Steven J. Green School of International & Public Affairs — Department of Economics

Early Childhood Human Capital Formation at Scale

Johannes M. Bos

Akib Khan

Saravana Ravindran

Abu Shonchoy

Department of Economics

Florida International University

Working Paper 2115

First Version: August 2021

Current Version: April 2022

11200 SW 8th Street, Miami, Florida 33199

https://economics.fiu.edu/

Early Childhood Human Capital Formation at Scale*

Johannes M. Bos, Akib Khan, Saravana Ravindran and Abu Shonchoy†

April 8, 2022

Abstract

Can governments leverage existing service-delivery platforms to scale early childhood development (ECD) programs? We experimentally study a large-scale home-visiting intervention providing materials and counseling — integrated into Bangladesh's national nutrition program without extra financial incentives for the service providers (SPs). We find SPs partially substituted away from nutritional to ECD counseling. Intent-to-treat estimates show the program improved child's cognitive (0.17 SD), language (0.23 SD), and socio-emotional developments (0.12-0.14 SD). Wasting and underweight rates also declined. Improved maternal agency, complementary parental investments, and higher take-up of the pre-existing nutrition program were important mechanisms. We estimate a sizeable internal rate-of-return of 19.6%.

JEL codes: J13, J24, I25, H11.

Keywords: Early childhood development, Human capital formation, Bangladesh

*We are grateful to the World Bank for funding the research through its Strategic Impact Evaluation Fund (SIEF) and hosting the project data on its web site. Ravindran is funded by a startup grant at the Lee Kuan Yew School of Public Policy, National University of Singapore. We also thank AIR's partners in Bangladesh, including Save the Children, Data International, the BRAC University, and the International Centre for Diarrhoeal Disease Research. We thank all AIR staff who contributed to this project including Marjorie Chinen (Project Director), Julia Lane (co-PI), Rosa Castro-Zarzur, and Matthew Murray. We are also grateful to (alphabitically) Douglas Almond, Niklas Bengtsson, Gabriella Fleischman, Esther Gehrke, Reshmaan Hussam, Daniel Maggio, Laia Navarro-Sola, Yoko Okuyama, Jonathan De Quidt, Abhijeet Singh, Anna Tompsett, and seminar participants at the CSAE Conference, ASSA Annual Meeting, NEUDC Conference, SEA Annual Conference, ASWEDE Conference, EUDN Conference, Japan Empirical Economics Seminar (JEES), National University of Singapore - Singapore Management University Development Economics Workshop, Korea University Applied Microeconomics Conference, Nordic Conference in Development Economics, Uppsala University, and the Lahore University of Management Sciences (LUMS) for their insightful comments. Most of all, we are grateful to the NNS workers and the families they served for allowing us to observe and document their work and collect survey and anthropometric data from them in several survey waves. This research could not have been done without them. This study was approved by AIR IRB under IRB00000436 and is registered in the AEA RCT Registry with the identifying number AEARCTR-0008001. All views and any errors are our own.

†Bos: American Institutes for Research (AIR), USA, jbos@air.org, Khan: Uppsala University, Sweden, akib.khan@nek.uu.se, Ravindran: Lee Kuan Yew School of Public Policy, National University of Singapore, saravana@nus.edu.sg, Shonchoy: Florida International University, USA, shonchoy@fiu.edu.

1 Introduction

An important policy challenge for governments of low- and middle-income countries (LMICs) is to find sustainable and effective ways to implement human capital formation programs — at scale — while minimizing potential unintended consequences on outcomes that are not directly targeted. Delivery of large-scale quality early childhood development (ECD) interventions, however, imposes demanding infrastructure and personnel requirements to reach all parents and children in both urban and rural areas. One practical solution to this issue is to integrate ECD programs into the existing public service infrastructure (Richter et al., 2017). While this can help reap economies of scale and reduce per-child costs of program provision, challenges relating to delivery of the program (such as distribution of program materials) can dilute program impacts. Such decreases in effect size (or "voltage drops,") when scaling up have been documented in other settings (List et al. 2021, List 2022). Furthermore, service providers may reallocate time away from their regular duties and towards the additional tasks assigned to them, potentially worsening child development outcomes in areas that the intervention does not target directly. This can be particularly salient when these providers are not offered compensation for the additional duties.[1]

We study at-scale service delivery in the context of ECD programs. A growing body of research highlights the importance of ECD in human capital formation (Knudsen et al. 2006, Almond and Currie 2011, Black et al. 2017). Evidence from targeted early childhood interventions with limited samples administered in Jamaica (Grantham-McGregor et al. 1991, Gertler et al. 2014) and the U.S. (Campbell and Ramey 1994, Heckman et al. 2010) suggests sizable short- and long-run improvements in a range of welfare and labor market outcomes (Walker et al. 2011, Gertler et al. 2014, Campbell et al. 2014). Early life investments are especially important for LMICs, since missing investments in early childhood can lead to lower human capital accumulation and intergenerational transmission of poverty. Lu et al. (2016) estimate that 250 million or 43% of under-five children in LMICs suffer from sub-optimal physical and mental growth, limiting their learning and earning possibilities.

Can governments in developing countries leverage existing service-delivery platforms to scale ECD interventions? We address this question using a unique randomized controlled trial (RCT) of an early childhood stimulation program in Bangladesh. Three key features of the intervention were (1) it was delivered at scale, targeting more than 18,000

[1] For the remainder of the paper, we use the terms 'service providers' and 'providers' interchangeably.

children in three out of seven divisions in Bangladesh, (2) it was integrated into the government's flagship National Nutrition Services (NNS) program and used existing NNS staff to deliver the intervention instead of deploying a separate group of ECD workers, and (3) it did not provide any additional financial incentives to service providers. Our results show that ECD programs delivered using the existing public sector infrastructure, though partially compliant with significant time reallocation by providers, can still achieve sizable improvements in a range of child development outcomes.

Bangladesh offers an ideal setting to study large-scale ECD interventions. The need for ECD interventions is particularly strong in the country, where 42% of children aged five and below were stunted around the time of this study and only 13.4% of 3-5 years old children received any early childhood education (Bangladesh Bureau of Statistics and United Nations Children's Fund, 2015).[2] Furthermore, the government runs the flagship NNS program that promotes best nutritional practices through micro-nutrient supplements, growth monitoring, counseling support, and monthly home-visits by community health workers. This vast NNS public health infrastructure presents an opportunity to integrate an ECD program and deliver the intervention at scale. The intervention we study was implemented by Save the Children, an international humanitarian aid organization for children.

The intervention targeted children aged 3-18 months at baseline and was implemented in 78 community clinics across three districts located in three out of seven divisions in Bangladesh. Half of these community clinics and their catchment areas were randomly assigned to the treatment group which received both NNS and ECD programs, while the other half only received the pre-existing government NNS program. The ECD program was a bundled intervention comprising the following components: (i) training of NNS workers on ECD, (ii) provision of four types of program materials to parents, including a child development card containing information on age-appropriate stimulation practices, two picture books, and a booklet with key messages and clinic visit guidelines, and (iii) regular ECD counseling during routine monthly visits by NNS workers. The intervention was implemented over 15 months from June 2014–August 2015.

We find that the training led to higher perceived importance of ECD among service providers in the treatment communities. However, treated NNS workers did not increase the number of home visits nor their duration. Instead, they reallocated the time spent

[2] A child is defined to be stunted if his/her height-for-age is less than two standard deviations below the median height-for-age as per WHO growth standards (Bangladesh Bureau of Statistics and United Nations Children's Fund, 2015).

with parents by partially substituting away from discussions of child health and nutrition to spend more time on ECD. The intervention also faced scale-up challenges owing to compliance and delivery. For example, only half of the households in the treatment group received any three of the four program materials. We find that compliance is strongly negatively correlated with distance, or time to the nearest clinic.

Despite these challenges, our results show that the ECD program delivered through the existing public delivery platform led to sizable improvements across a range of child development outcomes. Intent-to-treat estimates using endline surveys conducted 1-3 months after the end of all intervention activities for 15 months show improvements in child outcomes along dimensions of cognitive (0.17 SD), language (0.23 SD), and socio-emotional development (0.12-0.14 SD). Moreover, we find evidence of positive intra-household spillovers: immediately older siblings (aged 3-5 years at baseline) were more likely to be attending school in the treatment communities at endline. The program was highly cost-effective at a cost of 6.84 USD (2014 values) per targeted child. Driven by economies of scale and sizeable impacts, we estimate an internal rate-of-return to the program of 19.6%.

We find that the reallocation of service-provider time away from child nutrition and towards ECS did not lead to adverse effects. We find better nutritional outcomes, with the intervention reducing wasting and underweight rates by 4.8 and 2.9 percentage points, respectively. This is consistent with models of multiple investment inputs in the production function for skills formation, where inputs exhibit diminishing marginal returns (see, for example, Cunha et al. 2010). We also find evidence of complementary parental investments, with parents in the treatment group providing a significantly greater variety of play materials and learning activities for their child. Improvements in the latter are driven exclusively by mothers. The intervention led to greater involvement of mothers in household decision-making, especially in the domains of child health and stimulation. The intervention also induced higher take-up of the pre-existing NNS program whose staff delivered the early childhood simulation intervention. Parents in the treatment communities were more likely to have and show the NNS growth development card (during the endline survey) and also took their children for a greater number of NNS growth monitoring check-ups.

A mediation analysis suggests that the complementary parental investments played a major role in explaining the treatment impacts. Our unique service-provider survey also enables investigating treatment-effect heterogeneity by service-provider characteristics:

we find larger gains for children in communities where service providers were more educated but less experienced (and thereby possibly more open to changing their interactions with parents).

We contribute to the literature in at least three ways. First, we study an integrated model of delivering ECD interventions at scale *without* providing any extra financial incentives for the incumbent workers. This is a key policy consideration for resource-constrained governments in developing countries. Attanasio et al. (2022) and Sylvia et al. (2020), to our knowledge, are the only other studies that consider a similar intervention design. Attanasio et al. (2022) study a structured early stimulation curriculum introduced to the Family, Women, and Childhood program (FAMI), an existing government program in Colombia. Sylvia et al. (2020) study an ECD program in rural China delivered by the Family Planning Commission cadres. Both studies note significant improvements in child's cognitive skills. Our paper contributes to these findings by investigating a lighter-touch intervention with laxer enforcement protocols that could be easier to replicate in other settings.[3]

Second, we contribute to the broader evidence base on community health-provider behavior and productivity in low-income countries (Deserranno 2019, Nyqvist et al. 2019, Ashraf et al. 2020). In addition to households, we surveyed the service providers to understand their response, focusing on beliefs and behaviors such as the reallocation of time spent with parents. This allows us to study possible unintended consequences relating to the crowding-out of tasks that were not targeted by the intervention. We also study impact heterogeneity by service provider characteristics, which could inform policy decisions pertaining to the hiring and placement of service providers when scaling up such interventions.

Third, we contribute to the growing literature studying at-scale delivery of ECD interventions in developing countries. Our experiment conducts and evaluates an ECD intervention at a scale larger than many of the existing studies (Attanasio et al. 2014, Sylvia et al. 2020, Yousafzai et al. 2014).[4] For example, Araujo et al. (2021) study a home-visiting

[3] For example, Attanasio et al. (2022) use an intervention that included weekly group sessions for an average of 10 months, and a monthly nutritional supplement.

[4] Attanasio et al. (2014) study an ECD program delivered using paid local female volunteers to marginalized families who were beneficiaries of an ongoing conditional cash transfer program in Colombia. Yousafzai et al. (2014) study a community-based ECD program in Pakistan that provided a salary increase to the "Lady Health Workers" of the existing national family planning and primary health care program. Attanasio et al. (2014): N = 1,420, Central Colombia; Sylvia et al. (2020): N = 592, one province in rural Shaanxi, China; Yousafzai et al. (2014): N = 1,489, one district in Sindh, Pakistan.

intervention implemented by paid para-professionals in Peru that targeted over 67,000 children in rural areas. We build on this work by considering a poorer LMIC setting and an intervention that did not hire additional service providers. The program was also scaled simultaneously in three divisions in Bangladesh, strengthening the generalizability and external validity of the results.[5]

The rest of the paper is organized as follows. Section 2 provides more details on the background and experimental design. Section 3 presents descriptive statistics from the data and outlines the empirical strategy. Section 4 presents treatment effects on service providers, households, and children. Section 5 discusses compliance and the dose-response relationship, while Section 6 investigates potential mechanisms. Section 7 presents a cost-benefit analysis of the intervention and Section 8 concludes.

2 Experimental Context, Sampling, & Design

2.1 Program Description

Our at-scale ECD program was designed and managed by Save the Children and integrated into the existing National Nutrition Services (NNS) in Bangladesh.[6] The program targeted children aged 0-3 years and had two components: educational materials for the household and counseling of mothers or caregivers by the NNS workers. The materials are: the Child Development Card reporting age-specific recommendations for play and communication, two picture books illustrating objects from household and nature, and the Key Message Booklet summarizing the information shared via counseling. The Child Development Card highlighted the importance of everyday play and communication for child development and offered, via easily understandable visual aids, age-appropriate ideas to exercise them (please see Figures 1-2 in Appendix A). These aids were included

[5] See Table 16 for a summary of studies. More broadly, our study speaks to the growing body of work exploring different determinants of early childhood development in low-income countries (Currie and Vogl, 2013). In Bangladesh, Barham (2012) studies the Matlab Maternal and Child Health and Family Planning Program, and finds significant positive impacts on cognitive functioning, height, and educational attainment among children aged 8-14. Joshi and Schultz (2013) also study the Matlab program and document lower child mortality in areas that received the program. Hamadani et al. (2014) study participants of a pregnancy supplementation trial in Bangladesh and find that the effects of poverty on children's cognition are mostly mediated through parental education, size at birth, growth in the first 24 months, and home stimulation in the first 5 years.

[6] The NNS was designed to enhance access to nutritional services and counseling for pregnant women and lactating mothers and children under the age of five. For instance, through its network of community health workers, the program distributes micro-nutrient supplements and deworming medication, organizes growth monitoring sessions and malnutrition screening, and offers referral services for maternal and child malnutrition.

so that mothers from low-literacy communities could understand and implement the recommendations. In addition, these ECD materials were distributed to help mothers use the card as a picture book, potentially showing illustrations from the card to their children. For instance, the books illustrated different games that mothers could play with their child, potential play materials they could offer, and ways to respond to the child's cues. The picture books were designed to help mothers teach their children new words and provide topics for communication and play. These books contained names and illustrations of 30 regular objects present in households (e.g., window, glass, and chair) or in nature (e.g., tree, cow, and dog) (please see Figures 3-4 in Appendix A). Lastly, the Key Message Booklet recorded the main suggestions to be conveyed during the counseling sessions. These suggestions spanned the following broad themes: love and affection, play and games, talk and communication, positive discipline, care during pregnancy, responsive feeding, health and hygiene, and sharing knowledge with others in and outside the household (please see Figure 5 in Appendix A).

The NNS workers were expected to offer early stimulation counseling and distribute the aforementioned materials during their monthly household visits, as well as during parental visits to the community clinics and other related events such as the Expanded Program of Immunization (EPI) and Growth Monitoring Campaigns. During these visits, the program intended the NNS workers to show mothers or primary caregivers how to use the Child Development Card and the two picture books to create a stimulating environment at home. Save the Children developed home visit and clinic visit guidelines for the NNS workers (please see Figures 6-7 in Appendix A). These workers received four days of training on early stimulation and responsive care. The training oriented the workers on the mechanics of child development, highlighted the importance of early stimulation at home in addition to nutrition and health, and instructed them on how to counsel mothers or caregivers to enable them in using the aforementioned materials to provide better learning opportunities for their children.[7] The NNS workers did not receive any additional cash or in-kind incentives to deliver the ECD intervention under the experimental setting.

[7] This session was added to the training for the NNS program, which the workers received in both treatment and control communities.

2.2 Study Site, Sample, and Randomization

Our study was executed in three districts in the Barisal, Chittagong, and Sylhet divisions in Bangladesh.[8] The research team selected the three districts based on the existing government-run NNS program supported by Save the Children under the Tackling Childhood Malnutrition (TCM) project. These three study sub-districts have a total of 41 unions.[9] From these unions, six unions were excluded due to incomplete health assistance records. From the remaining 35 unions, 30 unions were selected that had at least two community clinics. This process selected 78 community clinics in Muladi, Kulaura, and Satkania sub-districts and its catchment areas.

Within the selected areas, households with children aged between 3-18 months were eligible to participate in the study. The minimum age restriction was imposed due to the validity of the main developmental assessment tool used for the evaluation (the Bayley-III; Bayley, 2006), which had not been validated (at the time of the study) on children below the age of three months in Bangladesh. Moreover, the Bayley-III test is only valid for children aged up to 42 months. Hence, the upper age restriction was imposed to collect end-line assessment 24 months after the baseline. We sampled 33 eligible households from each of the 78 community clinic catchment areas to reach a total sample size of 2,574 households.

A detailed baseline survey was conducted between November 2013–January 2014. These baseline data included data on households' socio-economic conditions as well as assessments of anthropometric, parenting stimulation, and child development outcomes. Upon completion of the baseline surveys, we performed a community-clinic-level clustered randomization, where half of the community clinics (39) were assigned to treatment and the rest (39) to the control or "business as usual" arm. Treated community clinics and their catchment areas received a bundled ECD intervention in addition to regular NNS activities. The ECD intervention included training of NNS workers to provide dedicated counseling on ECD practices, as well as delivery of the ECD materials to households in the treatment group. These NNS workers included community health care providers, health assistants, and family welfare assistants. They were trained to discuss key mes-

[8] The three districts are Barisal, Chittagong, and Moulvibazar, and exhibit notable heterogeneity in child outcomes. For instance, incidence of low birth-weight ranges from 29-37% (nationally, 26%); share of children aged 36-59 months developmentally on track varies from 53-77% (nationally, 64%); literacy rate among 15-24 year-old women ranges from 74-95% (national average is 82%) (Bangladesh Bureau of Statistics and United Nations Children's Fund, 2015).

[9] Unions are the lowest administrative unit in Bangladesh. Public-funded community clinics are located in all unions.

sages of ECD practices during their regular home visits as well as visits by parents and children to community clinics and EPI events. End-line surveys with households and service providers were conducted from September–November 2015. The end-line household surveys included all questions and anthropometric assessments conducted at baseline, and additionally collected child development measures using the Bayley-III and Wolke Behavioral Rating Scales. The final sample size for the end-line survey was 2,486 households, with a low attrition rate of 3.4%. The service provider survey, on the other hand, was administered to 201 NNS workers, collecting information on their age, education, work and training experience, and perception of different aspects of child health and development, and on the duration and content of their recent household visit.

3 Data and Empirical Methods

3.1 Descriptive Statistics and Randomization Balance

Table 1 reports the summary statistics for the primary outcomes and a set of covariates measured at baseline and stratified by experimental assignment. None of the differences in means are significant at conventional levels. The mean age of study children is about 12 months, whereas the rates of stunting, wasting, and underweight are about 28%, 6%, and 19%, respectively. 17% of the children live in single-parent households (primarily because of migrant fathers). Mothers are, on average, 25 years old, predominantly Muslim (86%), and more likely to have continued beyond primary schooling (they, on average, completed an additional year of schooling compared to the fathers). Their rate of wage employment, on the other hand, is 5.5%. It takes close to half an hour for the average household to travel to the nearest community clinic. Finally, attrition is balanced across the two arms.

3.2 Empirical Methods

To estimate intent-to-treat impacts of treatment assignment on service providers, we estimate the following specification for service provider i in community clinic j and union (stratum) s:

$$Y_{i,j,s} = \gamma_0 + \delta_s + \gamma_1 T_j + \epsilon_{i,j,s} \tag{1}$$

where $Y_{i,j,s}$ refers to the outcome of interest, δ_s refers to strata (union) fixed effects, and T_j is an indicator for assignment to the treatment group. γ_1 is the coefficient of interest and standard errors are clustered by community clinic.

To estimate intent-to-treat impacts of treatment assignment for child i in community clinic j, we use an Analysis of Covariance (ANCOVA) specification:[10]

$$Y_{i,j,t+1} = \beta_0 + \beta_1 T_j + \beta_2 Y_{i,j,t} + \beta_3 \mathbf{X_{i,j,t}} + \varepsilon_{i,j,t+1} \tag{2}$$

The control variables $\mathbf{X_{i,j,t}}$ include the following baseline variables: (i) child-level controls: gender and age in months; (ii) parent-level controls: mother's age, weight, religion, education, and employment status; father's education and an indicator variable for the presence of the father for 6 months or more in the last year; (iii) household-level controls: measures of household composition including household size, the number of individuals in the households aged 18 years and below, an indicator for single parent households, an asset index, and the time taken to travel to the nearest community clinic; and endline-survey-month fixed effects. Union (strata) fixed effects are included in all regressions. Standard errors are clustered by community clinic, the unit of randomization.[11]

We address problems of multiple inference by creating indices for our three broad families of outcomes: child cognitive and language development, Wolke behavioral rating, and anthropometrics. Outcome variables are transformed into z-scores (relative to the endline control distribution) and then aggregated to form a standardized average across each outcome in the family (i.e., an index). We then test the overall effect of the treatment on the index (see Kling et al. 2007). To further account for multiple hypothesis testing across the three indices, we also report the sharpened False Discovery Rate (FDR) adjusted q-values (Anderson, 2008).

For the spillover analysis using older siblings of the index child, we use a specification similar to equation 2. We look at two age groups, namely 5-7 and 10-14 years, that represent two key margins of schooling decisions for the study households. For children aged 5-7, the parents decide when to enroll them in the primary school whereas for those aged 10-14, the choice entails whether to continue beyond primary school or drop out (primary completion rate is over 98%). We study the following outcomes: school enrollment (5-7 years old); school enrollment, completed years of schooling, caregiver-reported ability to

[10] Except in a small number of cases where the outcome was not measured at baseline, such as socio-economic skills. Table notes identify these cases.

[11] In an alternative specification, we choose controls applying the double-lasso method introduced by Urminsky et al. (2016), without penalizing the strata fixed effects and outcome measured at baseline (when available). The set of potential controls comprises the aforementioned covariates as well as an additional 63 variables: all cognitive, language, and anthropometric outcomes; additional parental and household demographics; child outcomes entailing feeding practices, illness, and healthcare seeking behavior; parental ECD knowledge and practices; maternal agency and healthcare-seeking behavior; shocks to the household since baseline; and Upazila/sub-district fixed effects. We present the results for cognitive, language, and socio-emotional skills, and for anthropometric outcomes in Appendix C.

write a letter, and labor force participation (10-14 years old).

4 Results

Our theory of change hypothesizes that service providers assigned to the treatment group attended early childhood development training, updated their beliefs on the importance of ECD, and changed their behaviors in the form of the time spent and topics discussed with parents. Households in the treatment group received the program materials and discussed ECD with service providers during home visits. We test these hypotheses starting with the impacts of treatment assignment on service providers, followed by households.

4.1 Treatment Impacts on Service Providers

Overall, the majority of the NNS workers were female (71%) and completed secondary education (82%). The median worker was aged 30-39 years, with 5 years of relevant experience.[12] Service providers in the treatment group were allocated to receive a four-day training program on early childhood stimulation delivered by Save the Children. In addition, they were asked to deliver the four program materials to households and discuss ECD practices with parents during their home visits.

Table 2 presents treatment impacts on various types of training received by the service providers. Column (1) highlights that 98% of service providers in the treatment group attended Save the Children's ECD program while none in the control group received the training. To assess the concern of potential crowd-out of other training provided to the service providers, we study the likelihood of providers to report receipt of training on child health, child feeding and nutrition, and other child-related training in columns (2), (3), and (4) of Table 2, respectively. Providers in the treatment group were 6.5 percentage points less likely to receive training on child health. This result is statistically significant at the 5% level. We do not find statistically significant treatment impacts on training on child feeding and nutrition or other child-related training.

The ECD program had important impacts on service providers' perceptions of the importance of early childhood stimulation. Table 3 presents ordered logit regressions on five different perceptions, each measured on a 3-point scale ranging from "neither important nor unimportant" to "very important." Columns (1)-(3) study perceptions relating to early childhood stimulation while columns (4)-(5) study perceptions relating to child health and nutrition. Service providers in the treatment group were more likely to report that it was important to teach mothers how to talk with children, respond to children's

[12] For service providers, age was recorded in 10-year age bands.

cues, and play games with their children (statistically significant at the 1% level). These results are in line with the curriculum of the ECD training attended by service providers in the treatment group. A potential concern might be a decline in the perceived importance of child health and nutrition owing to the lack of training on these topics (as seen in Table 2). However, service providers in the treatment group were *more* likely to report that it was important to teach mothers about child health and the types of food to feed their children, as seen in columns (4) and (5), respectively. This suggests that the training on ECD, in addition to training on child feeding and nutrition (which all service providers in the treatment group received), was sufficient to change perceptions on the importance of child health without the need to attend separate training on child health.

The change in perceptions among service providers in the treatment group also translated into more time spent talking to parents about early childhood stimulation. Columns (1)-(6) of Table 4 present the time spent (in minutes per household visit in the last working day) discussing six different topics with parents during their household visits. Column (7) presents treatment impacts on the total time spent with households, while column (8) presents impacts on the number of unique topics discussed.

Column (1) shows that NNS workers in the treatment group spent approximately three more minutes per household visit discussing early childhood stimulation, on a control group base of 0.04 minutes or two seconds (statistically significant at the 1% level). This is one of the key first stage impacts of the intervention. The increase in time spent on early childhood stimulation came at the expense of time spent discussing general nutrition and health of children, however. Service providers in the treatment group spent approximately 0.8 fewer minutes discussing general nutrition ($p < 0.01$) and 0.5 fewer minutes discussing child health ($p < 0.05$). They also spent approximately 0.5 fewer minutes discussing the Expanded Program of Immunization (EPI) ($p < 0.05$). We also observe decreases in time spent discussing family planning and issues relating to pregnancy for pregnant women, although these decreases are not statistically significant at conventional levels. Column (7) shows that there were no significant treatment impacts on the total time spent with households. Thus, we interpret these results as a reallocation of time away from child health and nutrition and towards early childhood stimulation. This reallocation is in line with the crowd-out of child health training for service providers observed in Table 2. Service providers in the treatment group discussed, on average, 0.4 more topics with households, on a control mean base of 3.7. This was largely driven by the increase in the number of service providers discussing ECD during their visits.

Time spent on different topics, however, is one of many possible measures of health-worker effort and may speak more to the quantity rather than the quality of their effort. Theoretically, it is plausible that non-pecuniary motivators such as training for some tasks can incentivize all tasks that are complementary (Holmstrom and Milgrom (1991); see Rowe et al. (2018) for a review of evidence on the role of non-financial incentives in improving health-worker productivity).

To verify the accuracy of the service provider reports, we also asked households about health worker visits along both the extensive (any visits, number of visits) and intensive margins (duration and topics discussed conditional on visits). Column 1 of Table 5 shows that households in the treatment group were 7.4 percentage points more likely to report being visited by a service provider in the last six months, on a control mean base of 18.6% (statistically significant at the 5% level). While we observe an increase in the total number of visits in column (2), this is not significant at conventional levels.

On the intensive margin, we asked households about the duration and topics discussed during their last visit by service providers, conditional on any visit. We do not observe a significant increase in the duration of the visits, a result consistent with the service provider reports. Households reported that service providers in the treatment group were significantly more likely to talk about playing with, and talking to children, in addition to talking about the program materials (child development card, picture books, and the key message booklet). All results on the various ECD topics discussed are positive and statistically significant at the 1% level, which is also consistent with the service provider reports.

In sum, service providers in the treatment group were significantly more likely to attend training on ECD, change their perceptions on the importance of ECD, and change their behavior, as reflected in their reallocation of time towards ECD during home visits. Without additional monetary incentives, however, they did not increase the total number of home visits nor the duration of each visit.

4.2 Receipt of Program Materials by Households

As per the intervention guidelines, households in the treatment group should have received four materials: a child development card, a household picture book, a nature picture book, and a key message booklet. However, Table 6 shows that due to imperfect compliance, the differential likelihood of receipt of the child development card, household picture book, and nature picture book between treatment and control households was approximately 49 percentage points (instead of 100 under perfect compliance). Furthermore, 2-3% of households in the control group received these materials. Finally, on a base of 1% in the control group, households in the treatment group were 16 percentage points more likely to have received the key message booklet, partially because of its late distribution starting in March 2015 (compared to June 2014 for the first three materials above).

Overall, column (5) shows that treatment group households were 50 percentage points more likely to have received any of the four intervention materials on a control group base of 2.6%. The receipt of materials by a small number of households in the control group may bias our estimates downward and we view our treatment impacts in the sections that follow as lower-bound estimates of the true impacts of the intervention.

4.3 Direct Impacts

We summarize the direct intent-to-treat (ITT) impacts of the intervention on the primary outcomes of interest in Figure 1. The outcomes we study are organized across the two broad families of child cognitive and language development, and the Wolke behavioral rating. Each diamond in the figure plots the point estimate for β_1 estimated using equation (2). The line corresponding to each point estimate reflects the 95% confidence interval for the outcome. Standard errors are clustered by community clinic. Each outcome variable has been standardized using the mean and standard deviation of the endline control group.

The intervention had statistically significant impacts across all measures of child cognitive and language development: cognition, language, expressive communication, and receptive communication. Panel A of Table 7 reports the ITT estimates for the Bayley measures. Children in treatment households scored 1.2 points – or 0.17 SD – higher on their cognitive composite score than children in control households. The corresponding estimate for language development is higher, specifically, 2.24 points or 0.23 SD. This is consistent with the program's emphasis on developing the child's communication skills. Between the two components of the language composite outcome, receptive communication exhibits a larger treatment effect, 0.24 SD, compared to 0.18 SD for expressive communication. These impacts are all statistically significant at the 1% level.

Across the five measures considered under the Wolke behavioral rating, we observe statistically significant impacts for three outcomes: approach, emotion, and activity. ITT estimates for the Wolke behavioral measures are reported in Panel B of Table 7, where we observe effect sizes ranging from 0.12 to 0.14 SD for these three measures. Taken together, the program had a positive impact on children's socio-emotional development.

We address problems of multiple inference by creating indices for the two broad families of outcomes: child cognitive and language development, and the Wolke behavioral rating. We observe a 0.21 SD gain in the index summarizing the Bayley outcome measures in the treatment group over the control group (95% CI: 0.11-0.32 SD). For the Wolke index, we observe a 0.12 SD gain (95% CI: 0.03-0.21 SD). Both indices are statistically significant at the 1% level.

Figure 2 presents treatment impacts on child anthropometrics. We observe higher weight-for-height and weight-for-age z-scores among treated children relative to the control group that also translate into significant and meaningful reductions in the rates of wasting ($p < 0.05$) and underweight ($p < 0.1$). Table 8 shows that weight-for-height and weight-for-age z-scores are higher in the sample households in treatment clinic catchment areas by 0.22 SD and 0.14 SD, respectively. This translates into a 4.8 percentage-point reduction in wasting. The decline in the prevalence of underweight, on the other hand, is 2.9 percentage points. We do not find any significant effects on height-for-age z-scores or stunting. The results on anthropometrics are important given concerns over possible crowd-out of resources from the NNS that might adversely affect child nutrition outcomes. Instead, we observe improvements in child anthropometrics in the treatment group. Overall, we observe a 0.12 SD gain (95% CI: 0.07-0.16 SD) in the anthropometrics index in the treatment group (statistically significant at the 1% level). To further account for multiple hypothesis testing across the three indices, we estimate the sharpened False Discovery Rate (FDR) adjusted q-values (Anderson, 2008), all of which are below 0.01 (0.001, 0.003, and 0.001 for the Bayley, Wolke, and Anthropometric indices, respectively; reported in Table 1, Appendix B).

We compare the magnitudes of our estimates with those of related ECD studies at-scale in Table 16. Despite differences in the nature of interventions and target populations, we find that our estimates are comparable with these studies. For example, the study in Peru by Araujo et al. (2021) finds an estimated 0.1 Standard deviation (SD) ITT impact on infant skill as measured by the Ages and Stages Questionnaire, which is lower than our estimates. Meanwhile, Sylvia et al. (2020) report magnitudes in the range of 0.22-0.26 SD for cognition and receptive language in China, which are slightly larger than our estimates. An outlier in the estimated impacts seems to be the study by Yousafzai et al. (2014) in Pakistan, where estimated impacts for cognition, language, and motor development range from 0.5-0.7 SD. A notable difference is that all studies listed in Table 16, except that by Sylvia et al. (2020), provided financial incentives to service providers delivering the intervention.

4.4 Indirect Impacts: Sibling Spillovers

We study the indirect impacts of the intervention on the older siblings of treated children in Table 9. We look at siblings aged 5-7 at endline in column (1) and 10-14 at endline in columns (2)-(5). Siblings aged 5-7 were 6.2 percentage points or 0.13 SD more likely to attend school in the treatment group (significant at the 5% level). While siblings of treated children aged 10-14 were more likely to attend school, have more years of schooling, and able to write a letter, these outcomes were noisily measured and the point estimates are not statistically significant. In column (5), we see that siblings of treated children aged 10-14 years at endline were 1.5 percentage points less likely to be in the labor force (significant at the 5% level).

Overall, the treatment had positive schooling spillovers on siblings that are closer in age to the treated children. This is likely due to the relevance of the program materials to the siblings and parents spending more time on ECD activities with younger children.

5 Compliance & Dose-response Relationship

On average, households in the treatment group owned the child development card for 6 months. In Table 10, we assess the drivers of incomplete compliance, regressing the number of months in which households in the treatment group owned the child development card against a large set of variables including child-level (age, gender), parent-level (age, education), and household-level (wealth, size) characteristics. Overall, compliance is strongly negatively correlated with distance, or time to the nearest clinic. On average, households owned the childhood development card for 0.06 fewer months, or 2 fewer days, for every additional minute located further away from the nearest clinic ($p < 0.01$). Since households in the treatment group were located, on average, 30 minutes away from the nearest clinic, this implies that the average household owned the child development card for 2 fewer months than households located very close to clinics.

In column (2), we study compliance in ownership of at least one of the four intervention materials. While we observe a statistically significant relationship with distance ($p < 0.01$), the magnitude of the correlation is small – households owned any program materials for 0.004 fewer months, or 0.1 fewer days, for every additional minute located further away from the nearest clinic. None of the other variables considered were statistically significant at the 5% level.

Given imperfect compliance, we present estimates of the dose-response relationship between the number of months that households owned the childhood development card and the seven outcomes of primary and second interest that displayed statistically significant impacts in Figures 1 and 2. Specifically, we use the following specification for household i in community clinic j:

$$Y_{i,j,t+1} = \alpha_0 + \alpha_1 Compliance_{i,j} + \alpha_2 Y_{i,j,t} + \alpha_3 \mathbf{X_{i,j,t}} + \eta_{i,j,t+1} \tag{3}$$

where $Y_{i,j,t+1}$ is the outcome of interest measured at endline, $Compliance_{i,j}$ represents the number of months that household i owned the childhood development card, $\mathbf{X_{i,j,t}}$ comprises the set of baseline covariates defined in Section 3.2, and $\eta_{i,j,t+1}$ is the error term. To account for potentially endogenous take-up of treatment, we instrument for compliance with treatment assignment using two-stage least-squares (2SLS). α_1 is the coefficient of interest and standard errors are clustered by community clinic.

Table 11 shows the IV estimates of the dose-response relationships. Assuming a linear relationship, we also present effect sizes in SD units of the endline control group, if households were to own the childhood development card for 1 year (in our experiment, households only held the card on average, for 6 months). These 12-month effect sizes show that ownership of the card for a full year could lead to 0.33-0.44 SD gains in Bayley scores, 0.23-0.27 SD gains in Wolke scores, and 0.27-0.42 SD gains in anthropometric outcomes. Scale-up challenges and imperfect compliance, however, muted (in other-words "voltage dropped") the hypothetical gains that could have been achieved by the intervention.

6 Mechanisms

What are some mechanisms underlying the theory of change for the intervention, in addition to receipt of the program materials and ECS discussions with service providers? In this section, we consider three mechanisms: maternal agency, parental investments, and take-up of the NNS program. We then conduct a mediation analysis that explores the relative contributions of the different mediators in explaining our estimated impacts.

6.1 Mother's agency in the household

The first panel in Figure 3 reports ITT effects of the intervention on maternal decision-making in the following four domains: Money (household food expenditure and choice, buying important things for the family, and how to spend household earnings); Food (daily food preparation and child feeding practices); Child Health (responses to child's serious illness and taking child to health facility for growth monitoring checks or immunization); and Child Stimulation (expenditure on play materials for the child and taking him/her outside to visit family or friends). For every domain, we aggregate the number of components in which the mother usually makes decisions - either by herself or jointly with her husband. Consistent with the handful of ECD studies reporting results on this dimension (Evans et al., 2021), we find positive and significant effects on maternal decision-making agency, ranging from 0.08 to 0.14 SD, with larger effects in child health and stimulation. By equipping mothers with information on optimal stimulation habits, their benefits as well as materials to practise them, the intervention may have prompted them to engage more in decision-making around child development. This evidence also aligns well with Roy et al. (2019), who found similar impact on maternal agency of a behavior-change-communication program in Bangladesh that counseled mothers on child nutrition in addition to offering food or cash transfers.

6.2 Complementary Parental Investments

The second panel of Figure 3 presents ITT impacts of the intervention on five measures of parental investments. The first three variables relate to investments by both the mother and father: a stimulation knowledge scale, Home Observation for Measurement of the Environment (HOME) inventory scale (Bradley and Caldwell, 1979), and a measure of the variety of play materials and learning activities provided to the child. A short list of questions were asked of caregivers to study stimulation knowledge. These included agree/disagree questions such as, "It is important to teach the baby names of simple objects and colors." Questions under the HOME inventory scale were organized into three broad categories: (i) whether the caregiver promotes child development, (ii) organization of the physical environment for the child, and (iii) the provision of opportunities for variety in daily stimulation. Overall, we do not observe significant program impacts on parents' stimulation knowledge and the HOME inventory scale.

Program materials provided under the intervention are not included when studying the variety of play materials and learning activities provided to the child. Learning activities included reading books or showing pictured books, telling stories or nursery rhymes, singing songs or lullabies, playing toys, or naming, counting, and/or drawing things with the child. The composite score of play materials and learning activities counted the total number of such play materials and learning activities and ranged from 0-11. On an endline control mean of 6.484, households in the treatment group scored 0.26 points (0.094 SD) higher on this score relative to the control group at endline (see third measure of the 2nd panel in Figure 3).

The fourth and fifth measures in this panel capture the variety of learning activities provided to the child specifically by the mother and father, respectively. We observe that mothers in the treatment group scored 0.18 points (0.115 SD) higher on this score relative to the endline control group mean of 2.324 ($p < 0.01$). Notably, we do not observe any treatment impacts for fathers. This result is in line with the increase in maternal agency seen in the first panel Figure 3.

Overall, the intervention did not lead to significant impacts on parents' stimulation knowledge or the HOME inventory scale but led to complementary parental investments in the variety of play materials and learning activities for the child, specifically by mothers. This channel could help explain the positive treatment impacts that we observe on cognitive and language development, as well as the Wolke behavioral rating scales.[13]

6.3 Complementary Take-up of Nutrition Program

The intervention used existing National Nutrition Services (NNS) staff to deliver the intervention instead of a separate group of ECD workers. Did the treatment have any impact on the take-up of the NNS program?

The third panel of Figure 3 presents ITT impacts on take-up of the NNS program. We first study whether parents of treated children had and were able to show the NNS growth development card to the enumerator when surveyed. Households in the treatment group were 7.5 percentage points (0.17 SD) more likely to have and show the growth development card relative to the control group endline mean of 27.5%. Next, we study the number of NNS growth monitoring check-ups taken up by parents. Children in the treatment group took up 0.25 (0.21 SD) more growth monitoring check-ups relative to the control group base of 0.66 check-ups at endline. Overall, we find statistically significant complementary take-up of the NNS program.

6.4 Mediation Analysis

What are the relative contributions of the mechanisms outlined above in comparison with the direct intervention channel (receipt of program materials)? To explore this further, we conduct a mediation analysis, and consider (i) the receipt of any program materials, (ii) maternal agency on child stimulation, (iii) complementary parental investments in the

[13] We explore parental investments in the form of the diet provided to children in Table 7 of Appendix D. While we observe statistically significant increases in the provision of flesh foods (fish, poultry, or meat) and Khichuri (a protein-rich local dish) by parents in the treatment group, there is no overall increase in the number of food groups consumed by children.

form of mother's variety of learning activities, and (iv) complementary NNS take-up as measured by the number of growth monitoring check-ups, as possible mediators.

We follow Gelbach (2016) to conduct the mediation analysis.[14] The treatment effect of our intervention T on outcome Y can be decomposed as operating through the four mediators $\{m_j\}_{j=1}^4$:

$$\frac{dY}{dT} = \sum_{j=1}^{4} \frac{\partial Y}{\partial m_j} \frac{\partial m_j}{\partial T} + R \tag{4}$$

where the residual R is the part of the treatment impact that cannot be attributed to the four mediators. The analysis follows a four-step procedure for each outcome Y: first, obtain an estimate of $\hat{\beta}_1^U$ in an *unconditional* ITT regression: $Y = \alpha + \beta_1^U T + u_1$. Second, obtain estimates $\{\hat{\beta}_2^j\}_{j=1}^4$ of the impact of the four mediators on Y in a *conditional* ITT regression: $Y = \alpha + \beta_1^C T + M\beta_2 + u_2$. Third, obtain estimates $\{\hat{\gamma}^j\}_{j=1}^4$ of the treatment impact on each mediator j: $m_j = \alpha + \gamma^j T + u_3$. Fourth, compute mediator j's contribution to the treatment impact: $\frac{\hat{\gamma}^j \hat{\beta}_2^j}{\hat{\beta}_1^U}$. We repeat this four-step procedure for the two primary outcome indices, i.e. the Bayley index and Wolke index.

The methodology used has the advantage of invariance to the order in which the mediators are considered. However, the analysis does not represent causal mediation except under strong assumptions (Imai et al. 2010). One such assumption is the requirement that the mediators do not impact each other. This is unlikely to hold in our setting since maternal agency, investments, and NNS take-up likely interact with one other. By ignoring such interactions, we likely leave more of the overall treatment effect unexplained (R).

Table 12 presents the mediation analysis for the Bayley and Wolke indices. Column (1) shows that 33.5% of the estimated ITT impact on the Bayley index can be explained by the receipt of any of the four program materials. A further 7.5% can be explained by complementary parental investments in the form of an increase in the variety of learning activities provided by the mother. Column (2) shows that for the Wolke index, 14.2% of the estimated ITT impact can be explained by the complementary parental investments. This is likely because the learning activities, including playing toys and drawing things with the child, led to improvements in the approach, activity, and cooperation scores of the Wolke index, while activities such as singing songs and lullabies with the child led to

[14] This methodology has subsequently been applied in other studies, including a randomized evaluation of an early life intervention in Nigeria (Carneiro et al. 2021). In their study, a large fraction of the treatment effect (40%) remains unexplained, while 60% of the height-for-age impact at midline could be explained by information-related mediators.

gains in the vocalization score. In comparison, the program materials led to large direct gains in cognition, but not socio-emotional skills.

In comparison to the contributions of the program materials and maternal investments, maternal agency and complementary NNS take-up do not show statistically significant contributions to the Bayley and Wolke indices (the estimates of β_2 in the conditional ITT regressions were statistically insignificant with $p > 0.1$). This is likely because the increase in maternal agency leads to treatment impacts through maternal investments in children. While increases in NNS take-up may have led to nutritional impacts, this was unlikely to lead to significant gains in cognition and socio-emotional development, given the nutrition and health focus of the NNS.

6.5 Heterogeneity by Household & Service Provider Characteristics

Are the observed treatment impacts larger for particular sub-groups of the sample? In this section, we study treatment impact heterogeneity by four household characteristics (household wealth, gender of the child, mother's education, and father's education), in addition to two service provider characteristics (education and experience).

Table 13 presents results of the heterogeneity in treatment impacts by household characteristics for seven key outcome variables of interest along which we observe statistically significant impacts in the overall sample: cognitive and language composite scores, Wolke approach, emotion, and activity scores, as well as weight-for-height and weight-for-age z-scores.

In Panel A, we study heterogeneity by household wealth. The point estimates on the interaction term are consistently negative in six of the seven columns, although not always statistically significant. These interaction terms are significantly negative ($p < 0.05$) for the Wolke approach score as well as the two anthropometric z-score measures. Overall, treatment effects are larger for poorer households.

We study heterogeneity by the gender of the child in Panel B. Overall, we do not find significant impacts in the treatment effects by gender, with the exception of the Wolke activity score, in which girls in the treatment group saw larger effects (with boys showing no statistically significant impacts). Differences in treatment effects by maternal and paternal education are explored in panels C and D, respectively. We observe larger treatment effects on anthropometric z-scores for the children of fathers with primary school education or less.

Heterogeneity by service provider characteristics are presented in Table 14. Panel A explores differential treatment impacts by service provider education. Children in treatment groups where service providers had above-median levels of schooling saw larger gains in anthropometric outcomes ($p < 0.01$). Panel B shows that children in treatment groups where service providers had above-median levels of experience saw smaller gains in anthropometric outcomes and the Wolke activity score ($p < 0.01$). This is likely because service provider education and experience exhibit a negative correlation. More experienced workers may also be less likely to change their behaviors in comparison to younger, less experienced workers.

7 Cost-Benefit Analysis

As developing countries weigh the cost-effectiveness of various poverty alleviation strategies, understanding the rates of return to early childhood programs can have important policy implications for how governments choose to spend scarce resources. We present a cost-benefit analysis of the program and compute the internal rate of return (IRR), the discount rate at which the net present values of the costs and benefits are equal.

Table 15 presents a breakdown of intervention costs in 2014 U.S. dollars. Overall, the program cost $127,534. Personnel costs to Save the Children management staff was the largest cost category, accounting for 36% of total costs. Administration costs included office rent, utilities, communications, and equipment. Training of service providers formed the third largest cost category. Spreading this cost across 39 community clinics and 117 service providers, we obtain per clinic costs of $3,270 and per service provider costs of $1,090. Across 18,644 target beneficiaries, the cost per child was $6.84 (538 Taka, 2014 values). These costs are low given that the intervention relied on existing NNS infrastructure and service providers.

The benefits of the program include an increase in wages arising from improvements in cognitive and non-cognitive social skills. To convert the estimated treatment impacts into wage impacts, we use estimates from Deming (2017) for the U.S., who notes that a one standard deviation increase in cognitive skills increased wages by 15.1% in the NLSY97 and a one standard deviation increase in social skills yields a wage gain of 3.7% in the NLSY97.[15] Applying these gains to the 0.17 SD gain in the cognition composite score and 0.12 SD gain in Wolke index, respectively, we obtain a treatment impact of 3% on income, assuming additive separability of the cognitive and non-cognitive gains. We note that an alternative method of monetizing the benefits could have used estimates of the returns to education in developing countries (for example, the 9% return estimated by

[15]Unfortunately, comparable estimates of the cognitive and non-cognitive skill-wage premium are not available for developing countries.

Patrinos and Psacharopoulos 2020). However this strategy would require an assumption on the impact that increases in cognitive and non-cognitive skills have on education in developing countries, for which data is not readily available. As a result, we did not proceed with the latter approach.[16]

We assume that individuals enter the labor market at age 21 and work to age 59, the retirement age for public servants in Bangladesh. We further assume that individuals earn an annual income of 82,637 Taka (2014 values) and do not experience growth in income over their lifetime.[17] The benefits of the intervention are then calculated as the net present value of the gain in income over the individual's lifetime. All present values are discounted to the individual's year of birth.

Under these assumptions, we compute an IRR of 19.6% for the intervention. This is significantly larger than rates of return to early childhood development programs implemented in other countries. For example, Heckman et al. (2010) estimate that the High/Scope Perry Preschool Program in the U.S. had an estimated IRR in the range of 7-10 percent, while García et al. (2019) estimate that the Carolina Abecedarian Project (ABC) and the Carolina Approach to Responsive Education (CARE) had an IRR of 13.7%. Deming (2009) estimates an IRR of 7.9% for the Head Start early childhood program in the U.S.

There are alternative benchmarks to evaluate an ECD intervention such as unconditional cash transfer or book-delivery. While we agree that it would be interesting to have these as possible benchmarks, we argue that the absence of these benchmarks does not detract from the exercise we undertake for three reasons. First, there is still a lack of rigorous evidence on large-scale, potentially sustainable ECD interventions in resource-poor settings using high-quality and comprehensive measurements of outcomes and mechanisms. Second, the current evidence from the large literature on unconditional cash transfer is rather mixed on child health and nutrition (de Groot et al., 2017; Pega et al., 2017). Moreover, given the size of transfers in programs that have a positive effect on household

[16] Alternatively, we can use estimates from the 20 and 30-year follow-ups of the Jamaica early childhood stimulation trial, where Gertler et al. (2014) and Gertler et al. (2021) find that the short-run gains for the treated children in cognition (0.91 SD) and language (0.59 SD) translate to 25% and 37% higher earnings at age 22 and 31 years, respectively. Hence the estimate we use to monetize a 1 SD gain in cognitive skills, namely 15.1% (at age 21 years, and assuming no growth in earnings thereafter), is probably a lower bound for low-income settings. Note, however, that the study population for the Jamaica trial was stunted in the baseline.

[17] The Bangladesh Labor Force Survey 2016-2017 notes that average monthly income for individuals in the labor force was 13,258 Taka, while the labor force participation rate was 58.2%. Taka values are deflated to 2014 values using inflation rates available at `https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=BD`.

consumption (e.g., US$ 285 in Haushofer and Shapiro (2016)), it is unlikely that a one-time transfer of $7 (equivalent to the per-child cost of the ECD program we study) would have any substantive effect on child development. Third, while our results point towards the potential for materials distribution as a scalable and leaner approach, it was probably not obvious ex-ante.

While the benefits of the program are comparable to other interventions, the costs are significantly lower due to the economies of scale and implementation of the program by NNS service providers. The intervention thus highlights that it is possible for governments to leverage existing public service delivery platforms to implement early childhood development interventions with high rates of return at scale.

8 Conclusion

A multi-disciplinary body of work shows that a child is not born with a fixed brain capacity. The human brain builds and develops continuously during the first five years of life and complex neural advances and connections are formed by a myriad of factors including nutrition, the parenting and home environment, and exposure to positive stimulation (Almond and Currie 2011, Shonkoff et al. 2012, Black et al. 2017). Investments by governments and parents are therefore essential to lay the foundations for human capital accumulation from an early age. However, the provision of ECD interventions at scale remains a key policy challenge in developing countries.

We experimentally study the delivery of an ECD program at scale in three divisions in Bangladesh. Utilizing a unique implementation model, we used existing government NNS staff to implement the program at scale without any additional financial incentives for the providers. Similar to other public sector programs delivered at scale in developing countries, the execution of the program faced several operational and compliance challenges (Sylvia et al., 2020). Despite the implementation challenges, ITT estimates at endline show improvements in child outcomes along various dimensions of cognitive (0.17 SD), language (0.23 SD), and socio-emotional development (0.12-0.14 SD). The intervention also led to improvements in nutritional outcomes, reducing wasting and underweight rates by 4.8 and 2.9 percentage points, respectively. The program was very cost-effective, with an IRR of 19.6%.

A limitation of the study design is that we can not disentangle the effects of materials from visits. Although we do not see any difference in the number of visits by health-workers, the visits succeeding the distribution of materials can serve as monitoring or reminders to use them. This probably confounds the impact of higher exposure to materials, which we explore in a LATE framework. The visits can also reinforce the message on complementarities in investments in stimulation and nutrition which, our analysis suggests, is a key mechanism. This offers concrete directions for new research that can further experiment with the design of such programs with emphasis on understanding parental beliefs about the human capital production function (Cunha et al., 2013; Attanasio et al., 2020; Giannola, 2021).

Our study demonstrates promising possibilities to scale up ECD programs in developing countries and opens up several areas for future research. Further work is needed to understand the challenges pertaining to compliance of program delivery. The improvement of ECD interventions using the public sector infrastructure, alternative models of program delivery, and targeting of the most vulnerable children remain important areas for future work.

References

Almond, Douglas and Janet Currie, "Human Capital Development before Age Five," *Handbook of Labor Economics*, 2011, *4b* (15), 1315–1486.

Anderson, Michael L, "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 2008, *103* (484), 1481–1495.

Araujo, M. Caridad, Marta Dormal, Sally Grantham-McGregor, Fabiola Lazarte, Marta Rubio-Codina, and Norbert Schady, "Home visiting at scale and child development," *Journal of Public Economics Plus*, 2021, *2* (100003).

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S Lee, "Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services," *American Economic Review*, 2020, *110* (5), 1355–94.

Attanasio, Orazio, Helen Baker-Henningham, Raquel Bernal, Costas Meghir, Diana Pineda, and Marta Rubio-Codina, "Early Stimulation and Nutrition: The Impacts of a Scalable Intervention," *Journal of the European Economic Association*, 2022, *jvac005.*

Attanasio, Orazio P, Camila Fernández, Emla O A Fitzsimons, Sally M Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina, "Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial," *BMJ*, 2014, *349* (5785).

Attanasio, Orazio, Teodora Boneva, and Christopher Rauh, "Parental Beliefs about Returns to Different Types of Investments in School Children," *Journal of Human Resources*, 2020, pp. 0719–10299R1.

Bangladesh Bureau of Statistics and United Nations Children's Fund, "Final Report of the Bangladesh 2012-13 Multiple Indicator Cluster Survey," *MICS*, 2015.

Barham, Tania, "Enhancing Cognitive Functioning: Medium-Term Effects of a Health and Family Planning Program in Matlab," *American Economic Journal: Applied Economics*, 2012, *4* (1), 245–273.

Black, Maureen M, Susan P Walker, Lia CH Fernald, Christopher T Andersen, Ann M DiGirolamo, Chunling Lu, Dana C McCoy, Günther Fink, Yusra R Shawar, Jeremy Shiffman et al., "Early childhood development coming of age: science through the life course," *The Lancet*, 2017, *389* (10064), 77–90.

Bradley, R. H. and B. M. Caldwell, "Home observation for measurement of the environment: A revision of the preschool scale," *American Journal of Mental Deficiency*, 1979, *84*, 235–244.

Campbell, Frances A. and Craig T. Ramey, "Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families," *Child Development*, 1994, *65* (2), 684–698.

Campbell, Frances, Gabriella Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan, "Early Childhood Investments Substantially Boost Adult Health," *Science*, 2014, *343* (6178), 1478–1485.

Carneiro, Pedro, Lucy Kraftman, Giacomo Mason, Lucie Moore, Imran Rasul, and Molly Scotty, "The Impacts of a Multifaceted Pre-natal Intervention on Human Capital Accumulation in Early Life," *American Economic Review, forthcoming*, 2021.

Cunha, Flávio, Irma Elo, and Jennifer Culhane, "Eliciting Maternal Beliefs about the Technology of Skill Formation," *NBER Working Paper 19144*, 2013.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78* (3), 883–931.

Currie, Janet and Tom Vogl, "Early-Life Health and Adult Circumstance in Developing Countries," *Annual Review of Economics*, 2013, *5* (1), 1–36.

de Groot, Richard, Tia Palermo, Sudhanshu Handa, Luigi Peter Ragno, and Amber Peterman, "Cash transfers and child nutrition: pathways and impacts," *Development Policy Review*, 2017, *35* (5), 621–643.

Deming, David, "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start," *American Economic Journal: Applied Economics*, 2009, *1* (3), 111–134.

Deming, David J., "The Growing Importance of Social Skills in the Labor Market," *The Quarterly Journal of Economics*, 2017, *132* (4), 1593–1640.

Deserranno, Erika, "Financial incentives as signals: Experimental evidence from the recruitment of village promoters in Uganda," *American Economic Journal: Applied Economics*, 2019, *11* (1), 277–317.

Evans, David K, Pamela Jakiela, and Heather A Knauer, "The impact of early childhood interventions on mothers," *Science*, 2021, *372* (6544), 794–796.

García, Jorge Luis, James J. Heckman, Duncan Ermini Leaf, and María José Prados, "Quantifying the Life-cycle Benefits of a Prototypical Early Childhood Program," *NBER Working Paper 23479*, 2019.

Gelbach, Jonah B., "When Do Covariates Matter? And Which Ones, and How Much?," *Journal of Labor Economics*, 2016, *34* (2), 509–543.

Gertler, Paul, James J. Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor, "Labor market returns to an early childhood stimulation intervention in Jamaica," *Science*, 2014, *344* (6187), 998–1001.

— **, James J Heckman, Rodrigo Pinto, Susan M Chang, Sally Grantham-McGregor, Christel Vermeersch, Susan Walker, and Amika Wright**, "Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31," Working Paper 29292, National Bureau of Economic Research 2021.

Giannola, M, "Parental investments and intra-household inequality in child human capital: Theory, measurement, and evidence from a lab-in-the-field experiment," *University College London*, 2021.

Grantham-McGregor, Sally M, Christine A Powell, Susan P Walker, and John H Himes, "Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study," *The Lancet*, 1991, *338* (8758), 1–5.

Hamadani, Jena D, Fahmida Tofail, Syed N Huda, Dewan S Alam, Deborah A Ridout, Orazio Attanasio, and Sally M Grantham-McGregor, "Cognitive deficit and poverty in the first 5 years of childhood in Bangladesh," *Pediatrics*, 2014, *134* (4), e1001–e1008.

Haushofer, Johannes and Jeremy Shapiro, "The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya," *The Quarterly Journal of Economics*, 2016, *131* (4), 1973–2042.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz, "The Rate of Return to the High/Scope Perry Preschool Program," *Journal of Public Economics*, 2010, *94* (1-2), 114–128.

Holmstrom, Bengt and Paul Milgrom, "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design," *JL Econ. & Org.*, 1991, *7*, 24.

Imai, Kosuke, Luke Keele, and Dustin Tingley, "A General Approach to Causal Mediation Analysis," *Psychological Methods*, 2010, *15* (4), 309–334.

Joshi, Shareen and T. Paul Schultz, "Family Planning and Women's and Children's Health: Long-Term Consequences of an Outreach Program in Matlab, Bangladesh," *Demography*, 2013, *50* (1), 149–180.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Knudsen, Eric I., James J. Heckman, Judy L. Cameron, and Jack P. Shonkoff, "Economic, neurobiological, and behavioral perspectives on building America's future workforce," *PNAS*, 2006, *103* (27), 10155–10162.

List, John A., *The Voltage Effect*, Crown Publishing Group, 2022.

— **, Dana Suskind, and Lauren H. Supplee**, *The Scale-Up Effect in Early Childhood and Public Policy*, Routledge, 2021.

Lu, Chunling, Maureen M Black, and Linda M Richter, "Risk of poor development in young children in low-income and middle-income countries: an estimation and analysis at the global, regional, and country level," *The Lancet Global Health*, 2016, *4* (12), e916–e922.

Nyqvist, Martina Björkman, Andrea Guariso, Jakob Svensson, and David Yanagizawa-Drott, "Reducing child mortality in the last mile: Experimental evidence on community health promoters in Uganda," *American Economic Journal: Applied Economics*, 2019, *11* (3), 155–92.

Patrinos, Harry Anthony and George Psacharopoulos, "Chapter 4 - Returns to education in developing countries," *The Economics of Education (Second Edition)*, 2020, pp. 53–64.

Pega, Frank, Sze Yan Liu, Stefan Walter, Roman Pabayo, Ruhi Saith, and Stefan K Lhachimi, "Unconditional cash transfers for reducing poverty and vulnerabilities: effect on use of health services and health outcomes in low-and middle-income countries," *Cochrane Database of Systematic Reviews*, 2017, (11).

Richter, Linda M, Bernadette Daelmans, Joan Lombardi, Jody Heymann, Florencia Lopez Boo, Jere R Behrman, Chunling Lu, Jane E Lucas, Rafael Perez-Escamilla, Tarun Dua, Zulfiqar A Bhutta, Karin Stenberg, Paul Gertler, Gary L Darmstadt, Paper 3 Working Group, and The Lancet Early Childhood Development Series Steering Committee, "Investing in the foundation of sustainable development: pathways to scale up for early childhood development," *The Lancet*, 2017, *389* (10064), 103–118.

Rowe, Alexander K, Samantha Y Rowe, David H Peters, Kathleen A Holloway, John Chalker, and Dennis Ross-Degnan, "Effectiveness of strategies to improve health-care provider practices in low-income and middle-income countries: a systematic review," *The Lancet Global Health*, 2018, *6* (11), e1163–e1175.

Roy, Shalini, Melissa Hidrobo, John Hoddinott, and Akhter Ahmed, "Transfers, behavior change communication, and intimate partner violence: Postprogram evidence from rural Bangladesh," *Review of Economics and Statistics*, 2019, *101* (5), 865–877.

Shonkoff, Jack P, Andrew S Garner, Benjamin S Siegel, Mary I Dobbins, Marian F Earls, Laura McGuinn, John Pascoe, David L Wood, Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood, Adoption, and Dependent Care et al., "The lifelong effects of early childhood adversity and toxic stress," *Pediatrics*, 2012, *129* (1), e232–e246.

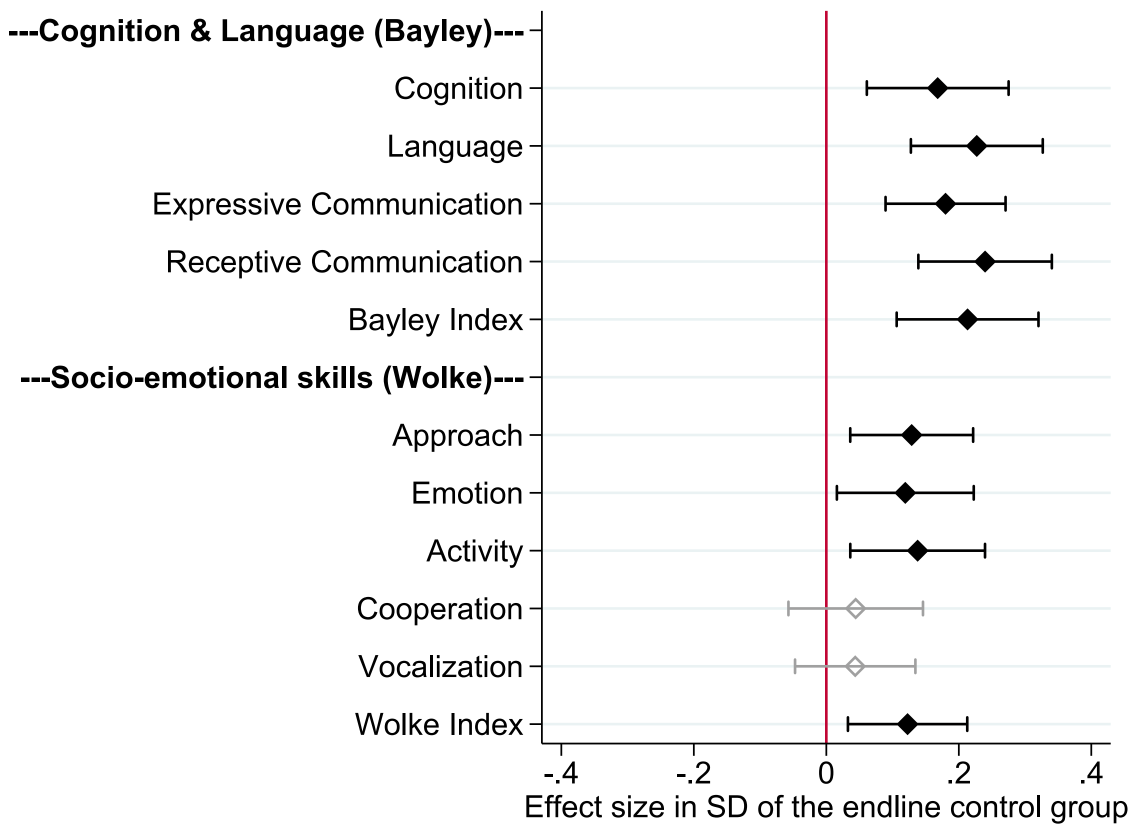
Sylvia, Sean, Nele Warrinnier, Renfu Luo, Ai Yue, Orazio Attanasio, Alexis Medina, and Scott Rozelle, "From Quantity to Quality: Delivering a Home-Based Parenting Intervention Through China's Family Planning Cadres," *The Economic Journal*, 2020, *ueaa114.*

Urminsky, Oleg, Christian Hansen, and Victor Chernozhukov, "Using double-lasso regression for principled variable selection," *Available at SSRN 2733374*, 2016.

Walker, Susan P., Susan M. Chang, Marcos Vera-Hernández, and Sally Grantham-McGregor, "Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior," *Pediatrics*, 2011, *127* (5), 849–857.

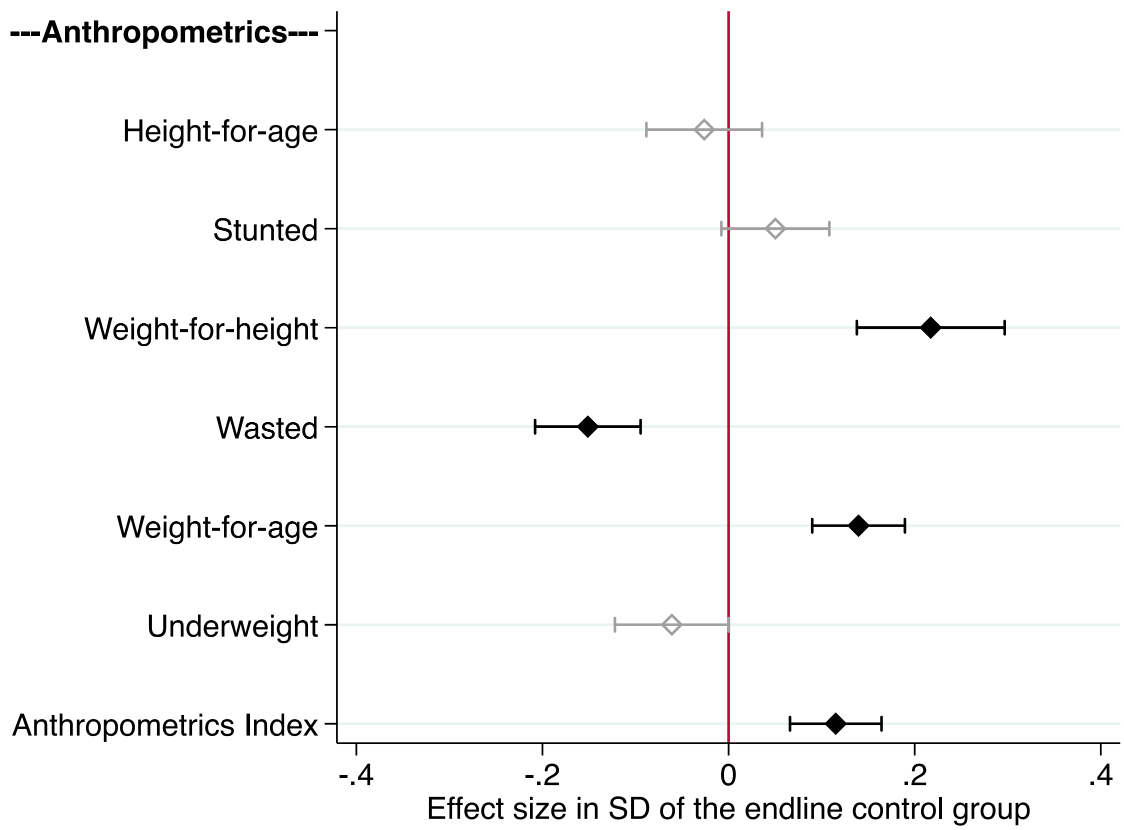
Yousafzai, Aisha K, Muneera A Rasheed, Arjumand Rizvi, Robert Armstrong, and Zulfiqar A Bhutta, "Effect of integrated responsive stimulation and nutrition interventions in the Lady Health Worker programme in Pakistan on child development, growth, and health outcomes: a cluster-randomised factorial effectiveness trial," *Lancet*, 2014, *384* (9950), 1282–93.

Figure 1: Direct Impacts on Primary Outcomes of Interest



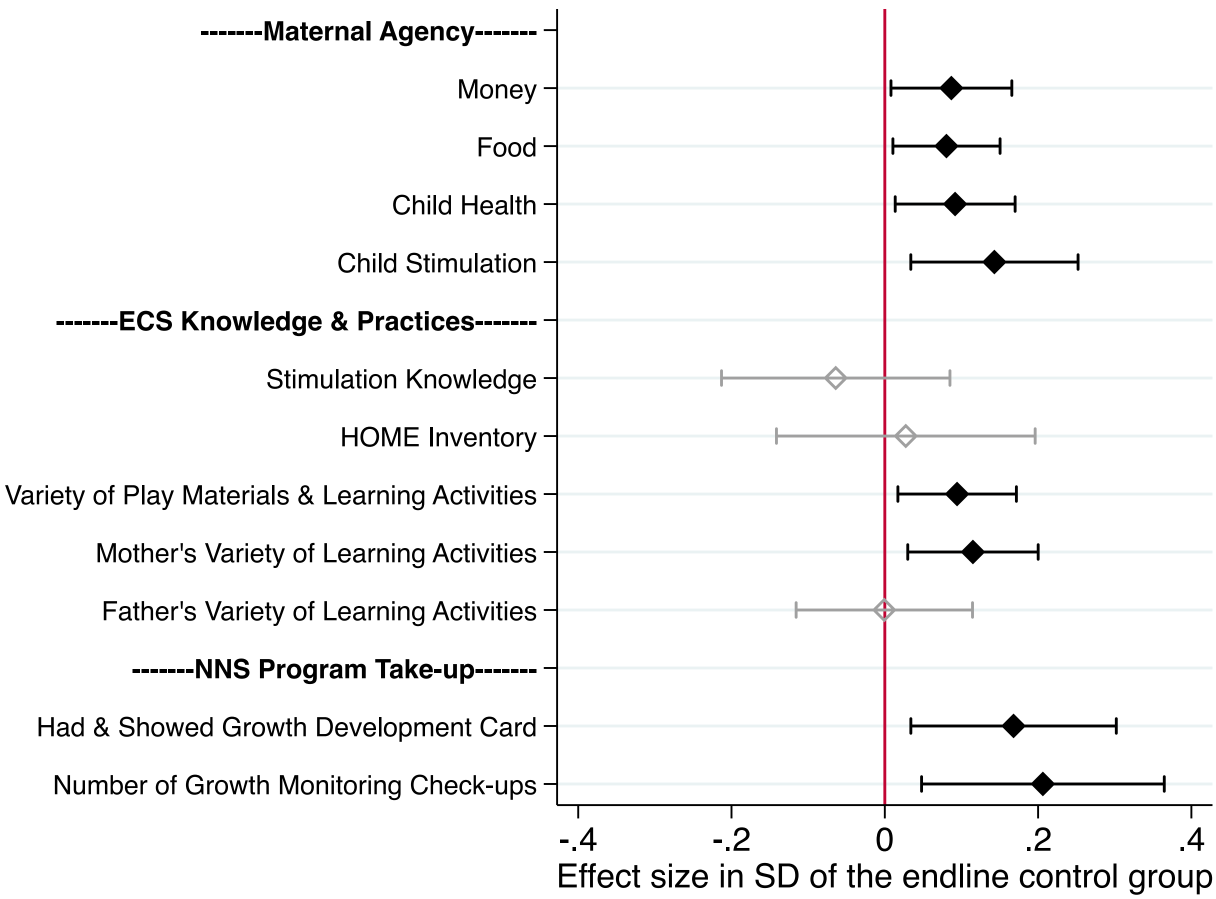
Notes: Each diamond plots the point estimate for β_1 estimated using specification (2). The line corresponding to each point estimate reflects the 95% confidence interval for the outcome. Standard errors are clustered by community clinic. Each outcome variable has been standardized using the mean and standard deviation of the control group at endline. All regressions include the outcome variable at baseline (except socio-emotional skills), union (strata) and endline-survey-month fixed effects, and the following: child's gender and age in months; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Figure 2: Direct Impacts on Anthropometric Outcomes



Notes: Each diamond plots the point estimate for β_1 estimated using specification (2). The line corresponding to each point estimate reflects the 95% confidence interval for the outcome. Standard errors are clustered by community clinic. Each outcome variable has been standardized using the mean and standard deviation of the control group at endline. All regressions include the outcome variable at baseline, union (strata) and endline-survey-month fixed effects, and the following: child's gender and age in months; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Figure 3: Mechanisms



Notes: Each diamond plots the point estimate for β_1 estimated using specification (2). The line corresponding to each point estimate reflects the 95% confidence interval for the outcome. Standard errors are clustered by community clinic. Each outcome variable has been standardized using the mean and standard deviation of the control group at endline. All regressions include the outcome variable at baseline (except variables representing NNS program take-up), union (strata) and endline-survey-month fixed effects, and the following: child's gender and age in months; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Table 1: Balance on Observables at Baseline

Variable	(1) Control N/[Clusters]	(1) Control Mean/SE	(2) Treatment N/[Clusters]	(2) Treatment Mean/SE	T-test P-value (1)-(2)
Cognitive composite score	1287	98.757	1287	99.751	0.505
	[39]	(1.079)	[39]	(1.033)	
Language composite score	1287	95.526	1287	96.921	0.358
	[39]	(1.004)	[39]	(1.140)	
Height-for-Age (z-score)	1279	-1.377	1278	-1.329	0.472
	[39]	(0.043)	[39]	(0.050)	
Stunted	1279	0.287	1278	0.272	0.539
	[39]	(0.017)	[39]	(0.017)	
Weight-for-Height (z-score)	1275	-0.356	1266	-0.328	0.660
	[39]	(0.049)	[39]	(0.041)	
Wasted	1275	0.067	1266	0.069	0.845
	[39]	(0.008)	[39]	(0.007)	
Weight-for-Age (z-score)	1286	-0.993	1287	-0.956	0.537
	[39]	(0.043)	[39]	(0.043)	
Underweight	1286	0.199	1287	0.186	0.455
	[39]	(0.014)	[39]	(0.010)	
Female child	1287	0.491	1287	0.474	0.416
	[39]	(0.018)	[39]	(0.011)	
Child age (months)	1287	11.796	1287	11.328	0.153
	[39]	(0.261)	[39]	(0.197)	
Mother's age (years)	1283	25.610	1286	25.631	0.911
	[39]	(0.147)	[39]	(0.121)	
Mother employed	1283	0.060	1284	0.050	0.500
	[39]	(0.011)	[39]	(0.010)	
Mother is Muslim	1236	0.845	1250	0.875	0.416
	[39]	(0.026)	[39]	(0.025)	
Mother's weight (kg)	1252	47.076	1251	47.697	0.373
	[39]	(0.489)	[39]	(0.498)	
Single parent household	1287	0.172	1287	0.169	0.914
	[39]	(0.014)	[39]	(0.016)	
Father present 6+ months	1286	0.757	1283	0.767	0.767
	[39]	(0.023)	[39]	(0.027)	
Wealth Index	1266	-0.000	1267	0.000	0.997
	[39]	(0.060)	[39]	(0.059)	
Time to the nearest clinic (minutes)	1287	28.684	1287	30.243	0.395
	[39]	(1.182)	[39]	(1.404)	
Household size	1287	5.938	1287	6.017	0.550
	[39]	(0.096)	[39]	(0.092)	
Household members 0-18 years old	1287	2.818	1287	2.927	0.197
	[39]	(0.052)	[39]	(0.066)	
Mother has primary education or less	1284	0.379	1287	0.364	0.660
	[39]	(0.023)	[39]	(0.025)	
Father has primary education or less	1067	0.564	1069	0.561	0.936
	[39]	(0.024)	[39]	(0.027)	
Present at endline	1287	0.960	1287	0.971	0.254
	[39]	(0.006)	[39]	(0.007)	
F-test of joint significance (p-value)					0.203

Notes: The values displayed for t- and F-tests are p-values. Standard errors are clustered by community clinic. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 2: Treatment Impact on Service Provider Training

	(1) Attended training on ECD	(2) Received training on child health	(3) Received training on child feeding and nutrition	(4) Received other child-related training
Treatment	0.984***	-0.065**	0.011	-0.011
	(0.015)	(0.028)	(0.008)	(0.008)
Adjusted R^2	0.960	0.211	0.021	0.119
Control Mean	0.000	0.056	0.989	0.011
Observations	201	200	201	201

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses and clustered by community clinic. Union (strata) fixed effects are included in all regressions.

Table 3: Treatment Impact on Service Provider Perceptions

	How important is it to teach mothers how to/what:				
	(1) Talk with Children	(2) Respond to Children's Cues	(3) Play Games	(4) Child Health	(5) Food to Feed Children
Treatment	1.536***	1.528***	1.643***	0.700*	0.736**
	(0.369)	(0.362)	(0.404)	(0.359)	(0.350)
Pseudo R^2	0.26	0.20	0.27	0.19	0.19
Control Mean	2.484	2.341	2.297	2.593	2.549
Observations	201	201	201	201	201

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses and clustered by community clinic. Ordered logit regressions run with responses recorded on a 3-point scale where 1 = Neither important nor unimportant, 2 = Important, and 3 = Very important. Union (strata) fixed effects are included in all regressions.

Table 4: Reallocation of Service Provider Time Spent with Households

	Number of minutes spent on:							Number of topics:
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ECD	General Nutrition	Child Health	Family Planning	EPI	Pregnant Women	Total	Number of Topics Discussed
Treatment	3.010***	-0.805***	-0.503**	-0.215	-0.506**	-0.268	0.714	0.356**
	(0.271)	(0.209)	(0.215)	(0.369)	(0.212)	(0.335)	(0.544)	(0.138)
Adjusted R^2	0.608	0.156	0.094	-0.090	-0.110	0.309	0.276	0.315
Control Mean	0.038	2.25	1.135	2.827	1.692	3.423	11.365	3.654
Observations	122	122	122	122	122	122	122	122

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses and clustered by community clinic. Units for dependent variables: minutes per household visit in the last working day. Union (strata) fixed effects are included in all regressions.

Table 5: Household-reported health-worker visits and their duration and content

	Frequency of FWA/HA visits in the last 6 months		Duration (last visit)	Topic discussed in the last visit				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	At least once	No of total visits	Minutes	Playing with child	Talking to child	Child development card	Picture books	Key messages booklet
Treatment	0.074**	0.127	0.342	0.334***	0.319***	0.442***	0.422***	0.253***
	(0.034)	(0.315)	(1.263)	(0.068)	(0.063)	(0.068)	(0.068)	(0.045)
Adjusted R-squared	0.264	0.254	0.114	0.337	0.322	0.351	0.352	0.283
Endline Control Mean	0.186	0.924	11.550	0.214	0.153	0.013	0.017	0.009
Observations	2486	2486	593	590	590	590	590	586
Effect Size	0.191	0.051	0.052	0.813	0.883	3.877	3.217	2.686

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered by community clinic. Effect sizes are estimated using the endline control SD. All regressions include union (strata) and endline-survey-month fixed effects, and the following: child's gender and age in months; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Table 6: Receipt of Materials by Households

	Mother/Household received				
	(1)	(2)	(3)	(4)	(5)
	Child Development Card	Household Picture Book	Nature Picture Book	Key Message Booklet	Any of the four SC materials
Treatment	0.491***	0.489***	0.484***	0.164***	0.497***
	(0.034)	(0.035)	(0.035)	(0.018)	(0.035)
Adjusted R^2	0.356	0.355	0.349	0.152	0.364
Control Mean	0.024	0.026	0.026	0.011	0.026
Observations	2479	2479	2479	2479	2479

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered by community clinic. Effect sizes are estimated using the endline control SD. All regressions include union (strata) and endline-survey-month fixed effects, and the following: child's gender and age in months; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Table 7: Intent to Treat Impacts on Primary Outcomes of Interest

Panel A: Cognitive and Language Development					
	(1) Cognitive composite score	(2) Language composite score	(3) Expressive communication scaled score	(4) Receptive communication scaled score	
Treatment	1.202***	2.240***	0.308***	0.477***	
	(0.384)	(0.493)	(0.078)	(0.101)	
Adjusted R-squared	0.211	0.216	0.197	0.185	
Endline Control Mean	84.316	88.384	7.681	8.272	
Observations	2484	2484	2484	2484	
Effect Size	0.168	0.227	0.180	0.240	
Panel B: Wolke Behavioral Rating Scales					
	(1) Approach	(2) Emotion	(3) Activity	(4) Cooperation	(5) Vocalization
Treatment	0.197***	0.167**	0.199***	0.061	0.077
	(0.071)	(0.073)	(0.074)	(0.070)	(0.081)
Adjusted R-squared	0.232	0.160	0.180	0.130	0.129
Endline Control Mean	4.858	5.163	3.312	5.192	4.418
Observations	2485	2485	2485	2485	2485
Effect Size	0.129	0.119	0.138	0.044	0.044

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered by community clinic. Effect sizes are estimated using the endline control SD. All regressions include the outcome variable at baseline (except socio-emotional skills in Panel B), union (strata) and endline-survey-month fixed effects, and the following: child's gender and age in months; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Table 8: Intent to Treat Impacts on Anthropometric Outcomes

	(1) Height-for-age z-score	(2) Percent stunted	(3) Weight-for-height z-score	(4) Percent wasted	(5) Weight-for-age z-score	(6) Percent under-weight
Treatment	-0.024	0.023	0.230***	-0.048***	0.137***	-0.029*
	(0.035)	(0.015)	(0.042)	(0.009)	(0.025)	(0.014)
Adjusted R-squared	0.457	0.248	0.207	0.049	0.447	0.247
Endline Control Mean	-1.848	0.441	-0.867	0.118	-1.644	0.330
Observations	2482	2482	2481	2481	2485	2485
Effect Size	-0.022	0.047	0.217	-0.149	0.140	-0.061

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered by community clinic. Effect sizes are estimated using the endline control SD. All regressions include the outcome variable at baseline, union (strata) and endline-survey-month fixed effects, and the following: child's gender and age in months; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Table 9: Treatment Spillovers on Siblings

	5-7 years old	10-14 years old			
	(1) Attending school	(2) Attending school	(3) Completed years of schooling	(4) Can write a letter	(5) Not in the labor force
Treatment	0.062**	0.010	0.014	0.000	0.015**
	(0.027)	(0.012)	(0.064)	(0.031)	(0.007)
Adjusted R^2	0.187	0.423	0.766	0.475	0.163
Endline Control Mean	0.650	0.942	4.114	0.698	0.981
Observations	749	804	800	804	803
Effect Size	0.130	0.041	0.008	0.000	0.112

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered by community clinic. Sample comprises siblings of study children. Effect sizes are estimated using the endline control SD. All regressions include outcome variable at baseline, union (strata) and endline month fixed effects, and the following covariates: sibling's gender, birth order, and age in years; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Table 10: Determinants of Compliance Within Treatment Group

	(1) Months household owned Childhood development card	(2) Months household owned at least one of the four materials
Female child	0.063	0.005
	(0.275)	(0.023)
Child age (months)	-0.078 (0.058)	-0.006 (0.005)
Mother age (years)	0.007 (0.053)	-0.001 (0.004)
Mother employed	-1.342	-0.138^{*}
	(0.958)	(0.076)
Mother is Muslim	0.383	-0.001
	(0.758)	(0.060)
Mother's weight (kg)	0.006 (0.023)	0.001 (0.002)
Single parent household	-0.357 (0.655)	0.000 (0.000)
Father present 6+ months	-0.077 (0.569)	-0.015 (0.042)
Household size	0.134	0.012
	(0.155)	(0.011)
Household members 0-18 years old	-0.178 (0.230)	-0.008 (0.017)
Wealth Index	-0.070	-0.011
	(0.274)	(0.019)
Time to the nearest clinic (minutes)	-0.058^{***} (0.016)	-0.004^{***} (0.001)
Mother has primary education or less	-0.119 (0.525)	0.000 (0.040)
Father has primary education or less	-0.689^{*} (0.403)	-0.042 (0.030)
Share of HHs treated	0.53	0.53
Adjusted R-squared	0.167	0.171
Observations	1250	1248

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered by community clinic. All regressions include union (strata) and endline-month fixed effects.

Table 11: Dose-response Impacts on Child Development Outcomes

	(1) Cognitive composite score	(2) Language composite score	(3) Approach	(4) Emotion	(5) Activity	(6) Weight-for-height z-score	(7) Weight-for-age z-score
Months household owned	0.195***	0.364***	0.032***	0.027**	0.032***	0.037***	0.022***
Childhood development card	(0.060)	(0.078)	(0.012)	(0.012)	(0.012)	(0.007)	(0.004)
Endline Control Mean	84.316	88.384	4.858	5.163	3.312	-0.867	-1.644
Observations	2484	2484	2485	2485	2485	2481	2485
Effect Size (12 months)	0.328	0.443	0.252	0.233	0.269	0.424	0.273
Weak-IV robust F statistic	189.21	188.65	189.51	189.51	189.51	187.87	188.75
C-statistic p-value	0.068	0.010	0.068	0.111	0.033	0.000	0.000

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered by community clinic. Effect sizes are estimated using the endline control SD. Two-stage least-squares (2SLS) regressions shown. All regressions include outcome variable at baseline, union (strata) and endline-month fixed effects, and the following: child's gender and age in months; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Table 12: Contribution of Mediators to ITT Impacts on Primary Outcomes

	(1) Bayley Index	(2) Wolke Index
Receipt of Any Program Materials	33.5%	0%
Maternal Agency on Child Stimulation	0%	0%
Complementary parental investments (Mother's variety of learning activities)	7.5%	14.2%
Complementary NNS Take-up (Number of growth monitoring check-ups)	0%	0%

Percentage contributions are estimated using the four-step algorithm outlined in Section 6.4. Statistically insignificant point estimates with $p > 0.1$ are set to zero. All regressions include outcome variable at baseline (excluding the Wolke index in column 2), union (strata) and endline month fixed effects, and the following: child's gender and age in months; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Table 13: Heterogeneity in Treatment Impacts by Household Characteristics

Panel A: Household Wealth

	(1) Cognitive composite score	(2) Language composite score	(3) Approach	(4) Emotion	(5) Activity	(6) Weight-for-height z-score	(7) Weight-for-age z-score
Treatment	1.344***	2.374***	0.149**	0.148*	0.135*	0.203***	0.117***
	(0.438)	(0.550)	(0.072)	(0.084)	(0.075)	(0.049)	(0.028)
Wealth Index	0.581**	0.922**	0.135***	0.051	0.030	0.055*	0.080**
	(0.241)	(0.350)	(0.044)	(0.052)	(0.049)	(0.032)	(0.031)
Treatment∗Wealth Index	-0.064	0.093	-0.126**	-0.013	-0.019	-0.107**	-0.094**
	(0.366)	(0.492)	(0.062)	(0.067)	(0.069)	(0.042)	(0.037)
Observations	1965	1965	1965	1965	1965	1938	1965

Panel B: Child Gender

	(1) Cognitive composite score	(2) Language composite score	(3) Approach	(4) Emotion	(5) Activity	(6) Weight-for-height z-score	(7) Weight-for-age z-score
Treatment	1.566***	2.615***	0.162*	0.188*	0.027	0.202***	0.122***
	(0.548)	(0.704)	(0.093)	(0.100)	(0.088)	(0.066)	(0.045)
Female child	-0.301	0.188	-0.040	-0.011	-0.307***	-0.060	-0.175***
	(0.387)	(0.537)	(0.090)	(0.076)	(0.067)	(0.061)	(0.053)
Treatment∗Female child	-0.434	-0.515	0.008	-0.078	0.224**	0.030	0.015
	(0.581)	(0.715)	(0.117)	(0.102)	(0.105)	(0.084)	(0.074)
Observations	1965	1965	1965	1965	1965	1938	1965

Panel C: Maternal Education

	(1) Cognitive composite score	(2) Language composite score	(3) Approach	(4) Emotion	(5) Activity	(6) Weight-for-height z-score	(7) Weight-for-age z-score
Treatment	1.396***	2.900***	0.182*	0.188*	0.195**	0.239***	0.088**
	(0.497)	(0.729)	(0.096)	(0.098)	(0.090)	(0.063)	(0.035)
Mother has primary education or less	-1.009**	-0.983	-0.164	-0.151	-0.082	0.109	-0.045
	(0.453)	(0.647)	(0.101)	(0.097)	(0.083)	(0.067)	(0.052)
Treatment∗Mother has primary education or less	-0.110	-1.344	-0.040	-0.095	-0.143	-0.056	0.104
	(0.647)	(0.958)	(0.135)	(0.125)	(0.122)	(0.090)	(0.067)
Observations	1965	1965	1965	1965	1965	1938	1965

Panel D: Paternal Education

	(1) Cognitive composite score	(2) Language composite score	(3) Approach	(4) Emotion	(5) Activity	(6) Weight-for-height z-score	(7) Weight-for-age z-score
Treatment	1.408**	3.108***	0.139	0.166	0.129	0.117*	-0.001
	(0.550)	(0.732)	(0.098)	(0.102)	(0.091)	(0.070)	(0.047)
Father has primary education or less	-0.660	-0.994	-0.183*	-0.118	-0.073	-0.160**	-0.200***
	(0.526)	(0.684)	(0.094)	(0.098)	(0.090)	(0.061)	(0.056)
Treatment∗Father has primary education or less	-0.098	-1.329	0.047	-0.030	0.014	0.178**	0.232***
	(0.682)	(0.916)	(0.116)	(0.122)	(0.111)	(0.070)	(0.060)
Observations	1965	1965	1965	1965	1965	1938	1965

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered by community clinic. All regressions include outcome variable at baseline (excluding socio-emotional skills i.e., for models in columns 3-5), union (strata) and endline month fixed effects, and the following: child's gender and age in months; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Table 14: Heterogeneity in Treatment Impacts by Service Provider Characteristics

Panel A: Service Provider Education							
	(1) Cognitive composite score	(2) Language composite score	(3) Approach	(4) Emotion	(5) Activity	(6) Weight-for-height z-score	(7) Weight-for-age z-score
Treatment	3.077***	4.102***	0.224	0.272	0.091	-0.138*	-0.020
	(0.790)	(1.054)	(0.201)	(0.186)	(0.167)	(0.072)	(0.052)
Above-median Schooling	1.497	1.747	-0.063	-0.009	0.131	-0.234***	-0.023
	(0.911)	(1.159)	(0.185)	(0.182)	(0.167)	(0.076)	(0.055)
Treatment ∗ Above-median Schooling	-2.115**	-1.627	0.111	0.011	0.306	0.479***	0.245***
	(1.062)	(1.332)	(0.240)	(0.210)	(0.230)	(0.116)	(0.078)
Observations	2421	2421	2422	2422	2422	2418	2422

Panel B: Service Provider Experience							
	(1) Cognitive composite score	(2) Language composite score	(3) Approach	(4) Emotion	(5) Activity	(6) Weight-for-height z-score	(7) Weight-for-age z-score
Treatment	0.866	2.987***	0.316***	0.326***	0.453***	0.355***	0.251***
	(0.882)	(0.994)	(0.108)	(0.123)	(0.126)	(0.064)	(0.041)
Above-median Experience	-2.054**	-0.347	-0.096	0.202	0.172	0.093	0.024
	(0.930)	(1.223)	(0.164)	(0.230)	(0.145)	(0.106)	(0.054)
Treatment ∗ Above-median Experience	0.676	-0.930	-0.184	-0.258*	-0.432***	-0.298***	-0.218***
	(0.947)	(1.194)	(0.156)	(0.150)	(0.162)	(0.091)	(0.055)
Observations	2421	2421	2422	2422	2422	2418	2422

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered by community clinic. All regressions include outcome variable at baseline (excluding socio-emotional skills i.e., for models in columns 3-5), union (strata) and endline month fixed effects, and the following: clinic-level means of service provider's age, training (no of topics), and experience in the current union; number of local service providers surveyed; child's gender and age in months; mother's age, weight, religion, education, and employment status; father's education and whether the father was present for 6 months or more in the last year; household composition (single parent household, household size, number of people in the household between 0-18 years old); an asset index; and time to travel to the nearest community clinic (in minutes).

Table 15: Intervention Cost Categories

Type of Cost	Cost	Per Clinic	Per Service Provider	Per Child	Percentage of Total
Personnel	45,589	1,169	390	2.45	35.7%
Administration	25,237	647	216	1.35	19.8%
Training	24,534	629	210	1.32	19.2%
Knowledge sharing, incl. program materials	21,312	546	182	1.14	16.7%
Travel and accommodation	10,862	279	93	0.58	8.5%
Total	**127,534**	**3,270**	**1,090**	**6.84**	**100%**

Notes: All costs are adjusted for inflation to report values in 2014 U.S. dollars.

Table 16: At-scale ECD Studies & Comparison of ITT Impacts

Study	Country	Design	Intervention	Estimated Impact
Attanasio et al. (2022)	Colombia	N = 1,460 T = 10 months	ECS with materials, training of mothers, monthly nutritional supplement	Cognition: 0.16 SD
Attanasio et al. (2014)	Colombia	N = 1,440 T = 18 months	CCT program delivered by hired community mothers	Cognition: 0.26 SD, Receptive language: 0.22 SD
Yousafzai et al. (2014)	Pakistan	N = 1,489 T = 24 months	Built on Lady Health Worker program, additional $ provided	Cognition: 0.6 SD, Language: 0.7 SD, Motor development: 0.5 SD
Sylvia et al. (2020)	China	N = 592 T = 6 months	Delivered by Family Planning Commission cadres	Infant skill: 0.25 SD
Araujo et al. (2021)	Peru	N = 4,685 T = 24 months	CCT program & delivered by paid para professionals	Infant skill: 0.10 SD
Ganimian et al. (2021)	India	N = 4,675 T = 16 months	Built on India's Integrated Child Development Services, additional $ provided	Executive function: 0.18 SD