NBER WORKING PAPER SERIES

EARLY CHILDHOOD INTERVENTION FOR THE POOR: LONG TERM OUTCOMES

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

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

Authors' names in alphabetical order. 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 original trial was funded by a personal donation from Mr. Rushton Turner, by a grant from the Waterloo Foundation and by the National Institutes of Health, USA (Grant R01 HD 72120). The present follow-up study was funded by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant 695300 - HKADeC - ERC-2015-AdG/ ERC-2015-AdG) and the Economic and Social Research Council (ESRC)(Grant ES/ M010147/1). 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). 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.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Alison Andrew, Orazio Attanasio, Britta Augsburg, Lina Cardona-Sosa, Monimalika Day, Michele Giannola, Sally Grantham-McGregor, Pamela Jervis, Costas Meghir, and Marta Rubio-Codina. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

ABSTRACT

Early childhood interventions aim to promote skill acquisition and poverty reduction. While their short-term success is well established, research on longer-term effectiveness is scarce, particularly in LDCs. We present results of a randomized scalable intervention in India, that affected developmental outcomes in the short-term, including cognition (0.36 SD p=0.005), receptive language (0.26 SD p=0.03) and expressive language (0.21 SD p=0.03). After 4.5 years, when the children were on average 7.5 years old, IQ was no longer affected, but impacts persisted relative to the control group in numeracy (0.330 SD, p=0.007) and literacy (0.272 SD, p=0.064) driven by the most disadvantaged.

Alison Andrew
Department of Economics
Oxford University
alison.andrew@economics.ox.ac.uk

Orazio Attanasio
Department of Economics
Yale University
87 Trumbull Street
New Haven, CT 06511
and FAIR, BREAD and CEPR
and also NBER
orazio.attanasio@yale.edu

Britta Augsburg
The Institute for Fiscal Studies
7 Ridgmount Street
London WC1E 7AE
britta_a@ifs.org.uk

Lina Cardona-Sosa
World Bank
1818 H St NW
Washington, DC 20433
lcardonasosa@worldbank.org

Monimalika Day
School of Education Studies
Dr. B.R Ambedkar University Delhi
Lodhi Road Campus
Aliganj, B.K Dutt Colony
New Delhi, 110003
India
monimalika@aud.ac.in

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

Sally Grantham-McGregor
University College London
Gower Street
London - WC1E 6BT
sallymcgregor@yahoo.com

Pamela Jervis
Department of Industrial Engineering
University of Chile
Beauchef 851
Santiago, Chile
and Institute for Fiscal Studies
pjervisr@uchile.cl

Costas Meghir
Department of Economics
Yale University
87 Trumbull Street
New Haven, CT 06511
and IZA
and also NBER
c.meghir@yale.edu

Marta Rubio-Codina
Inter-American Development Bank
Social Protection and Health Division
Washington, DC 20577
martarubio@iadb.org

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

1 Introduction

Poverty is transmitted intergenerationally, and a key channel is the developmental deficits accumulated by children growing up in deprived environments. The cognitive development gap between children from poor and middle-class families increases steadily with age (Rubio-Codina et al. (2015), Rubio-Codina and Grantham-McGregor (2020)) leading to poor school performance, limited skill acquisition, and consequently the perpetuation of poverty. Since the first *Lancet Series on Child Development* (Grantham-McGregor et al., 2007), which highlights the vast number of children that do not meet their developmental potential in low and middle-income countries (LMICs), there has been a growing interest in designing scalable interventions that can help break this cycle (Richter et al., 2017). However, while evidence on the effectiveness of early child development (ECD) interventions is growing, even when implemented at scale, little is known about their impacts on child outcomes in the longer-term, and which children benefit the most. A notable exception is a parenting intervention developed and evaluated in Jamaica (Grantham-McGregor et al., 1991), that led to strong short-term and long-term impacts for particularly vulnerable children (Gertler et al., 2014). Beyond this, few such programs are followed up in the longer term, (Jeong et al., 2021b) particularly in LMICs, including those with larger sample sizes and scalability in mind.

In this paper, we partly fill this gap by presenting medium run impacts of an ECD intervention focusing on parenting, and implemented in the slums of Cuttack, a mid-sized Indian city in the eastern state of Odisha. The intervention we evaluate was designed to be *scalable*, both in terms of cost and the approach to delivery, which is based on local community workers. The slum residents are an important population to consider. They were typically migrants from rural areas with limited local support networks and are extremely poor. However, as baseline data confirmed, they had high aspirations for their children. We can thus expect them to be open to interventions to improve human development for their

children, which might increase their motivation to engage with the program.

Our focus is on ECD because it is well established that deficits in child development, once they have set in, are hard to reverse and thus play a key role in the intergenerational transmission of poverty. Millions of children around the world are at risk of not achieving their potential because of their exposure to poverty and deprivation, making it a major social and public health problem (Bhutta et al., 2022). Hence, successful early interventions in childhood have the potential of becoming central elements of policies promoting development and fighting poverty.

More specifically, disadvantage in the first three years of life affects the development of children's brain and endocrine system (Jensen et al., 2021). Delays in development, which become apparent from the first year, are associated with poor school achievement and adult functioning, contributing to the intergenerational transmission of poverty. While these issues are relevant wherever poverty is present, they are particularly salient in LMICs (Black et al. (2016); Hamadani et al. (2014)). The consequences of the early delays in development have therefore important consequences for individual well-being over the life cycle and, more generally, for economic development.

To provide solutions to this problem we implemented an adaptation of the home-visiting program Reach-Up, a parenting intervention originally developed for the Jamaica study, which focuses on caregiver-child interactions and on the home environment.[1] In most societies, very young children spend far more time with their mother at home, than with anyone else. For this reason, children's home environment – which we take to include both caregiver-child interactions and the material environment, such as access to play materials – is paramount in shaping children's development during these earliest years (Attanasio et al., 2020a, 2022). However, poverty, beyond limiting material resources that can be used for child learning and stimulation, can also cause stress and anxiety for parents, making it hard

[1]See https://reachupandlearn.com/

for them to prioritize creating a stimulating and nurturing environment (Engle and Black, 2008). Very low levels of education may also limit parents' understanding of age-appropriate parenting and its role in the process of child development. For mothers living in poverty in India, all of these challenges are likely to be exacerbated by a high degree of social isolation limiting the support they can draw on (Kandpal and Baylis, 2019; Andrew et al., 2023), by highly unequal gendered distribution of power and resources (Calvi, 2020) and by the fact that many women become mothers for the first time at a young age.

Over the short-term, many programs have been found to be effective at improving the quality of adult-child interaction, increasing parents' investments and improving child development (see, for instance, Attanasio et al. (2022)). However, there is little – and mixed – evidence on whether such impacts are sustained over the longer-term.[2] Both in LMICs and higher income countries, in the few studies that have been followed up over the long-term (Jeong et al., 2021b), it has often been observed that measured impacts fade out partially or completely in the medium-term only to reappear in adult outcomes later on (see Bailey et al. (2017) for an example in high-income countries). This pattern is observed in the Jamaica experiment cited above, as well as in other well-conducted trials, such as those described in Heckman et al. (2013) and Elango et al. (2016). This evidence suggests that following up with children who participated in such studies over a substantial part of their life-cycle is key for a deeper understanding of how early interventions can benefit people and how sustainable there impacts are.

Understanding whether impacts persist is crucial for determining whether particular interventions can make a significant difference to the opportunities of underprivileged children. On the one hand, there is evidence that the process of child development is such that later investments are complements with the stock of accumulated human capital, which would suggest the possibility of sustained and lasting impacts (as pointed out, for instance

[2]For example, in an ECD intervention in Colombia (Attanasio et al., 2014) showed substantial impacts on child development and parental investments at endline. However, Andrew et al. (2018) found that these impacts had disappeared after two years.

by Heckman et al. (2013); Garcia et al. (2021); Attanasio et al. (2020b)). On the other hand, it is possible that the short-term impacts are influenced by factors that may not be permanent. For instance, if an intervention exposes children to the types of tasks that are subsequently used to measure their development, certain measured impacts could be reflecting "teaching to the test" and be only of short-run duration (see Heckman et al., 2020). Moreover, parental behavioral changes, which seem key to mediating the effects of interventions (Attanasio et al., 2020a), may not be sustained or may no longer be age appropriate as children grow.

Evidence that early years interventions have long-lasting effects is more abundant in high income countries (see for example Campbell et al. (2014); Heckman et al. (2010); Doyle (2020)), but much less so in developing countries, where the most important follow up is from Jamaica (Gertler et al., 2014). There are good reasons to believe that long-term sustainability of impacts may be very different in environments with fundamentally different institutions, such as quality of schools and welfare programs and also very different initial conditions. Hence, our paper helps to fill an important gap in this literature by considering the medium term impacts of a scalable intervention among some of the most deprived people in India, a low income country setting.

The target of the intervention we study were mothers and their infant children living in slums around Cuttack. The program, which we describe in detail below, lasted eighteen months and aimed to help mothers improve the quality of interactions with the young child and more broadly the home environment. In previous work (Andrew et al., 2019), we found that this program led to substantial short-run improvements in children's cognitive and language skills, increased parental investments in their children and improved maternal mental health. In this paper, we study whether the impacts observed in the short run were sustained when children were 7-8 years old, four and a half years later, between December 2019 and January 2020. We find that children had sustained benefits in overall achievement, in

numeracy and in literacy, which in turn are likely to feed into better school performance and stronger accumulation of skills, with all the economic benefits that this will entail. IQ was not significantly improved. We show that impacts are driven almost exclusively by children that were particularly vulnerable at baseline, as measured by their height-for-age at baseline. The impacts for this group of children helps them to catch-up with their less vulnerable peers.

This medium-term follow-up occurred at a particularly critical age, as it is close to the beginning of children's schooling. In the context of hundreds of millions of children in LMICs leaving school without basic numeracy and literacy skills (Snilstveit et al., 2017) and evidence that children's skills at school entry are important determinants of their later school attainment (Heckman, 2006), early-years programs that have sustained positive impacts up until the entry into formal schooling appear particularly promising.

In the Indian context, schooling without learning has been an important concern as indicated by various surveys of school children (ASER, 2023). A longitudinal study of 14000 preschool children in three states of India found that children had low scores in cognition and language at age 5 years which was associated with their performance in the primary grades (Kaul and Bhattacharjea, 2019). This motivates the need to ensure improved developmental outcomes that are expected to promote school readiness and future successful skill accumulation, with the expectation that this will provide an escape route from poverty.

In what follows we describe the intervention and the experimental design (Section 2). We then present the Follow up study (Section 3) and the results (Section 4). We end with a short discussion and concluding remarks on the broader implications (Section 5).

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 took place 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 by us to the socio-cultural context of Odisha.

The intervention, implemented in collaboration with Pratham[3], consisted of weekly one-hour home visits with the target child and primary caregiver over an eighteen-months period. 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 (Grantham-McGregor et al., 1991).[4]

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 two-to-three days refresher training. 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 child and caregivers and introducing play activities, picture books, images to stimulate story-telling and conversation, and homemade toys from household surplus materials, all involving

[3]https://www.pratham.org/about/

[4]Countries in which Reach Up has been adapted include Bangladesh, Brazil, China, Colombia, India, 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. (2020a) and Jeong et al. (2021a) for the importance of encouraging parenting practices to support the socio-emotional development of infants and young children.

close to no financial cost. The HVs left the toys in the home until the next visit, and encouraged caregivers to continue play and chat 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 household at the end of the intervention.

The local women that delivered the intervention were between 18 and 55 years old, and 74.1% had completed high school. Three mentors with field experience were trained in key principles of child development to train, mentor and assist the HVs through weekly meetings. These mentors were employed by our project partner, Pratham, India's largest educational NGO, which ran and managed the delivery of the intervention.[5] A psychologist supported the intervention activities throughout. Over the 18-months period, the cost per child in the treatment group was US$251, which is comparable to the cost reported in Grantham-McGregor et al. (2020).[6]

In addition to the cost being relatively low, 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, from which between 7-9 children per cluster were randomly selected for inclusion in the study. Consent was obtained before treatment status was revealed. This

[5]Pratham is the major NGO in India on educational matters. They were our partner in this and other experiments in India.

[6]Indian GDP per capita in 2012 was US$1434.

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 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.

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, we measured child development 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). At endline, after 18 months of treatment, children's cognition had benefited by 0.357 SD (p-value =0.005), receptive language by 0.256 SD (p-value=0.034) and expressive language by 0.209 SD (p-value=0.033).[7] 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). Also children who were moderately stunted at baseline (height-for-age <-2 standard standard deviations of WHO standard) improved significantly more than non-stunted children (the effect size on the Bayley was 0.794 (standard error 0.266) versus 0.158

[7]All these p-values have been adjusted for multiple testing based on the Romano-Wolf step-down procedure (Romano and Wolf, 2005b). There were four hypotheses being tested, the fourth one being fine-motor skills which showed no significant improvement (0.11 SD, p-value=0.36).

(standard error 0.140), p-value of the difference = 0.19).

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 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

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 answers by parents.

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 communication score 0.039. The remaining 10 p-values range from 0.088 to 0.913.[8]

3.2 Outcome measures

The key outcomes that the intervention targets are numeracy, language development and cognitive development. Now we describe the nature of these 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 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 et al. (2014)). The task measures the ability to solve simple 60 addition and subtraction problems in a set time of 3 minutes. For

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

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 intervention (Meghir et al. (2023))[9]. All sub-scales were combined into the Full-Scale-IQ (FSIQ).

All tests were administered in the children's homes. Testing was conducted by 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

[9]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.

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' 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.[10]

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.

[10]See the description of the procedure in section B.4 of the online appendix of Attanasio et al. (2020a) 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.

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.[11] 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 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

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

numeracy and literacy summary indices, which form part of the overall achievement index. The impact on numeracy is 0.33 SD and strongly significant. The impact on literacy, is 0.272 SD and significant at the 6% level.

In Panel C, we further break down the impacts on individual components of both the numeracy and literacy summary indices. The math facts fluency test is significant at the 3% level, while the early numeracy scale is significant at 5.4%. Again the two literacy tests are less significant, especially the early literacy scale.

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 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 which are directly relevant to the intervention stand out: parents were more likely to report doing stimulating activities with their child (FCI and HOME activities, p-value=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.

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 behind the intervention's impacts on child development and investments, consistent with Baranov et al. (2020).

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).

Table 3: Treatment Effects on Intermediate Outcomes

Panel A:	Maternal depression (Z-score)	Maternal Self-efficacy (Z-score)	Materials (Z-score)	Activities (Z-score)	Educational expenditure
	(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).

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. (2020a) 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 (weight-for-age Z score of –2 SDs of the WHO standard). In our study, we did not screen for malnutrition, but rather selected 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,

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: The table reports impact heterogeneity by stunting status at baseline. Controls include child gender, age, baseline ASQ scores, maternal education, and randomization strata fixed effects. 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).

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.[12]

The heterogeneity analysis is presented in Table 4 and shows that the effect sizes at age 7-8 are much larger for children who were stunted at baseline with p-values for the difference ranging from 8.4% overall to 9.4% for numeracy and 15% for literacy. This shows that the differences we noted at endline have persisted. This result is important for various reasons. First, it shows that the intervention can help the most deprived children to catch up. However, the other side of the same coin is that we were not able to improve as much the children that were slightly better off, albeit still slum dwellers. Therefore, 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 the stunted children we obtain results similar to those of the original Jamaica intervention both at endline and at mid-term follow-up. Hence, our evidence reinforces the message that ECD interventions for the extremely

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

deprived can have life-changing impacts, and the Jamaica experimental results do replicate both at endline and at follow-up, with an intervention that was designed to be scalable. This evidence constitutes some of the strongest results on *scalable* ECD interventions in developing countries to date. However, it is important that this result is followed up in future experiments designed for power in that comparison.

Finally, as discussed in section 2, the endline results showed impact heterogeneity by mother's education and (suggestively) also by child gender. In this long-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 demonstrate a remarkable persistence of the benefits four and a half years after the intervention ended. 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 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 improvement in the home environment and the extra money parents spent on education suggest that the parents were also affected and provided more nurturing homes and better education which may be part of the mechanism. And returning to the original motivation for ECD, the improved nurturing may have permanent neurological effects on the brain (Jensen et al., 2021).

The results are also important because there can be no suggestion that the benefits were taught as part of the intervention. The children were too young at the end of intervention (28 to 38 months old) to be taught literacy and numeracy, on which we report sizable 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 minimal resources based on locally available women, without any specific qualifications, trained for this intervention just for a few weeks. This demonstrates the potential for policy to dramatically improve child outcomes without insuperable scaling-up constraints.

References

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

Andrew, A., Attanasio, O., 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., 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* .

Andrew, A., Attanasio, O., 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.

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

Attanasio, O., Cattan, S., and Meghir, C. (2022). Early childhood development, human capital, and poverty. *Annual Review of Economics* 14(1):853–892.

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

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.

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.

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.

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).

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.

Elango, S., García, J., Heckman, J., and Hojman, A. (2016). Early childhood education. In R. A. Moffitt, ed., *Economics of Means-Tested Transfer Programs in the United States*, volume 2, chapter 4. University of Chicago Press, Chicago, IL.

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., 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.

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. (2021a). 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. (2021b). 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).

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.

Meghir, C., Attanasio, O., 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., 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.

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

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.

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.