluency and early numeracy. For math facts fluency, we developed and piloted a task based on math facts fluency sub-test of the Woodcock-Johnson-IV "Tests of Achievement - 20" (Schrank, Mather, and McGrew (2014)). The task measures the ability to solve 60 simple addition and subtraction problems in a set time of 3 minutes. For early numeracy we selected age-appropriate items from the "Number Concepts" sub-scale of the same Screening-for-School-Readiness test adapted for India (Meghir et al. (2023)) that includes 16 items for naming numbers, matching quantities with numbers, greater or less, and addition/subtraction questions. To assess IQ, we selected sub-scales from the Wechsler Preschool and Primary Scale of Intelligence-IV (WPPSI-IV) (Wechsler, 2012), which has been adapted for India to evaluate a pre-school

[10]These results are presented in Table form in Appendix Tables A.2 and A.3.

intervention (Meghir et al. (2023))[11]. All sub-scales were combined into the Full-Scale-IQ (FSIQ).

All tests were administered in the children's homes. Testing was conducted by XXX trained enumerators who were blinded to the treatment allocation of the child. All instruments were previously adapted to be culturally appropriate without changing the underlying constructs. The instruments showed good inter-tester reliability (Cohen's kappa was above 0.9 for all instruments).

To better understand how the intervention operated and the extent to which it changes the behavior of the parents and the investments they make in their children we measure several outcomes relating to parent child interactions and the environment in the home relating to child upbringing and stimulation.

Such secondary outcomes were collected through a structured interview conducted with the child's primary caregiver to obtain details on family characteristics, standard of living, and stimulation in the home together with home environment quality. For this purpose, we use the United Nation's Children's Fund family care indicators (FCI; Kariger et al. (2012)) and The Early Childhood Home Observation for the Measurement of the Environment (EC-HOME; Bradley et al. (1988)). The FCI includes a play-materials scale, which records the presence of toys and books, and a play-activities scale, which includes adult involvement with the child in certain play activities. For the EC-HOME, we selected items from the scales Acceptance, Responsivity, Physical Environment, Language Stimulation and Academic Stimulation. These items ask about the type and variety of activities the child engaged in with an adult over the preceding days (such as whether the child received help with homework, or whether an adult discussed schoolwork with the child).

We also collected information on educational expenditure and the primary caregivers'

[11]The following subscales were used: Verbal Comprehension, Visual Spatial, Working Memory, and Processing Speed Indices and Matrix Reasoning sub-scale, which is used internationally to measure children's general intelligence.

expectations and aspirations regarding their children's skills and educational attainment. Finally, we collected measures of the primary caregiver's self-efficacy (Pedersen et al., 1989; Porter, 2003) and depression (Center for Epidemiologic Studies Depression; CES-D; Radloff (1977)). For the self-efficacy and depression scales we constructed summary indices as simple averages of the responses.

3.3 Empirical Analysis

We present impact estimates on four individual child assessments, two for numeracy (math facts fluency and early numeracy) and two for literacy (foundational reading skills and early literacy), as well as the general IQ.

To construct these primary outcome variables, we standardize internally the outcome variables related to child numeracy, literacy, and IQ to remove age and tester effects. Specifically, for each test, we separately regress the raw test scores on tester dummies and take the residuals. Those residuals are standardized using the empirical age-conditional means and standard deviations each estimated using non-parametric regression methods on the control group sample.[12]

We also assess impacts on summary indices computed by combining the outcomes related to the same dimension of child development, one for numeracy and one for literacy. In this case, we take the average of the standardized test scores we are combining. Then we re-standardize to have a mean 0 and standard deviation of 1 in the control group.[13] We also report impacts on a overall achievement index combining the two literacy and two numeracy scores with the same procedure described above. Finally, the item scores of the FCI and HOME scales were created as the sum of different questions from the FCI and HOME

[12]See the description of the procedure in section B.4 of the online appendix of Attanasio et al. (2020) found here: `https://www.aeaweb.org/content/file?id=11163`. Briefly, we regress the raw scores on tester dummies and compute the residuals from this regression. We then use a kernel-weighted local polynomial regression to estimate the age-specific mean and standard deviation of these residuals in the control group and use these to create age-standardized z-scores.

[13]When one of the two scores was missing the index was set to missing.

questionnaires and then standardized as above. Given our standardization, the estimated effect sizes for all outcomes are measured in standard deviation units (SD) in the control group.

We obtain the estimates by regressing the standardized scores on treatment status and including baseline controls for ASQ-3, child gender, maternal education and strata fixed effects. All the estimates should be interpreted as intention to treat, as we do not control for compliance with the intervention or the number of visits each households received, for which we have no information.

Standard errors have been computed using 5000 bootstrap replications. The resulting p-values are two-sided and have been adjusted for multiple testing using the stepdown procedure described in Romano and Wolf (2005a,b).

4 Results

Primary Outcomes. In Table 2, we report the estimated impacts on our primary outcomes, four and a half years after the end of the intervention, when the children in the sample are between 6.8 and 8 years of age. The effects are in units of standard deviation of the control group (SD). In panel A, we consider an overall index of achievement in numeracy and literacy. The intervention caused a highly significant improvement of 0.332 SD (p-value=0.009) in the overall achievement index.

In Panel B of Table 2, we report the impacts of the intervention separately on the numeracy and literacy summary indices, which form part of the overall achievement index. The impact on numeracy is 0.33 SD and strongly significant (p-value 0.007). The impact on literacy is 0.272 SD with a p-value of 0.06, which has been adjusted for the fact that we are testing two hypotheses (numeracy and literacy).

In Panel C, we further breakdown the impacts on individual components of the sum-

mary indices of numeracy and literacy, and the p-values reported adjust for testing the four hypotheses in that panel. The estimates show substantial impacts on all subscales ranging from 33% to 17% of a standard deviation. The math facts fluency test is significant at the 3% level, while the early numeracy scale is significant at 5.4%.

Table 2: Treatment Effects at Second Follow-up

	Effect	Observations
Panel A:		
Achievement index	0.332	279
	(0.127) [0.009]	
Panel B: Summary Indices		
Numeracy	0.330	314
	(0.117) [0.007]	
Literacy	0.272	279
	(0.133) [0.064]	
Panel C: Tests		
Math facts fluency scale	0.330	314
	(0.130) [0.03]	
Early numeracy scale	0.256	314
	(0.109) [0.054]	
Foundational reading scale	0.269	279
	(0.130) [0.065]	
Early literacy scale	0.173	314
	(0.124) [0.142]	
Panel D: Full-Scale-IQ		
Full-Scale-IQ	0.083	314
	(0.146) [0.289]	

Notes: The outcome variables were internally standardized, removing tester and age effects and have a mean of zero and standard deviation of one in the control group. Controls include child gender, age, baseline ASQ scores, maternal education, and randomization strata fixed effects. Panel A shows the results for the achievement index. Panel B shows the results for the summary indices for numeracy and literacy as described in the text. Panel C shows the results on individual tests. Standard errors in parentheses are computed using the bootstrap with 5000 replications and are clustered at the slum level. Romano-Wolf step down p-values in square brackets are reported for the treatment effects on the numeracy and literacy indices (2 hypotheses) and separately for the four estimates in the Panel C (4 hypotheses). The number of observations is lower for the literacy scale (i.e., 279) and, as a result, for the achievement index that includes it, due to missing data on 35 children's ASER test (that measures foundational reading skills).

Finally, in the last row of the table we report the Full-Scale-IQ as measured by the

WPPSI-IV. While the estimated coefficient is positive, the effect is statistically insignificant and hence suggests that the program had no long run effects on children's general IQ.

Secondary Outcomes. Table 3 shows the impacts on a number of secondary outcomes measured four and a half years after the intervention ended relating to parental investments in children, aspirations, and maternal mental health. In Panel A, two results that are directly relevant to the intervention stand out: parents were more likely to report engaging in stimulating activities with their child (FCI and HOME activities 0.293 with a p-value of 0.038), and parents increased educational expenditures (0.62SD, p-value=0.006). Other secondary outcomes related to parents' investment decisions, including material investments in toys and learning materials, school enrollment, school absenteeism, perceived ability of the child, and parental aspirations for the child's educational attainment, did not show evidence of improvement.

Table 3: Treatment Effects on Intermediate Outcomes

Panel A:	Maternal depression (Z-score)	Maternal Self-efficacy (Z-score)	Materials (Z-score)	FCI and HOME Activities (Z-score)	Educational Expenditures (Z-score)
	(1)	(2)	(3)	(4)	(5)
Treatment	-0.033	0.044	-0.097	0.293	0.618
	(0.155)	(0.136)	(0.095)	(0.141)	(0.227)
P-value	0.834	0.745	0.305	0.038	0.006
Mean outcome control	-0.009	0.003	0.000	0.010	8.083
Observations	312	312	312	312	312

Panel B:	Enrolled in school	Days absent from school	Perceived child ability	Educational attainment wished
	(6)	(7)	(8)	(9)
Treatment	0.017	-0.313	0.083	0.172
	(0.017)	(0.465)	(0.205)	(0.113)
P-value	0.302	0.501	0.687	0.126
Mean outcome control	0.975	2.660	8.951	13.521
Observations	287	303	312	310

Notes: The table report treatment effects on intermediate outcomes. Standard errors are computed using the bootstrap with 5000 replications and are clustered at the slum level. Z-scores are standardized to have a mean 0 and standard deviation 1 in the control group. Days absent from school refers to the 20 days preceding the interview (excluding holidays).

We did not find any improvements in primary caregiver's self-efficacy or depression, although we found an improvement in the latter at endline, immediately after the intervention ended (0.266 SD on the CESD-10 score, p-value=0.013). We note that despite not finding persistent improvements, addressing maternal depression in the short term, possibly by the provision of social support (Andrew et al., 2023), could have been an important mechanism

underlying the initial and sustained impact of the intervention on child development and investments, consistent with Baranov et al. (2020).

These results speak to the mechanisms underlying the effects: the program got parents to do more and indeed this improved level of investments persists even four years later. By contrast, in the Colombia intervention where the effects did not persist, Andrew et al. (2018) report that parental investments also declined relative to the control group. Moreover, Attanasio et al. (2020) show that the impact of the Colombia intervention operated exclusively through its impact on parental investments. It seems therefore that a key to persistence of the effects may be changing parental behavior permanently, at least in these first few years.

Heterogeneity analysis: Stunting and the impact of the intervention. Earlier interventions, such as the Jamaica study (Grantham-McGregor et al., 1991) as well as a more recent study in Bangladesh (Hamadani et al., 2019), have shown endline results that are at least twice as large as the ones we reported for our own endline (Andrew et al., 2019). However, these other study populations were selected to include only malnourished children, measured by the extent of stunting or being underweight[14]. In this study, there was no screening for malnutrition, but rather the selected group was from a general population in poverty (defined by living in a slum), which includes about 20% stunted children; (see also Jervis et al., 2023b, on this issue). Indeed, as we reported at the end of Section 2, our intervention shows similar impacts to that of Jamaica at endline once we focus on stunted children. Beyond the scientific interest, the question is interesting for the better understanding of targeting and for designing interventions that can benefit a broad range of children. We thus examine whether the increased benefit for stunted children persists.[15]

The heterogeneity analysis is presented in Table 4 and shows that the effect sizes at 7-8 years of age are much larger for children who were stunted at baseline with p-values for the

[14]Stunting: height -for-age Z score of $\leq -2SDs$ of the WHO standard. Underweight : weight-for-age Z score $\leq -2SDs$ of the WHO standard

[15]The baseline characteristics are also balanced conditional on stunting.

Table 4: Heterogeneity in Treatment Effects at Second Follow-up

	Heterogeneity by Stunting at Baseline		
	Stunted	Treatment	Non-stunted × Treatment
Panel A:			
Achievement index	-0.649	0.682	-0.486
	(0.181)	(0.270) [0.009]	(0.283) [0.084]
Panel B: Summary Indices			
Numeracy	-0.550	0.692	-0.502
	(0.170)	(0.231) [0.004]	(0.247) [0.094]
Literacy	-0.729	0.542	-0.391
	(0.180)	(0.295) [0.094]	(0.303) [0.15]

Notes: Controls include child gender, age, baseline ASQ scores, maternal education, and randomization strata fixed effects. Each row represents a separate regression with the dependent variable stated in the first column. The table reports impact heterogeneity by stunting status at baseline. Standard errors (in parentheses) are computed using the bootstrap with 5000 replications and are clustered at the slum level. We report Romano-Wolf stepdown p-values (in square brackets) separately for panel A (two hypotheses) and for panel B (4 hypotheses).

difference ranging from 8.4% overall to 9.4% for numeracy and 15% for literacy. This shows that the differences documented right at the end of the intervention when the children were approximately 3 years old, have persisted. This result is important for various reasons. First, it shows that the intervention can help the most deprived children catch up. In contrast, the intervention did not seem to be able to improve as much the children who were slightly better off, albeit still living in poverty within the slum. Understanding how we can extend the benefits to the broader poor population is a challenge of central policy and scientific importance. Second, we show that for stunted children we obtain results that are not too different from those of the Jamaica intervention.[16] Therefore, the results of this experiment, designed to be scalable, reinforce the message that ECD interventions at least for the most deprived children can have life-changing impacts and should form a key component of human capital policy.

Finally, as discussed in section 2, the endline results showed impact heterogeneity by

[16] of the Jamaican studies some of its benefits at 7 years of age, but they reappeared in IQ in the medium term follow up at 11 to 12 years.

mother's education and (suggestively) also by child gender. In this medium-term follow up, we see no significant differences in these impacts. This suggests that conditional on stunting status, the children of less educated mothers did catch up.

5 Discussion and Conclusions

The medium-term impacts of the intervention we have analyzed demonstrate a remarkable persistence of the benefits four and a half years after it was completed. The initial trial was randomized, the retrieval rates of participants was reasonable with a balanced loss and testers who were blind to the children's group assignment.

These results are important for several reasons. First, very limited evidence exists on the medium- term effects of child development interventions from LMICs. Second, governments in LMICs are particularly concerned about poor literacy and numeracy levels even when children are enrolled in school. Evidence from six longitudinal studies in the US found that preschool math and reading ability on school enrollment are the strongest predictors of later school reading and math achievement, suggesting that the participants may continue to show improved school achievement, with subsequent better life chances (Duncan et al., 2007).

The results are also important because there can be no suggestion that the benefits can be attributed to "teaching to the test". The children were too young at the end of the intervention (28 to 38 months) to be taught literacy and numeracy, for which we report significant impacts. Therefore, the intervention must have improved their ability to learn new skills and knowledge, which could be an example of dynamic complementarities, in which skills beget skills, suggested as a mechanism to explain late onset benefits.

Finally, all this was achieved with local human resources, run by locally available women, without any specific qualifications, trained for this intervention only for a few weeks.

This demonstrates the potential for policy to dramatically improve child outcomes for the most vulnerable without insuperable scaling-up constraints.

The precise mechanism whereby benefits are sustained is unknown, but the size of the effect at the immediate end of the intervention and the quality of the subsequent environment may play a role. The improved mental health of the mothers at endline, the overall improvement in the home environment, and the extra money parents spent on education suggest that the parents were also affected and thus were able to provide better nurturing care and education, which is likely the key mechanism.Attanasio et al. (2020). And returning to the original motivation for ECD, the improved nurturing may have permanent neurological effects on the brain (Jensen et al., 2021).

References

Aboud, F. and Yousafzai, A. (2015). Global health and development in early childhood. *Annual Review of Psychology* 66:433–57.

Aghion, P., Almås, I. and Meghir, C. (2025). Human capital and development. Working Paper 34602, National Bureau of Economic Research.

Akyeampong, K., Andrabi, T., Banerjee, A., Banerji, R., Dynarski, S., Glennerster, R., Grantham-McGregor, S., Muralidharan, K., Piper, B., Ruto, S., Saavedra, J., Schmelkes, S., and Yoshikawa, H. (2023). 2023 cost-effective approaches to improve global learning: What does recent evidence tell us are "smart buys" for improving learning in low- and middle-income countries? Technical report, World Bank Group.

Andrew, A., Attanasio, O. P., Augsburg, B., Behrman, J., Day, M., Jervis, P., Meghir, C., and Phimister, A. (2023). Mothers' social networks and socioeconomic gradients of isolation. *Economic Development and Cultural Change* .

Andrew, A., P., O. A., Fitzsimons, E., Grantham-McGregor, S., Meghir, C., and Rubio-Codina, M. (2018). Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in Colombia. *PLoS medicine* 15(4):e1002556.

Andrew, A., P.Attanasio, O., Augsburg, B., Day, M., Grantham-McGregor, S., Meghir, C., Mehrin, F., Pahwa, S., and Rubio-Codina, M. (2019). Effects of a scalable home-visiting intervention on child development in slums of urban India: Evidence from a randomised controlled trial. *Journal of Child Psychology and Psychiatry* .

Angrist, N., Djankov, S., Goldberg, P. K., and Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature* 592(7854):403–408.

ASER, C. (2012). *Annual Status of Education Report 2011*.

ASER, C. (2023). *Annual Status of Education Report*. https://asercentre.org/wp-content/uploads/2022/12/ASER-2023˙Main-findings-1.pdf.

Attanasio, O. P. (2025). The first 1000 days and beyond: The process of child development. *in preparation for the Journal of Economic Literature* .

Attanasio, O. P., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020). Estimating the production function for human capital: Results from a randomized controlled trial in colombia. *American Economic Review* 110(1):48–85.

Attanasio, O. P., Cattan, S. and Meghir, C. (2022). Early childhood development, human capital, and poverty. *Annual Review of Economics* 14:853–892.

Attanasio, O. P., Fernández, C., Fitzsimons, E. O. A., Grantham-McGregor, S. M., Meghir, C., and 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. *British Medical Journal* 349:g5785.

Attanasio, O. P., Meghir, C. and Nix, E. (2020). Human capital development and parental investment in India. *The Review of Economic Studies* 87(6):2511–2541.

Bailey, D., Duncan, G. J., Odgers, C., and Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Res. Educ. Eff.* 10((1):7–39.

Banerjee, A. V., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M., and Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives* 31(4):73–102.

Baranov, V., Bhalotra, S., Biroli, P., and Maselko, J. (2020). Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial. *American economic review* 110(3):824–859.

Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development-Third Edition: Technical manual*. San Antonio, TX: Harcourt Assessment.

Bhutta, Z. A., Boerma, T., Black, M. M., Victora, C. G., Kruk, M. E., and Black, R. E. (2022). Optimising child and adolescent health and development in the post-pandemic world. *The Lancet* 399(10337):1759–1761.

Black, M. M., Walker, S. P., Fernald, L. C. H., Andersen, C. T., DiGirolamo, A. M., Lu, C., McCoy, D. C., Fink, G., Shawar, Y. R., Shiffman, P. J., Devercelli, A. E., Wodon, Q. T., Vargas-Barón, E., and Grantham-McGregor, S. (2016). Early childhood development coming of age: Science through the life course. *Lancet. 6736* 1–14.

Bradley, R. H., Caldwell, B. M., Rock, S. L., Hamrick, H. M., and Harris, P. (1988). Home observation for measurement of the environment: Development of a home inventory for use with families having children 6 to 10 years old. *Contemporary Educational Psychology* 13(1):58–71.

Calvi, R. (2020). Why are older women missing in india? the age profile of bargaining power and poverty. *Journal of Political Economy* 128:2453–2501.

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., and Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science* 343:1478 – 1485. DOI: 10.1126/science.1248429.

Cunha, F., Heckman, J. and Schennach, S. (2010). Estimating the technology of cognitive and non-cognitive skill formation. *Econometrica* 78(3):883–931.

Cunha, F. and Heckman, J. J. (2007). The technology of skill formation. *Journal of Political Economy* 115(2):187–208.

Dillon, M. R., Kannan, H., Dean, J. T., Spelke, E. S., and Duflo, E. (2017). Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics. *Science* 357(6346):47–55.

Doyle, O. (2020). The first 2,000 days and child skills. *Journal of Political Economy* 128(6).

Duflo, E., Hanna, R. and Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review* 102(4):1241–78.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., et al. (2007). School readiness and later achievement. *Developmental psychology* 43(6):1428.

Engle, P. L. and Black, M. M. (2008). The effect of poverty on child development and educational outcomes.

Garcia, J., Bennhoff, F., Lean, D., and Heckman, J. (2021). The dynastic benefits of early childhood education.

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeerch, C., Walker, S., Chang, S., and Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science* 344:998–1001.

Grantham-McGregor, S., Adya, A., Attanasio, O. P., Augsburg, B., Behrman, J., Caeyers, B., Day, M., Jervis, P., Kochar, R., Makkar, P., Meghir, C., Phimister, A., Rubio-Codina, M., and Vats, K. (2020). Group sessions or home visits for early childhood development in India: A cluster rct. *Pediatrics* 146:e2020002725.

Grantham-McGregor, S., Cheung, Y., Cueto, S., Glewwe, P., Richter, L., and Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *Lancet* 369(9555):60–70.

Grantham-McGregor, S., Powell, C., Walker, S., and Himes, J. (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: The Jamaican study. *Lancet* 338(758):1–5.

Grantham-McGregor, S. and Smith, J. A. (2016). Extending the jamaican early childhood development intervention. *Journal of Applied Research on Children* 7(2):4.

Hamadani, J., Alam, F. T. S. H. D., Ridout, D., Attanasio, O., and Grantham-McGregor, S. (2014). Cognitive deficit and poverty in the first 5 years of childhood in bangladesh. *Pediatrics* 134(4):e1001–8.

Hamadani, J., Mehrin, F., Tofail, F., Hasan, I., Huda, S., Baker-Henningham, H., Ridout, D., and Grantham-McGregor, S. (2019). Integrating an early childhood development programme into Bangladeshi primary health-care services: An open-label, cluster-randomised controlled trial. *The Lancet Global Health* 7(3):366–375.

Heckman, J., Liu, B., Lu, M., and Zhou, J. (2020). The impacts of a prototypical home visiting program on child skills. Working Paper 27356, National Bureau of Economic Research.

Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., and Yavitz, A. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the highscope perry preschool program. *Quantitative Economics* 1(1):1–46.

Heckman, J., Pinto, R. and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103(6):2052–86.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science* 312(5782):1900–1902.

Hossain, S. J., Tofail, F., Mehrin, S. F., and Hamadani, J. D. (2023). Six-year follow-up of childhood stimulation on development of children with and without anemia. *Pediatrics Supplement* 151(Suppl 2):e2023060221E. PMID: 37125884.

Jensen, S., Xie, W., Kumar, S., Petri, R. H. W., and 3rd, C. N. (2021). Associations of socioeconomic and other environmental factors with early brain development in bangladeshi infants and children. *Dev. Cogn. Neurosci.* ;50:100981.

Jeong, J., Franchett, E., de Oliveira, V. R., Rehmani, K., and Yousafzai, A. (2021). Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis. *PLoS Med* 18(5):10:e1003602.

Jeong, J., Pitchik, H. and Fink, G. (2021). Short- term, medium-term and long- term effects of early parenting interventions in low- and middle- income countries: a systematic review. *BMJ Global Health* 6.

Jervis, P., Coore-Hall, J., Pitchik, H. O., Arnold, C. D., Grantham-McGregor, S., Rubio-Codina, M., Baker-Henningham, H., Fernald, L. C., Hamadani, J., Smith, J. A., Trias, J., and Walker, S. P. (2023a). The Reach Up Parenting Program, Child Development, and Maternal Depression: A Meta-analysis. *Pediatrics* 151(Supplement 2):e2023060221D.

Jervis, P., Coore-Hall, J., Pitchik, H. O., Arnold, C. D., Grantham-McGregor, S., Rubio-Codina, M., Baker-Henningham, H., Fernald, L. C., Hamadani, J., Smith, J. A., et al. (2023b). The reach up parenting program, child development, and maternal depression:

a meta-analysis. *Pediatrics* 151(Supplement 2).

Kakwangire, P., Muhoozi, G., Ngari, M., Matovu, N., Westerberg, A., Iversen, P. O., and Atukunda, P. (2024). 8-year follow-up of a maternal education trial in a low-resource setting. *Pediatrics* 153(4):e2023063352.

Kandpal, E. and Baylis, K. (2019). The social lives of married women: Peer effects in female autonomy and investments in children. *Journal of Development Economics* 140:26–43.

Kariger, P., Frongillo, E. A., Engle, P., Britto, P. M. R., Sywulka, S. M., and Menon, P. (2012). Indicators of family care for development for use in multicountry surveys. *Journal of health, population, and nutrition* 30(4):472.

Kaul, V. and Bhattacharjea, S. (2019). *Early Childhood Education and School Readiness in India: Quality and Diversity*. Springer.

Lancet (2007). Early child development in developing countries 2007 .

Mani, A., Mullainathan, S., Shafir, E., and Zhao, J. (2013). Poverty impedes cognitive function. *Science* 341(6149):976–980.

Meghir, C., Attanasio, O. P., Jervis, P., Day, M., Makkar, P., Behrman, J., Gupta, P., Pal, R., Phimister, A., Vernekar, N., et al. (2023). Early stimulation and enhanced preschool: A randomized trial. *Pediatrics* 151(Supplement 2).

Pedersen, F. A., Bryan, Y. E., Huffman, L., and Del Carmen, R. (1989). Construction of self and offspring in the pregnancy and early infancy periods. Technical report.

Porter, . H. H.-C., C. L. (2003). First-time mothers' perceptions of efficacy during the transition to motherhood: Links to infant temperament. *Journal of Family Psychology* 17(1):54–64.

Radloff, L. S. (1977). The ces-d scale: A self-report depression scale for research in the general population. *Applied psychological measurement* 1(3):385–401.

Richter, L., Daelmans, B., Lombardi, J., Heymann, J., Lopez-Boo, F., Behrman, J., Lu, C., Lucas, J., Perez-Escamilla, R., Dua, T., Bhutta, Z., Stenberg, K., Gertler, P., and Darmstadt, G. (2017). Investing in the foundation of sustainable development: pathways to scale up for early childhood development. *The Lancet* 389:103–118.

Romano, J. P. and Wolf, M. (2005a). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association* 100:94–106.

Romano, J. P. and Wolf, M. (2005b). Stepwise multiple testing as formalized data snooping. *Econometrica* 73:1237 – 1282.

Rubio-Codina, M., Attanasio, O. P., Meghir, C., Varela, N., and Grantham-McGregor, S. (2015). The socioeconomic gradient of child development: Cross-sectional evidence from children 6–42 months in bogota. *Journal of Human Resources* 50(2):464–483.

Rubio-Codina, M. and Grantham-McGregor, S. (2020). Predictive validity in middle childhood of short tests of early childhood development used in large scale studies compared to the bayley-iii, the family care indicators, height-for-age, and stunting: A longitudinal study in bogota, colombia. *PloS one* 15(4):e0231317.

Sah, R. K., Panda, P., Agrawal, S., Tripathy, S., Negi, S., and Mehta, V. (2024). Prevalence and factors contributing to stunted growth in young children of khordha district, odisha: A cross-sectional study. *Journal of Family Medicine and Primary Care* 13(7):2746–2752.

Schrank, F. A., Mather, N. and McGrew, K. S. (2014). *Woodcock-Johnson IV tests of achievement*. Riverside Publ.

Smith, J. A., Chang, S. M., Lopez Boo, F., Ferro, M. d. l. P., and Walker, S. P. (2021). Are

benefits from a parenting intervention delivered through the health services sustainable? follow-up of a randomized evaluation in jamaica. *Academic Pediatrics* 21(4):638–645. Epub ahead of print: 09 Jan 2021.

Snilstveit, B., Gallagher, E., Phillips, D., Vojtkova, M., Eyers, J., Skaldiou, D., Stevenson, J., Bhavsar, A., and Davies, P. (2017). Protocol: Interventions for improving learning outcomes and access to education in low- and middle-income countries: a systematic review. *Campbell Systematic Reviews* 13(1):1–82.

Squires, J. and Bricker, D. (2009). Baltimore: Paul H. Brookes Publishing Co. Inc.

Tomlinson, M., Skeen, S., Melendez-Torres, G. J., Hunt, X., Desmond, C., Morgan, B., Murray, L., Cooper, P. J., Rathod, S. D., Marlow, M., Fearon, P., et al. (2021). First 1,000 days: enough for mothers but not for children? long-term outcomes of an early intervention on maternal depressed mood and child cognitive development: follow-up of a randomised controlled trial. *Journal of Child Psychology and Psychiatry* .

UNESCO Institute for Statistics (2025). Unesco data api. `https://databrowser.uis.unesco.org/resources`. United Nations Educational, Scientific and Cultural Organization (UNESCO).

Vagh, S. (2010)). Validating the aser testing tools: Comparisons with reading fluency measures and the read india measures. Technical report.

Wechsler, D. (2012). *Wechsler preschool and primary scale of intelligence—fourth edition*. The Psychological Corporation San Antonio, TX.

Online Appendix

Power Calculations for the original trial Assuming an attrition rate of 10.7% (as at first follow up) and an intracluster correlation coefficient equal to 0.04, as in similar studies, we calculated the minimum detectable effect sizes (MDE) of our study, without accounting for efficiency gains from controlling for covariates, at 80% power and significance $\alpha = 0.05$. Our design had a MDE of 0.28 SD for testing the difference in means in the treatment group against the control group.

Table A.1: Baseline Balance for Non-attriters

	Treatment (T)		Control (C)		p-value: Treatment = Control
	Mean	S.D.	Mean	S.D.	
Age in months	14.934	3.049	15.115	3.113	0.589
Male %	0.579	0.495	0.457	0.500	0.053
Firstborn %	0.461	0.500	0.438	0.498	0.671
Mother's years of education	7.908	3.240	6.574	3.748	0.005
Asset index Z-score	0.131	0.910	-0.125	0.914	0.057
ASQ-3 problem solving Z-score	0.020	1.026	0.044	0.946	0.859
ASQ-3 communication Z-score	0.019	1.008	0.092	0.875	0.499
ASQ-3 fine motor Z-score	0.059	0.991	0.079	0.892	0.867
Stunted	0.200	0.401	0.261	0.440	0.180
Maternal knowledge of child development Z-score	-0.020	0.816	-0.032	0.863	0.928
Quality of home environment Z-score	0.084	0.914	-0.034	0.740	0.172
Maternal depressive symptoms Z-score	-0.030	0.874	-0.002	0.792	0.779
Below Urban Poverty Line	0.546	0.500	0.475	0.501	0.345
Income (Rs) per Capita per Day	98.865	192.223	103.511	234.789	0.877
Roof made from metal sheet/thatch/polyethylene	0.414	0.494	0.494	0.502	0.378
House has dirt floor	0.066	0.249	0.068	0.252	0.958
House has piped water connection	0.533	0.501	0.574	0.496	0.661
Household has electricity connection	0.980	0.140	0.988	0.111	0.585
Household owns a fridge	0.395	0.490	0.286	0.453	0.107
Observations	152		162		

Notes: The table reports balance in baseline characteristics among non-attriters between the treatment group and the control group. p-values are for tests of equality of the means across treatment and control groups. Z-scores have a mean 0 and standard deviation of 1 in the control group. [a] The Urban poverty line as defined by the Rangarajan committee is Rs. 47 per household member per day. [b] The average exchange rate during the baseline survey (November/December 2013) was Rs. 62/USD.

Table A.2: Attrition

	Follow-up 1 and 2	Follow-up 1	Follow-up 2
Treatment	0.025	-0.032	0.037
	(0.022)	(0.027)	(0.050)
Constant	0.047***	0.118***	0.236***
	(0.017)	(0.021)	(0.034)
Observations	421	421	421

Notes: The table shows attrition by treatment status. In column 1 we report the results of a regression of being lost at the first and second follow-up on treatment status, the other two columns look at attrition separately at the first and second follow-ups. Standard errors in parentheses.

Table A.3: Baseline Balance between Attriters and Non-attriters

	Attriters		Non-attriters		p-value: Attriters =
	Means	S.D.	Mean	S.D.	Non-attriters
Age in months	14.595	3.361	15.028	3.079	0.198
Male %	0.533	0.501	0.516	0.501	0.724
Firstborn %	0.542	0.501	0.449	0.498	0.088
Mother's years of education	7.935	3.924	7.220	3.569	0.156
Asset index Z-score	0.014	0.988	0.000	0.919	0.913
ASQ-3 problem solving Z-score	-0.095	1.021	0.032	0.984	0.311
ASQ-3 communication Z-score	-0.167	1.124	0.057	0.941	0.039
ASQ-3 fine motor Z-score	-0.202	1.119	0.069	0.940	0.041
Quality of home environment Z-score	-0.068	0.811	0.023	0.830	0.253
Maternal depressive symptoms Z-score	0.045	0.967	-0.015	0.831	0.580
Below Urban Poverty Line	0.411	0.494	0.510	0.501	0.136
Income (Rs) per Capita per Day	112.305	227.646	101.262	214.910	0.691
Observations	107		314		

Notes: The table reports balance in baseline characteristics between attiters and non-attriters. p-values are for tests of equality of the means across the two groups. Z-scores have a mean 0 and standard deviation of 1 in the control group. [a] The Urban poverty line as defined by the Rangarajan committee is Rs. 47 per household member per day. [b] The average exchange rate during the baseline survey (November/December 2013) was Rs. 62/USD.