Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia

Orazio Attanasio, Sarah Cattan, Emla Fitzsimons,

Costas Meghir, and Marta Rubio-Codina*

November 21, 2019

ONLINE APPENDIX

*Attanasio: Department of Economics, Yale University, Institute for Fiscal Studies, FAIR and NBER (email: orazio.attanasio@yale.edu); Cattan: Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, and IZA Institute of Labor economics (email: sarah_c@ifs.org.uk); Fitzsimons: UCL Institute of Education, 20 Bedford Way, London WC1H 0AL, and Institute for Fiscal Studies (e.fitzsimons@ucl.ac.uk); Meghir: Department of Economics, Yale University, 28 Hillhouse Avenue, New Haven, CT 06511, National Bureau of Economic Research, IZA Institute of Labor economics and Institute for Fiscal Studies (email: c.meghir@yale.edu); Rubio-Codina: Inter-American Development Bank,1300 New York Avenue, N.W. Washington, D.C. 20577 (email: martarubio@iadb.org). We thank four anonymous referees and the editor for detailed comments. We have benefited from extensive discussions with Don Andrews, Joseph Altonji and James Heckman. We are grateful to Sally Grantham McGregor who participated in designing and implementing the experiment we describe here. We also thank participants at the NBER Summer Institute, Barcelona GSE Summer Forum and Montreal CIREQ Applied Microeconomics on Fertility and Child Development and seminars at Stanford University, University of Chicago, Oxford University, Cornell University, Bristol University and the Institute for Fiscal Studies for their comments. The authors gratefully acknowledge the support of the British Academy (grant reference pf140104), the NIH (grant reference R01 HD072120), the Economic and Social Research Council's Centre for the Microeconomic Analysis of Public Policy at the IFS (grant reference ES/M010147/1), the Cowles Foundation and the ISPS at Yale. Attanasio's time was partly funded by an European Research Council (ERC) Advanced Grant (Grant Agreement No 695300). This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No 646917, 240910, and 249612). The authors thank the Economic and Social Research Council (grant RES-062-23-1548), Inter-American Development Bank, World Bank, and International Growth Center for funding the intervention and data collection. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. All errors are the responsibility of the authors.

A The Intervention

This appendix provides a detailed description of the intervention, its implementation and roll-out. The material in this section draws heavily on Attanasio et al. (2014).

A.1 Description of the Intervention

The integrated early child development intervention was designed so that it could be delivered by local people, readily identifiable through administrative infrastructures of social welfare systems (the conditional cash transfer program *Familias en Accion* (FeA) in our case). The intervention included psychosocial stimulation on its own, micronutrient supplementation on its own, and both combined. Each of these arms had 24 clusters (municipalities).

A.1.1 Psychosocial Stimulation

The psychosocial stimulation component was inspired by and based on the Jamaican home visiting model (Grantham-McGregor et al., 1991), the overarching aim of which is to facilitate developmentally appropriate learning activities between mothers (primary caregivers) and their children through demonstration of play activities centred around daily routines. Play activities draw on resources in the home, low cost home-made toys and the intervention toy kit. The toy kit included picture books, naming plates, conversation scenes, puzzles, lotteries, and blocks. Play activities followed steps aimed towards children's gradual mastery of a learning objective: (1) child observes the play activity (modeling), (2) mother and child do the activity together, (3) child attempts the play activity on his/her own, (4) mother prompts naming and verbalisation of objects and actions linked to the play activity, (5) the developmental level of the play activity is adjusted by increasing or decreasing the challenge based on the child's performance (scaffolding). Throughout the play activities, mothers are encouraged to provide children with contingent positive reinforcement for progress toward the learning goals (praise) and to follow the child's interest.

Materials for the stimulation component were adapted from the Jamaican intervention guide to the local cultural context and to the average educational level of home visitors and program beneficiaries. Such adaptations included (1) inclusion of local songs and rhymes, (2) modification of the home-made toys instruction manual to use local recyclable materials, (3) incorporation into the intervention toy kit of culturally relevant pictures, scenes and objects, (4) re-organization of the psychosocial stimulation guide in weekly instruction cards for specific age groups, and (5) ordering of play materials, via an index, by developmental stage to facilitate the scaffolding of activities.

Home visits lasted approximately one hour. Home visits took place with the child's biological mother or primary caregiver. Other adults in the household, where present, were also encouraged to participate in the home visit. In advance of the visit, the home visitor selected the weekly instruction card from the psychosocial stimulation guide according to the appropriate developmental level of the target child, and prepared the toys and materials for the visit. The home visit had three parts. At the beginning, the home visitor did an informal assessment of the child's progress in the sequence of play activities, by asking about the play activities the mother and child practiced during the previous week and identifying where challenges arose (or on the contrary, where activities were overly simplistic). In the second and main part of the visit, new play activities were shown and practised. At the end, the mother was prompted to summarize the play activities and agree a plan on how to practice them over the following week.

A.1.2 Micronutrient Supplementation

The micronutrient supplementation component consisted of micronutrient supplementation in the form of sprinkles (encapsulated micronutrients). Each single-dose sachet contains 12.5 mg iron, 5 mg zinc, vitamin A 300 mg RE, 160mg g folic acid and 30 mg vitamin C and each displayed a pictorial representation of use. In addition to the fortnightly provision of sprinkles, participating families received a booklet with detailed instructions for use and

storage and daily record forms to track use. Families were provided with enough sachets for all children below six years of age to prevent sharing with siblings.

A.2 Implementation

We obtained institutional cooperation from the National Director of FeA, the coordinator of the Strategic Planning and Monitoring Unit of FeA, and FeA administrative staff at the municipality level. From the latter, we obtained rosters of female community representatives (Madres Lideres) and selected three female community representatives in each of the 96 target municipalities.

In municipalities assigned to receive stimulation, 63% of selected female community leaders took on the role of home visitor. The remainder declined due to other work commitments or not meeting the minimum reading comprehension criteria (established using a short reading comprehension test designed by the study data collection team). Replacements were found through referral; women who were referred were screened for reading comprehension skills, motivation and availability. If they met established criteria, they were offered the opportunity to become home visitors.

The intervention staff included a local field coordinator, six mentors, and a team of home visitors. The field coordinator oversaw the roll out of the intervention and monitored the mentors from the central office (in Bogota) and also on site. Six mentors, with an undergraduate degree in psychology/social work or comparable fieldwork experience, were recruited to train and supervise home visitors throughout the study. They had six weeks' pre-service training focused on the home visiting curriculum and protocols, training and supervision skills, creating home-made toys, and supervised practice. Four short (one- to four-day) refresher and feedback sessions took place in Bogota during the course of the intervention. These also provided the mentors the opportunity to exchange experiences, challenges and solutions. In addition, the mentors were in regular email/phone communication with the field coordinator, and email communication with a member of the research team, as needed.

Each mentor trained and supervised 24 home visitors, covering eight municipalities. The home visitors' pre-service training on the stimulation component lasted two weeks, with an additional week of in-service training after the team of mentors had completed the first round of itinerant supervision (one to two months after the intervention roll-out). The home visitor training sessions included: (1) basic concepts of child development and early learning; (2) the aims of the stimulation intervention and the role of the home visitor in this; (3) an introduction to play activities, the steps to demonstrate and scaffold each activity and guidelines to assess the progress of the child towards the learning goals; (4) guidelines to help manage difficult children; (5) a toy making workshop; (6) guidelines for record keeping. The training of home visitors emphasized the importance of a good working relationship with the beneficiary mother, positive reinforcement, and listening skills.

In addition, home visitors participated in a 5-hour training session on the micronutrient component of the intervention before rollout. The training sessions included a description of the micronutrient supplement, storage requirements, instructions for use, potential side effects, toxicity risks, and safety protocols in case of side effects.

A.3 Rollout and Monitoring

The intervention was rolled out over four months from February through May 2010, and phased out 18 months later, from September through December 2011. Training and supervision was rolled out by geographical location, evenly across treatment groups, following baseline data collection. Once the intervention was up and running, mentors visited intervention communities once every 7 to 10 weeks to monitor implementation, provide support, and reinforce home visitors' motivation. At this time, mentors also distributed one-page bulletins to home visitors, with reminders of best practices in home visiting. In addition, mentors sent short text (SMS) messages to home visitors every month to reinforce key aspects of the stimulation protocol. Home visitors were encouraged to call mentors for advice where necessary (calling cards were provided by the study team).

To monitor the psychosocial stimulation component, home visitors filled in a form at the end of each visit with basic information on the visit (e.g. date, activities performed, who present), as well as a short assessment of children's performance and engagement with the activities. Home visitors were paid $100,000 Colombian pesos (COP) (19.4% of legal monthly minimum wage for 2010) per month.

To monitor the micronutrient component, the home visitor collected the empty sachets and intake charts (monitoring forms) from the household every two weeks. During these visits, mentors checked that tracking charts were filled in correctly, reminded the mother about the protocols of use, how to react if side effects arose and addressed any other queries and concerns. The home visitor was paid $25,000 Colombian pesos (COP) per month for these activities.

A.4 Balance

Table A.1 shows the characteristics of children who remained in the study at follow-up, their mothers, and their households in the group who received any stimulation (alone or in combination with the micronutrient supplementation) and in the group who received none (including those who received the micronutrient supplementation alone). No differences were apparent between the groups, with the exception of the fact that "no stimulation group" had a higher proportion of first borns than the group who received some stimulation.

Table A.1: Baseline characteristics of participant children, their mothers and their households by treatment status

	Control		Treatment		Difference		
	Mean	Std. Dev	Mean	Std. Dev	Mean	p-val	RW p-val
Child:							
Boy	0.510		0.481		0.029	0.267	0.988
Age (months)	18.02	(3.799)	18.04	(3.721)	-0.019	0.924	0.988
Premature	0.098		0.074		0.024	0.107	0.988
Mean birth weight (g)	3236	(526.4)	3246	(499.8)	-9.924	0.726	0.988
Stunted	0.141		0.134		0.007	0.710	0.988
First born	0.420		0.367		0.053	0.044	0.960
Bayley: cognitive	-0.005	(1.003)	-0.034	(1.049)	0.029	0.592	0.988
Bayley: receptive language	-0.009	(0.999)	-0.006	(0.934)	-0.003	0.949	0.988
Bayley: expressive language	-0.020	(0.992)	0.004	(1.018)	-0.024	0.655	0.988
Bayley: fine motor	-0.006	(1.003)	-0.066	(1.094)	0.060	0.284	0.988
MacArthur: words the child can say	-0.018	(0.933)	0.003	(0.974)	-0.021	0.685	0.988
ICQ: difficult (-)	-0.002	(1.001)	0.062	(1.000)	-0.064	0.228	0.988
ICQ: unsociable (-)	0.001	(0.997)	-0.035	(1.002)	0.037	0.489	0.988
ICQ: unadaptable (-)	0.001	(0.998)	-0.035	(1.036)	0.037	0.497	0.988
ICQ: unstoppable (-)	0.006	(0.996)	0.092	(1.026)	-0.087	0.107	0.988
Mother:							
Education (years)	7.575	(3.534)	7.351	(3.423)	0.224	0.236	0.988
Married	0.690		0.673		0.017	0.490	0.988
Age (years)	26.04	(0.253)	26.50	(0.250)	-0.463	0.193	0.988
CESD: bothered by things? (-)	0.006		0.085		-0.079	0.135	0.988
CESD: trouble keeping mind on doing? (-)	0.001		0.053		-0.052	0.331	0.988
CESD: feel depressed? (-)	0.002		0.156		-0.155	0.004	0.291
CESD: everything was an effort? (-)	-0.004		0.136		-0.140	0.009	0.817
CESD: feel fearful? (-)	0.007		-0.001		0.008	0.880	0.988
CESD: was sleep restless? (-)	0.001		0.070		-0.070	0.195	0.988
CESD: feel happy?	-0.001		0.066		-0.067	0.206	0.988
CESD: feel lonely? (-)	0.005		0.090		-0.085	0.112	0.988
CESD: feel you couldn't get going? (-)	0.005		0.055		-0.050	0.369	0.988
Household:							
No of rooms/household size	0.594	(0.281)	0.591	(0.310)	0.003	0.841	0.988
Wealth index	-0.010	(0.993)	0.016	(1.006)	-0.027	0.614	0.988
Home ownership	0.370		0.364		0.006	0.809	0.988
FCI: types of play materials	3.708	(1.688)	3.644	(1.697)	0.064	0.477	0.988
FCI: types of play activities in last 3 days	3.256	(1.609)	3.223	(1.513)	0.033	0.688	0.988
FCI: no. of books for adults	1.071	(1.044)	1.047	(1.059)	0.023	0.675	0.988
FCI: no. of magazines and newspapers	0.984	(1.140)	0.960	(1.160)	0.025	0.684	0.988
Log village-level toy prices	8.072	(0.189)	7.998	(0.221)	0.074	0.000	0.964
Log village-level food prices	8.075	(0.147)	8.102	(0.135)	-0.026	0.001	0.988
Maternal childhood exposure to conflict	0.056	(0.081)	0.069	(0.094)	-0.013	0.005	0.988

Note: "pval" refer to single hypothesis testing p-values and "RW pval" refer to stepdown p-values accounting for multiple hypotheses testing as in Romano and Wolf (2005).

B Measures of skills and investments

In this section, we provide detailed information on each of the instruments we used to measure children's cognitive and socio-emotional skills, mother's cognitive and socio-emotional skills, and parental investments. In subsection B.4, we describe the non-parametric procedure we followed to standardize the measures for age.

B.1 Measures on the target child

B.1.1 Bayley Scales of Infant and Toddler Development, 3rd edition [Bayley]

We administered the cognitive, receptive language, expressive language, fine motor and gross motor scales of the 3rd edition of the Bayley Scales of Infant and Toddler Development (Bayley-III) developed by Bayley (2006) both at baseline and follow-up, following standard procedures. The scales assess children from birth up to 42 months by direct observation of performance on a series of items and are considered by many the "gold standard" for the assessment of children of these ages (Fernald et al., 2009). Bayley-III subscales were translated into Spanish, back translated to English to ensure accuracy, and piloted by testers. Children were assessed in local community centers with their mothers present. Testers held degrees in psychology and had a six-week training, including practice sessions with children of the target age groups. Inter-rater reliability (intra-class correlation) was above 0.9 on each subscale. Furthermore, 5% of the measurements were supervised by the trainer (reliabilities above 0.9) and corrective feedback was given when appropriate.

B.1.2 MacArthur-Bates Communicative Development Inventories I, II and III [MacArthur]

We assessed language comprehension and production using the Spanish short-form versions of the MacArthur-Bates Communicative Development Inventories I, II and III (SFI, SFII, SFIII). This is a parent report inventory and was collected in the house as part of the

household survey. At baseline, we administered Spanish short forms of Inventories I and II to children of 12-18 and 19-24 months of age, respectively, which have been validated in Mexico (Jackson-Maldonado, Marchman, and Fernald (2012), Jackson-Maldonado et al. (2003)). At follow-up, we administered the Spanish short form of Inventory III to children 30-42 months. This form was under validation at the time of data collection (Jackson-Maldonado and Conboy (2011), Jackson-Maldonado (2011)). We collaborated with the developer of the test in Spanish in the identification of suitable words in Colombian Spanish, prior and during piloting activities. We administered the vocabulary checklist (words the child "understands" and words the child "can say") for all Short Forms (SFs) and sentence structure sections (for SFIII only), and counted the number of words the child could say (as reported by the mother/caregiver) and number of more complex sentence structures the child uses.

B.1.3 Infant Characteristics Questionnaire [ICQ]

The Infant Characteristics Questionnaire (Bates, Freeland, and Lounsbury, 1979) assesses the construct of "difficult" temperament by maternal (caregiver) report. As such it measures parents' perceptions of the infant, not necessarily the infant's behavior as it might be objectively recorded. Both at baseline and follow-up, we used those items (17 in total) in questionnaires for 13-months and 24-months old children that related to the following constructs: difficult, unadaptable, unstoppable, and unsociable, as part of the household questionnaire. We made minor adjustments to the Spanish translations of the forms in order to maximize comprehension and cultural adequacy. For simplicity, we converted the 7-point rating items into 5-point ratings. We discussed these modifications with the author over email correspondence and piloted them before use in the field.

B.1.4 Early Children's Behavior Questionnaire [ECBQ]

At follow up, we complemented the assessment of temperament with measures of attention and inhibitory control by maternal report using the attentional focusing (sustained duration

of orienting on an object of attention; resisting distraction), attentional shifting (the ability to transfer attentional focus from one activity/task to another) and inhibitory control (the capacity to stop, moderate, or refrain from a behavior under instruction) sub-scales in the short versions of the Spanish translation of the Early Children's Behavior Questionnaire (Putnam, Gartstein, and Rothbart (2006), Putnam et al. (2010)). The ECBQ is designed to measure temperament in children aged 3-7 years. As before, for simplicity, we converted the 7-point rating scale into a 5-point rating scale. Minor language modifications to wording and sentence structure, with the aim to better reflect Colombian Spanish, were extensively piloted in the field.

B.2 Measures on the mother

B.2.1 Maternal vocabulary

We assessed maternal receptive vocabulary in the first follow-up survey. For this, we used a selection of 50 words from the Spanish version of the Peabody Picture Vocabulary Test (PPVT), the Test de Vocabulario en Imagenes de Peabody (TVIP) (Dunn, Padilla, and Lugo, 1986). The reason why we used a selection of the words is because the Spanish version of the test is developed for ages 2.5-18 years. Hence, we could not use the test as designed (using established start and stopping rules) on our sample of mothers. Instead, we selected those words exhibiting a reasonable level of varying difficulty, after extensive piloting, and administered them all in the order in which they appear in the test. For each word, the subject points at the one picture (out of four) that best relates to the word (noun, action, abstract concept, adjective) that the tester calls out. The test was administered in the home by the interviewer at the end of the household interview

B.2.2 Standard Progressive Matrices [RPM]

We used Raven's Standard Progressive Matrices system Raven (1981) to measure mothers' reasoning ability or what is often referred to as general intelligence in the second follow-

up. This is a non-verbal test typically made of multiple choice items listed in order of difficulty, requiring ever greater capacity to encode and analyze the information, and which are organized in sets. In each test item, the subject is asked to identify the missing element that completes a pattern. Test items are presented in black ink on a white background. Out of the 5 sets of 12 matrices, we administered the middle 3 – hence, those with an intermediate level of difficulty. This decision was made after careful piloting.

We collected RPM at second follow up only under the presumption that maternal reasoning ability should not be affected by the intervention, and because of the monetary and time restrictions we faced in previous rounds.

B.2.3 Center for Epidemiological Studies Depression scale (CES-D)

We assessed maternal depressive symptoms by direct interview with the mother using the 10-item Spanish version of the CES-D, the CESD-10, developed by Radloff (1977). We used the same measure both at baseline and follow up.

B.3 Measures of parental investments

The quality of the home environment (or level of stimulation in the home) was measured using items in the Family Care Indicators [FCI] developed by UNICEF (Kariger et al., 2012). The FCI has been validated against the Home Observations for Measurement of the Home Environment (HOME) (Caldwell and Bradley, 2001), against which it was validated in Bangladesh (Hamadani et al., 2010).

Both at baseline and follow up, we collected by direct observation during the household survey the following information: the number of books for adults, the number of newspapers and magazines, and the number of varieties of play materials in the home that the child often played with. We collected by maternal (caregiver) report the number of play activities the child engaged in with an adult over the three days before the interview. Play materials include toys that make/play music; toys/objects meant for stacking, constructing or building;

things for drawing, writing, colouring, and painting; toys for moving around; toys to play pretend games; picture and drawing books for children; and toys for learning shapes and colours. Play activities include reading or looking at picture books; telling stories to child; singing songs with child; playing with child with her toys; spending time with child scribbling, drawing, or colouring; and spending time with child naming things or counting; and taking child out for a leisure walk.

B.4 Age standardization of the measures

Total raw scores are increasing in age. Since we are interested in within sample comparisons, we internally standardize scores to remove the effect of age (child's age for the childs' measures and mother's age for the mothers' measures). We compute internal z-scores using the empirical age-conditional means and standard deviations estimated using non-parametric regression methods. In particular, for each measure to standardize, we compute the age-conditional mean using the fitted values of the following regression, estimated by kernel-weighted local polynomial smoothing methods:

$$Y_i = f(X_i) + \epsilon_i \tag{1}$$

where Y_i is the raw score of individual i in a given scale and X_i is the age of the individual (in months for the child, in years for the mother). Next, we regress the square of the residuals in equation (1) on age of the child as shown in the following kernel-weighted local polynomial regression:

$$(Y_i - \hat{f}_i)^2 = g(X_i) + \upsilon_i \tag{2}$$

Our estimate of the age-conditional standard deviation is the square root of the fitted values in equation (2). Finally, we compute the internally age-adjusted z-score, ZY_i, by subtracting from the raw score the within sample age-conditional mean estimated in (1) and

dividing by the within sample age-conditional standard deviation obtained from (2). More specifically:

$$ZY_i = \frac{(Y_i - \hat{f}_i)}{\sqrt{\hat{g}_i}} \tag{3}$$

This procedure, less sensitive to outliers and small sample sizes within age category, resulted in smoothly distributed internally standardized scores, with mean zero across the age range.

B.5 Instruments

Toy prices and food prices were collected as part of the community questionnaire administered by surveyors to a community leader. These therefore only vary at the village level but not at the individual level within villages. Toy prices were only collected in the second follow-up survey, i.e. 2 years after the end of the intervention, while food prices were collected in every round of data collection. For the analysis presented in this paper, we use food prices collected in the first follow-up, i.e. just after the intervention ended. The questionnaire asks price information about around 40 different items related to children, including books, toys, clothing and health products. We construct the measure of toy prices we use as an instrument in the paper as the average of the log price of all different types of books and toys listed in the questionnaire. For food prices, the questionnaire asks the price about 40 different types of food at the market and at the supermarket. To construct our measure of food prices that we use as an instrument in the paper, we first average the market and supermarket prices of each item and then average these averages across all 40 items. If items are only sold at the market or at the supermarket, we only use the available price.

Our measure of maternal childhood exposure to conflict is constructed using data from the "Panel Municipal del CEDE" created by the Centre of Studies on Economic Development at the University de Los Andes, Colombia. The dataset includes year and municipality level

information on general characteristics of the municipality (population, altitude, geographical location...), conflict and violence, public investment, and the education and agricultural sectors. The earliest year of data available on municipality-level exposure to conflict and violence is 1993. Our instrument is the number of attacks against the civil population in 1993 in the family's municipality by 1000 individuals living in the municipality that year. Mothers in our sample were on average 10 years old in 1993.

C Measurement system

This appendix provides further details about the measurement system relating observed measures of child's skills, parental investments and maternal skills to the latent factors measuring these constructs. We start by providing details of the exploratory factor analysis (EFA) we performed to inform the specification of the measurement system described in Table 1 of the paper. We then present estimates of the measurement system, including estimates of the factor loadings, of the variance of measurement error and of the distribution of latent factors.

C.1 Exploratory factor analysis

Following the psychometric literature (Gorsuch, 1983, 2003), we aim to build a measurement system with dedicated measures (measures that only proxy one latent factor) as it makes interpretation of the latent factors easiest and most transparent. EFA consists of two main steps. First, we select the number of latent factors that should be extracted from all the measures we have on each of the aspects we want to measure (e.g. child's development at baseline, child's development at follow-up, etc.). Second, we allocate measures to factors, estimate factor loadings and discard measurements that load on multiple factors in order to achieve a dedicated measurement system.

C.2 Selecting the number of latent factors

The first step aims to determine how many latent factors should be extracted from each set of measures we have available to measure the child's development at baseline, the child's development at follow-up, parental investment at follow-up, the mother's skills, and the household's wealth at baseline. A variety of methods are available to select the number of factors, and here we implement four of the most popular methods developed in the literature. Below, we succinctly describe each one of them, before commenting on the number of factors

they suggest to extract.

C.2.1 Description of methods

Kaiser's eigenvalue rule The Kaiser's criterion consists of retaining only factors with eigenvalues greater than 1 (Kaiser, 1960). The intuition behind this rule is that unless a factor extracts at least as much variance as the equivalent of one original variable, it should be dropped.

Cattell's scree plot The scree test was proposed by Cattell (1966) and is based on the analyst's inspection of a plot of the eigenvalues associated with the data. Cattell's rule is such that the number of factors should be equal to the number of eigenvalues before which the smooth decrease of eigenvalues appears to level off to the right of the plot.

Velicer's minimum average partial (MAP) correlation rule Velicer (1976)'s minimum average partial (MAP) involves a complete factor analysis followed by the examination of a series of matrices of partial correlations. In the first step, the first factor is partialed out of the correlations between the variables of interest, and the average squared coefficient in the off-diagonals of the resulting partial correlation matrix is computed. In the second step, the first two factors are partialed out of the original correlation matrix and the average squared partial correlation is again computed. These computations are conducted for $k-1$, where k is the number of measurements. The number of components is determined by the step number in the analyses that results in the lowest average squared partial correlation. Intuitively, components are retained as long as the variance in the correlation matrix represents systematic variance. Components are no longer retained when there is proportionately more unsystematic variance than systematic variance.

Horn's parallel analysis Horn's parallel analysis, described in Horn (1965), involves extracting eigenvalues from random data sets that parallel the actual data set with regard to

the number of observations and variables. For example, if the original data set consists of n observations for each one of m variables, then a series of random data matrices of size $n \times m$ are generated, and eigenvalues are computed for the correlation matrices for the original data and for each of the random data sets. The eigenvalues derived from the actual data are then compared to the eigenvalues derived from the random data. Factors are retained as long as the i-th eigenvalue from the actual data is greater than the i-th eigenvalue from the random data.

C.2.2 Results

Appendix Table C.1 reports how many factors each method suggests we should extract from all the measures we have available to measure child's development at follow-up and at baseline, parental investments, mother's skills and household's wealth. Between 1 and 3 factors should be extracted from the measures of child's development at baseline and at follow-up, and between 1 and 4 factors should be extracted from the measures of investments. Most methods suggest that two factors should be extracted from the measures of maternal skills.

Table C.1: Exploratory factor analysis to determine the number of latent factors

	Number of factors according to the following methods			
	Kaiser's eigenvalue rule	Cattell's scree plot	Velicer's MAP rule	Horn's parallel analysis
Child's skills at follow up	3	3	2	3
Child's skills at baseline	3	2	1	3
Investments at follow up	4	1	2	4
Maternal skills at baseline	3	2	2	3

The results from this first step of the EFA suggests that the data we work with may be rich enough to support the model we set out in Section 5, which assumes two dimensions for the child's skills, two dimensions for the mother's skills, and two dimensions for parental

investments. We now need to estimate factor loadings to allocate groups of measures to different factors and identify measures that do not proxy one factor in order to finalize the configuration of measurement system and interpret each factor with precision. This is what we do in the second step of the EFA, which we describe now.

C.3 Specifying the dedicated measurement system

Once we have a strong indication of how many factors should be retained from the data, we search for dedicated measures for each factor by implementing an exploratory factor analysis with *quartimin* rotation. We first estimate the factor loadings in a measurement system for each of the elements we want to measure. We then rotate the factor loadings so as to identify measures that heavily load on one factor and are therefore good candidates for the dedicated measurement system.*

The aim of the *quartimin* rotation is to re-weight the factor loadings obtained from the EFA in a way that leads to a structure of factor loadings such that measures only heavily load on one factor. This helps in identifying good candidate measures for our system of dedicated measures. In contrast, if a measure does not load heavily on a factor or if it is not clearly related to only one factor, it cannot serve as a dedicated measure. In this case, we exclude it from our measurement system.

Tables C.2 - C.5 report rotated factor loadings for each measure. Note that we have assumed two factors for the measures of child development at baseline and at follow-up, two factors for the measures of mother's skills, and two factors for the measures of parental investments. Below we comment on these results and how they informed the final configuration of our measurement system.

*Several methods are available to rotate the factors. We focus on the results of a *quartimin* rotation because it is an oblique rotation and hence allows factors to be correlated. We also performed the same exercise using *geomin* rotation, since it is another type of oblique rotation, and reached similar conclusions with respect to the final measurement system.

Measures of child's development at t and $t+1$ The factor loadings on the measures of child's skills at follow-up ($t+1$) clearly suggest two groupings of measures. The Bayley measures and the MacArthur measures heavily load on a first factor, which we call cognitive skill. Some of the ICQ measures and some of the ECBQ measures heavily loads on a second factor, which we call socio-emotional skill. Note that the measures *ICQ: Unadaptable* and *ECBQ: Attentional shifting* load slightly more heavily on the first factor than on the second, but overall have two very small loadings of each of these factors (below 0.3) and do not clearly load on one of the two factors. As a result, we discard these two measures from the final measurement system. Based on these groupings, we interpret the first factor as measuring the child's cognitive skills at follow-up and the second factor as measuring the child's socio-emotional skill at follow-up.

Turning to the measures on the child's skills at baseline (t), we again obtain a pattern of factor loadings that clearly supports two groupings of measures. On the one hand, all the Bayley measures (cognitive, receptive language, expression language and fine motor) heavily load on a first factor, which we interpret as measuring the child's cognitive skill at baseline. The MacArthur-Bates measures of number of words understood has two very small loadings and does not clearly load on one of the two factors. As a result, we discard this measure as an unfit candidate for our dedicated measurement system. The MacArthur-Bates measures of number of spoken shows a clearer pattern, with a heavier loading on factor 1 than factor 2. Although this pattern is not as clear as the Bayley measures, we do retain the MacArthur-Bates measure of the number of words spoken by the child in our final measurement system, so as to have a measure of the child's vocabulary measuring cognitive skill both at baseline and at follow-up. Finally, the ICQ measures of the child's temperament at baseline clearly load on the second factor. In the final measurement system, we retain these four measures to proxy the factor that we call socio-emotional skill of the child at baseline.

Measures of parental investments The measures of parental investments from the follow-up survey indicate two clear groupings. On the one hand, the variables measuring the variety of play materials and the number of different types of play material, for the most part, load on a first factor. On the other hand, the variable measuring the variety of play activities and most of the variables measuring the frequency of each of these activities in the three days preceding the interview load on a second factor. One variable ("Number of home-made toys") does not clearly load on one of the factors, and for this reason, we exclude it from our system of dedicated measures.

Measure of maternal skills The measures of maternal skills that we have at our disposal in the survey indicate two clear groupings of measure. On the one hand, the mother's years of education, vocabulary and IQ score, along with two items from the FCI (the number of books in the house and the number of magazines and newspapers) seem to be highly correlated with each other, as they clearly load on a first factor (Factor 2 in the table). On the other hand, most items of the CES-D scale heavily load on a second factor. The only exception is the third measure asking the mother "Did you feel hopeful about the future?", and we exclude this measure from our final measurement system since it is not clearly dedicated to either of the two factors measuring maternal skills.

Table C.2: Estimated rotated loadings on child's measures of development at baseline

Bayley: cognitive	0.662*	0.049
Bayley: receptive language	0.803*	-0.015
Bayley: expressive language	0.767*	0.013
Bayley: fine motor	0.683*	-0.066*
MacArthur: words the child can say	0.370*	0.143*
MacArthur: words the child can understand	0.047	0.097*
ICQ: difficult (-)	0.118*	0.382*
ICQ: unsociable (-)	-0.003*	0.669*
ICQ: unadaptable (-)	0.091*	0.429*
ICQ: unstoppable (-)	-0.118*	0.424*

Table C.3: Estimated rotated loadings on child's measures of development at follow-up

Bayley: cognitive	0.786*	-0.028
Bayley: receptive language	0.762*	0.007
Bayley: expressive language	0.777*	-0.016
Bayley: fine motor	0.629*	0.006
MacArthur: words the child can say	0.580*	0.125*
MacArthur: complex phrases the child can say	0.480*	0.137*
ICQ: unsociable (-)	0.166*	0.250*
ICQ: difficult (-)	0.027	0.691*
ICQ: unadaptable (-)	0.193*	0.144*
ICQ: unstoppable (-)	-0.086*	0.722*
ECBQ: inhibitory control	-0.001	0.748*
ECBQ: attentional focusing	0.062*	0.329*
ECBQ: attentional shifting	0.203*	0.078*

Table C.4: Estimated rotated loadings on measures of investment at follow-up

FCI: no. of different types of play materials	0.989*	-0.009
FCI: no. of picture books	0.341*	0.045
FCI: no. of coloring and drawing books	0.444*	-0.025
FCI: no. of home-made toys	0.105*	0.092*
FCI: no. of shop-bought toys	0.514*	0.025
FCI: no. of toys to learn movement	0.513*	0.065*
FCI: no. of toys to learn shapes	0.628*	0.011
FCI: no. of different types of play activities in last 3 days	0.007	0.958*
FCI: no. of times told a story to child in last 3 days	-0.021	0.540*
FCI: no. of times read to child in last 3 days	-0.01	0.541*
FCI: no. of times played with toys in the last 3 days	0.041	0.540*
FCI: no. of times named things to child in last 3 days	-0.017	0.563*

Table C.5: Estimated rotated loadings on baseline measures of maternal skills

Mothers' years of education	0.684*	0.001
Mother's Raven's score (IQ)	0.738*	0.011
Mother's vocabulary	0.778*	-0.036
FCI: no. of books for adults in the home	0.408*	0.007
FCI: no. of magazines and newspapers in the home	0.248*	0.027
CESD: did you feel depressed? (-)	-0.008	0.493*
CESD: are you bothered by what usually don't? (-)	0.019	0.553*
CESD: did you feel hopeful about your future?	0.130*	0.065*
CESD: did you have trouble keep mind on doing? (-)	-0.011	0.732*
CESD: did you feel everything you did was an effort? (-)	0.097*	0.515*
CESD: did you feel fearful? (-)	0.002	0.539*
CESD: was your sleep restless? (-)	-0.143*	0.485*
CESD: did you feel happy?	0.023	0.513*
CESD: how often did you feel lonely in last 7 days? (-)	-0.022	0.573*
CESD: did you feel you couldn't get going? (-)	0.047	0.534*

C.4 Estimates of the measurement system

The following tables report the estimates of the measurement system. Appendix Table C.6 reports the estimates of the factor loadings in each measurement equation. Appendix Table C.7 reports the estimates of the variance of the measurement error contained in each measure.

Appendix Table C.8 reports the estimates of the means of the latent factor distributions for the treated group relative to the means of the latent factor distributions for the control group (which is normalized to 0). Appendix Table C.9 reports the weights assigned to each normal distribution in the joint distribution of latent factors. Appendix Table C.10 reports the correlation between the latent factors in each group.

Table C.6: Estimates of factor loadings

Measures	Latent factor							
	θ^C_{t+1}	θ^C_t	θ^S_{t+1}	θ^S_t	I^M_{t+1}	I^T_{t+1}	P^C_t	P^S_t
Bayley: cognitive	1	0	0	0	0	0	0	0
Bayley: receptive language	0.935	0	0	0	0	0	0	0
Bayley: expressive language	0.992	0	0	0	0	0	0	0
Bayley: fine motor	0.779	0	0	0	0	0	0	0
MacArthur: words the child can say	0.784	0	0	0	0	0	0	0
MacArthur: complex phrases the child can say	0.685	0	0	0	0	0	0	0
Bayley: cognitive	0	1	0	0	0	0	0	0
Bayley: receptive language	0	0.980	0	0	0	0	0	0
Bayley: expressive language	0	1.035	0	0	0	0	0	0
Bayley: fine motor	0	0.909	0	0	0	0	0	0
MacArthur: words the child can say	0	0.679	0	0	0	0	0	0
ICQ: difficult (-)	0	0	1	0	0	0	0	0
ICQ: unsociable (-)	0	0	0.535	0	0	0	0	0
ICQ: unstoppable (-)	0	0	0.837	0	0	0	0	0
ECBQ: inhibitory control	0	0	0.982	0	0	0	0	0
ECBQ: attentional focusing	0	0	0.477	0	0	0	0	0
ICQ: difficult (-)	0	0	0	1	0	0	0	0
ICQ: unsociable (-)	0	0	0	0.542	0	0	0	0
ICQ: unadaptable (-)	0	0	0	0.638	0	0	0	0
ICQ: unstoppable (-)	0	0	0	0.486	0	0	0	0
FCI: no. of different types of play materials	0	0	0	0	1	0	0	0
FCI: no. of coloring and drawing books	0	0	0	0	0.231	0	0	0
FCI: no. of toys to learn movement	0	0	0	0	0.593	0	0	0
FCI: no. of toys to learn shapes	0	0	0	0	0.824	0	0	0
FCI: no. of shop-bought toys	0	0	0	0	0.583	0	0	0
FCI: no. of different types of play activities in last 3 days	0	0	0	0	0	1	0	0
FCI: no. of times told a story to child in last 3 days	0	0	0	0	0	0.778	0	0
FCI: no. of times read to child in last 3 days	0	0	0	0	0	0.965	0	0
FCI: no. of times played with toys in the last 3 days	0	0	0	0	0	0.663	0	0
FCI: no. of times named things to child in last 3 days	0	0	0	0	0	0.658	0	0
Mothers' years of education	0	0	0	0	0	0	1	0
Mother's Raven's score (IQ)	0	0	0	0	0	0	1.021	0
Mother's vocabulary	0	0	0	0	0	0	1.162	0
FCI: no. of books for adults in the home	0	0	0	0	0	0	0.862	0
FCI: no. of magazines and newspapers in the home	0	0	0	0	0	0	0.591	0
CESD: did you feel depressed? (-)	0	0	0	0	0	0	0	1
CESD: are you bothered by what usually don't? (-)	0	0	0	0	0	0	0	0.686
CESD: did you have trouble keep mind on doing? (-)	0	0	0	0	0	0	0	0.773
CESD: did you feel everything you did was an effort? (-)	0	0	0	0	0	0	0	0.727
CESD: did you feel fearful? (-)	0	0	0	0	0	0	0	0.749
CESD: was your sleep restless? (-)	0	0	0	0	0	0	0	0.622
CESD: did you feel happy?	0	0	0	0	0	0	0	0.726
CESD: how often did you feel lonely last week? (-)	0	0	0	0	0	0	0	0.786
CESD: did you feel you couldn't get going? (-)	0	0	0	0	0	0	0	0.798

Table C.7: Estimated variance of the measurement error in each measure included in the measurement system

Factor	Measurements	Variance of the measurement error	
		Estimate	Stand. Error
θ_{t+1}^C	Bayley: cognitive	0.430	(0.029)
	Bayley: receptive language	0.443	(0.026)
	Bayley: expressive language	0.428	(0.032)
	Bayley: fine motor	0.633	(0.034)
	MacArthur: words the child can say	0.538	(0.026)
	MacArthur: complex phrases the child can say	0.658	(0.040)
θ_t^C	Bayley: cognitive	0.520	(0.029)
	Bayley: receptive language	0.444	(0.030)
	Bayley: expressive language	0.450	(0.035)
	Bayley: fine motor	0.677	(0.043)
	MacArthur: words the child can say	0.702	(0.070)
θ_{t+1}^S	ICQ: difficult (-)	0.458	(0.043)
	ICQ: unsociable (-)	0.757	(0.041)
	ICQ: unstoppable (-)	0.635	(0.040)
	ECBQ: inhibitory control	0.481	(0.042)
	ECBQ: attentional focusing	0.811	(0.049)
θ_t^C	ICQ: difficult (-)	0.511	(0.055)
	ICQ: unsociable (-)	0.818	(0.046)
	ICQ: unadaptable (-)	0.814	(0.042)
	ICQ: unstoppable (-)	0.889	(0.048)
I_t^M	FCI: no. of different types of play materials	0.161	(0.048)
	FCI: no. of coloring and drawing books	0.715	(0.100)
	FCI: no. of toys to learn movement	0.604	(0.037)
	FCI: no. of toys to learn shapes	0.825	(0.082)
	FCI: no. of shop-bought toys	0.586	(0.049)
I_t^T	FCI: no. of different types of play activities in last 3 days	0.352	(0.081)
	FCI: no. of times told a story to child in last 3 days	0.713	(0.061)
	FCI: no. of times read to child in last 3 days	0.798	(0.074)
	FCI: no. of times played with toys in the last 3 days	0.845	(0.079)
	FCI: no. of times named things to child in last 3 days	0.794	(0.109)
P^C	Mothers' years of education	0.584	(0.037)
	Mother's Raven's score (IQ)	0.599	(0.030)
	Mother's vocabulary	0.496	(0.037)
	FCI: no. of books for adults in the home	0.773	(0.040)
	FCI: no. of magazines and newspapers in the home	0.912	(0.036)
P^S	CESD: did you feel depressed? (-)	0.464	(0.031)
	CESD: are you bothered by what usually don't? (-)	0.711	(0.040)
	CESD: did you have trouble keep mind on doing? (-)	0.680	(0.040)
	CESD: did you feel everything you did was an effort? (-)	0.682	(0.035)
	CESD: did you feel fearful? (-)	0.685	(0.039)
	CESD: was your sleep restless? (-)	0.798	(0.038)
	CESD: did you feel happy?	0.672	(0.040)
	CESD: how often did you feel lonely last week? (-)	0.642	(0.036)
	CESD: did you feel you couldn't get going? (-)	0.728	(0.048)

Table C.8: Estimated means of the (log) latent factors in the treated group

	Estimate	Stand. error
Child's cognitive skill at follow-up	0.115	(0.051)
Child's cognitive skill at baseline	-0.022	(0.067)
Child's socio-emotional skill at follow-up	0.087	(0.044)
Child's socio-emotional skill at baseline	0.034	(0.058)
Material investment	0.227	(0.069)
Time investment	0.302	(0.068)
Mother's cognitive skill	-0.016	(0.042)
Mother's socio-emotional skill	0.105	(0.046)

Table C.9: Estimates of the mixture weights of the joint distribution of the latent factors

	Mixture weight 1		Mixture weight 2	
	Estimate	Stand. error	Estimate	Stand. error
Control group	0.442	(0.026)	0.558	(0.026)
Treated group	0.357	(0.024)	0.643	(0.024)

Table C.10: Estimated latent factor correlation

Control group

	θ^C_{t+1}	θ^C_t	θ^S_{t+1}	θ^S_t	I^M_{t+1}	I^T_{t+1}	P^C_t	P^S_t
θ^C_{t+1}	1	0.667	0.363	0.296	0.396	0.280	0.496	0.148
θ^C_t	0.667	1	0.257	0.261	0.267	0.225	0.345	0.044
θ^S_{t+1}	0.363	0.257	1	0.622	0.372	0.275	0.290	0.183
θ^S_t	0.296	0.261	0.622	1	0.290	0.235	0.406	0.242
I^M_{t+1}	0.396	0.267	0.372	0.290	1	0.624	0.603	0.123
I^T_{t+1}	0.280	0.225	0.275	0.235	0.624	1	0.394	0.053
P^C_t	0.496	0.345	0.290	0.406	0.603	0.394	1	0.070
P^S_t	0.148	0.044	0.183	0.242	0.123	0.053	0.070	1

Treated group

	θ^C_{t+1}	θ^C_t	θ^S_{t+1}	θ^S_t	I^M_{t+1}	I^T_{t+1}	P^C_t	P^S_t
θ^C_{t+1}	1	0.681	0.320	0.213	0.292	0.172	0.257	0.087
θ^C_t	0.681	1	0.245	0.203	0.213	0.134	0.180	0.034
θ^S_{t+1}	0.320	0.245	1	0.445	0.282	0.284	0.234	0.183
θ^S_t	0.213	0.203	0.445	1	0.086	0.045	0.274	0.342
I^M_{t+1}	0.292	0.213	0.282	0.086	1	0.451	0.457	0.021
I^T_{t+1}	0.172	0.134	0.284	0.045	0.451	1	0.276	0.054
P^C_t	0.257	0.180	0.234	0.274	0.457	0.276	1	0.098
P^S_t	0.087	0.034	0.183	0.342	0.021	0.054	0.098	1

D Supplemental evidence on the intervention impacts

Table D.1 reports the treatment effects of the stimulation and micronutrient supplementation components of the intervention, and of their interactions on all the raw measures we use in the measurement system to proxy the latent factors for children's cognitive skill at follow-up, children's socio-emotional skills at follow-up, material investments and time investments at follow-up. Standard errors are clustered at the municipality level. To obtain these effects we regressed the outcome on an indicator that takes the value of 1 if any stimulation (alone or in combination), an indicator that takes the value of 1 if any supplementation (alone or in combination), and an indicator that takes the value of 1 if both stimulation and supplementation. We controlled for tester effects (two for each region), sex of the child, and baseline level of the outcomes, except for items that were not measured at baseline. Specifically, for *MacArthur: Complex phrases the child can say*, we controlled for the MacArthur item measure the number of words the child can say at baseline. For ECBQ items, we controlled for the ICQ item measuring how "difficult" the child is at baseline (*ICQ: difficult*). We controlled for baseline *FCI: Number of different types of play materials* for all items in Panel C and for baseline *FCI: Number of different types of play activities* for all items in Panel D.

Table D.2 compares the treatment effects of receiving any stimulation on the raw measures of material and time investments at follow-up between the full sample and the sample where we exclude the households where the data collection was less than four days after the last home visit.

Table D.3 shows descriptive statistics of the play materials delivered by the intervention and recorded at follow-up in the homes of the treatment group.

Table D.1: Treatment impacts of the psychosocial stimulation and micronutrient supplementation components of the intervention

	Any stimulation	Any supplementation	Supplementation and stimulation
A - Child's cognitive skills at follow-up			
Bayley: cognitive	0.310	-0.033	-0.121
	(0.084)	(0.073)	(0.123)
Bayley: receptive language	0.236	-0.047	-0.122
	(0.091)	(0.095)	(0.125)
Bayley: expressive language	0.080	-0.088	-0.099
	(0.078)	(0.092)	(0.126)
Bayley: fine motor	0.135	-0.139	-0.127
	(0.089)	(0.084)	(0.119)
MacArthur: words the child can say	0.143	-0.118	-0.106
	(0.087)	(0.077)	(0.131)
MacArthur: complex phrases the child can say	0.005	0.067	0.109
	(0.078)	(0.073)	(0.106)
B - Child's socio-emotional skills at follow-up			
ICQ: difficult (-)	0.119	-0.002	-0.090
	(0.067)	(0.060)	(0.090)
ICQ: unsociable (-)	0.001	0.010	0.080
	(0.083)	(0.084)	(0.110)
ICQ: unstoppable (-)	0.045	-0.011	-0.026
	(0.077)	(0.074)	(0.116)
ECBQ: inhibitory control	0.029	0.007	-0.065
	(0.078)	(0.079)	(0.116)

Table D.1 (cont.): Treatment impacts of the psychosocial stimulation and micronutrient supplementation components of the intervention

	Any stimulation	Any supplementation	Supplementation and stimulation
C - Material investment at follow-up			
FCI: no. of different types of play materials	0.282	-0.116	-0.138
	(0.080)	(0.096)	(0.128)
FCI: no. of books to pain and draw	-0.124	0.014	-0.017
	(0.076)	(0.090)	(0.113)
FCI: no. of toys to learn movement	0.059	-0.160	-0.219
	(0.083)	(0.108)	(0.136)
FCI: no. of toys to learn shapes	0.443	-0.069	-0.057
	(0.099)	(0.100)	(0.174)
FCI: no. of shop-bought toys	0.046	-0.121	-0.046
	(0.082)	(0.099)	(0.130)
D - Time investment at follow-up			
FCI: no. of types of play activities last 3 days	0.287	-0.066	-0.021
	(0.075)	(0.068)	(0.105)
FCI: no. of times told a story last 3 days	0.208	-0.084	-0.143
	(0.085)	(0.071)	(0.119)
FCI: no. of times read last 3 days	0.353	0.041	0.020
	(0.092)	(0.079)	(0.120)
FCI: no. of times played the last 3 days	0.115	-0.034	0.119
	(0.078)	(0.070)	(0.119)
FCI: no. of times named things last 3 days	0.078	-0.022	0.118
	(0.061)	(0.069)	(0.099)

Note: Coefficients and standard errors clustered at the municipality level (in parentheses) from a regression of the dependent variable measured at follow-up on an indicator that takes the value of 1 if any stimulation (alone or in combination), an indicator that takes the value of 1 if any supplementation (alone or in combination), and an indicator that takes the value of 1 if both stimulation and supplementation. We also control for the child's sex; tester dummies; and baseline level of the outcome. ***, ** and * indicate significance at 1, 5, and 10%. All scores have been internally standardized non-parametrically for age and are therefore expressed in standard deviations (see Appendix B for details about the measures and the standardization procedure). Sample size ranges from 1,262 to 1,326 depending on the outcome (see Table 2 in the paper for exact sample size for each outcome).

Table D.2: Impacts of the intervention on investments in the full and restricted samples

	Full sample	Restricted sample
A- Material investment at follow-up		
FCI: no. of different types of play materials	0.215	0.216
	(0.064)	(0.066)
FCI: no. of coloring and drawing books	-0.133	-0.136
	(0.056)	(0.058)
FCI: no. of toys to learn movement	-0.048	-0.051
	(0.065)	(0.067)
FCI: no. of toys to learn shapes	0.416	0.421
	(0.088)	(0.092)
FCI: no. of shop-bought toys	0.024	0.048
	(0.061)	(0.062)
B - Time investment at follow-up		
FCI: no. of different types of play activities in last 3 days	0.277	0.273
	(0.050)	(0.052)
FCI: no. of times told a story to child in last 3 days	0.138	0.113
	(0.060)	(0.060)
FCI: no. of times read to child in last 3 days	0.362	0.350
	(0.062)	(0.065)
FCI: no. of times played with toys in the last 3 days	0.175	0.193
	(0.060)	(0.063)
FCI: no. of times named things to child in last 3 days	0.137	0.134
	(0.048)	(0.050)
Number of observations	*1,326*	*1,244*

Note: Asymptotic standard errors in parentheses clustered at the municipality. Restricted sample excludes all those households where the data collection was less than four days after the last home visit.

Table D.3: Materials from the intervention in treatment households at first follow-up

	Mean	St. Dev.
Toys given by the home visitor	4.265	(6.378)
Toys made with the home visitor	0.841	(2.638)
Intervention books	5.198	(3.390)

Note: Sample size is 667 observations.

E Additional results

E.1 Reduced form production functions

In this section we present the reduced form production functions, where the material and time investments have been substituted out. The estimates of the specification using toy prices, food prices, and maternal childhood exposure to conflict as instruments are reported in Table E.1.

Table E.1: Reduced form Production Functions

	Cognitive skill	Socio-emotional skill
Intercept	-0.019	-0.010
	[-0.097,0.062]	[-0.086,0.066]
Treatment	0.085	0.071
	[-0.026,0.193]	[-0.057,0.192]
Log child's cognitive skill (t)	0.726	0.111
	[0.642,0.819]	[0.022,0.214]
Log child's socio-emotional skill (t)	-0.002	0.527
	[-0.087,0.084]	[0.409,0.667]
Log mother's cognitive skill	0.279	0.073
	[0.169,0.411]	[-0.049,0.201]
Log mother's socio-emotional skill	0.110	0.047
	[0.039,0.181]	[-0.049,0.13]
Log number of children	0.026	0.063
	[-0.024,0.075]	[0.015,0.112]
Log toy price	-0.013	-0.022
	[-0.072,0.049]	[-0.075,0.032]
Log food price	0.079	-0.015
	[0.017,0.136]	[-0.082,0.044]
Maternal childhood exposure to conflict	0.054	-0.067
	[0,0.102]	[-0.116,0.002]

Note: Dependent variable in column 1 (2) is the log of cognitive (socio-emotional) skill of the child at follow up $(t+1)$. t refers to baseline/pre-treatment measurement. 95% confidence intervals in brackets based on 1,000 bootstrap replications of the entire estimation process taking into account clustering at the municipality level.

E.2 Specification and robustness checks

Table E.2 reports the estimates of the investment functions where all the variables of the investment function are interacted with the treatment indicator.

Table E.3 report the estimates of the Cobb-Douglas production functions for cognitive skills and socio-emotional skills where all the parameters are interacted with the treatment dummy. The production function for cognitive skills is estimated by IV where we use toy price and food price as exclusion restrictions. The production function for socio-emotional skills is estimated by OLS.

Table E.4 reports the estimates of the production functions in the case where the control group comprises of children who neither received the stimulation nor the micronutrient supplementation component of the intervention and the treated group only comprises of children who received the stimulation, but not the micronutrient supplementation component.

Table E.5 reports the estimates of the production functions where the joint distribution of latent factors has been estimated using measures from which we removed tester fixed effects before age standardizing them. Specifically, we first regressed the raw measures on tester dummies and then age standardized the residuals of this regression using the age standardization procedure described in Appendix B.4.

Stratum fixed effects The randomization was stratified by geography. There were three strata and the probability of assignment to treatment was equal in each stratum. The strata are "Central Region", "Oriental Region" and "Zona Cafetera". Consequently the experimental results do not require controlling for stratum fixed effects. However the estimation of the production functions and the investment equation could be affected in principle by their inclusion. In this subsection we report results that include the strata fixed effects in the investment functions, the reduced form equations and the production functions. Table E.6 reports the estimates of the investment functions, and Table E.7 reports the estimates of the production functions estimated by OLS and IV. The results are substantially the same

as those reported in the main paper and none of the conclusions change. Any differences that do exist are insignificant.

E.3 Invariance of the measurement system between the treated and control groups

Tables E.8 and E.9 report estimates of the factor loadings in the measurement system where we allow the loadings associated with measures of child's skills and parental investments at follow up to differ between control and treatment groups. We test for the equality of loadings associated with measures of the child's cognitive and socio-emotional skills, as well as measures of parental investments at follow-up. We find that the p-value is 0.064, thus we cannot reject that they are the same between treated and control at the 5% significance level. Table E.10 report the estimates of our preferred specifications for the production function for cognitive and socio-emotional skills where the latent factors are estimated using this measurement system.

Table E.2: Material and time investment equations where all parameters are interacted with treatment

	Material investment	Time investment
Intercept	-0.017	0.002
	[-0.112,0.086]	[-0.092,0.094]
Treatment	0.204	0.330
	[0.032,0.378]	[0.131,0.494]
Log child's cognitive skill (t)	-0.064	0.005
	[-0.374,0.29]	[-0.304,0.317]
x Treatment	0.130	0.039
	[-0.078,0.329]	[-0.177,0.249]
Log child's socio-emotional skill (t)	0.028	0.134
	[-0.287,0.402]	[-0.203,0.528]
x Treatment	-0.050	-0.087
	[-0.286,0.142]	[-0.343,0.117]
Log mother's cognitive skill	1.088	0.580
	[0.642,1.522]	[0.15,0.973]
x Treatment	-0.185	-0.126
	[-0.436,0.094]	[-0.344,0.098]
Log mother's socio-emotional skill	0.171	-0.072
	[-0.099,0.43]	[-0.333,0.183]
x Treatment	-0.061	0.068
	[-0.224,0.105]	[-0.111,0.242]
Log number of children	-0.122	-0.251
	[-0.276,0.029]	[-0.416,-0.089]
x Treatment	0.001	0.079
	[-0.098,0.112]	[-0.021,0.182]
Log toy price	-0.132	-0.151
	[-0.369,0.093]	[-0.347,0.012]
x Treatment	0.022	0.083
	[-0.128,0.171]	[-0.034,0.221]
Log food price	-0.106	-0.087
	[-0.405,0.147]	[-0.291,0.132]
x Treatment	0.133	0.088
	[-0.049,0.322]	[-0.074,0.241]
Maternal childhood exposure to conflict	0.072	0.002
	[-0.219,0.25]	[-0.142,0.252]
x Treatment	-0.056	-0.062
	[-0.172,0.145]	[-0.203,0.043]
F-test of joint significance of interactions with treatment dummy:		
p-value	0.369	0.099

Note: Dependent variable in column 1 (2) is the log of material (time) investment measured at follow-up $(t+1)$. t refers to baseline/pre-treatment measurement. 95% confidence intervals in brackets based on 1,000 bootstrap replications of the entire estimation process taking into account clustering at the municipality level.

Table E.3: Production functions where all parameters are allowed to vary with treatment

	Cognitive skill	Socio-emotional skill
	IV (prices)	OLS
Intercept	-0.014	-0.01
	[-0.106,0.072]	[-0.085,0.063]
x Treatment	0.18	-0.001
	[-0.065,0.385]	[-0.11,0.107]
Log child's cognitive skill (t)	0.489	0.077
	[0.189,0.816]	[-0.172,0.356]
x Treatment	0.144	0.02
	[-0.07,0.335]	[-0.148,0.169]
Log child's socio-emotional skill (t)	0.056	0.843
	[-0.179,0.339]	[0.529,1.159]
x Treatment	-0.033	-0.202
	[-0.229,0.131]	[-0.385,-0.027]
Log mother's cognitive skill	0.484	-0.291
	[-0.205,1.231]	[-0.702,0.122]
x Treatment	-0.221	0.127
	[-0.91,0.36]	[-0.118,0.376]
Log mother's socio-emotional skill	0.177	0.012
	[-0.105,0.411]	[-0.254,0.275]
x Treatment	-0.043	0.019
	[-0.185,0.125]	[-0.138,0.179]
Log number of children	0.135	0.13
	[-0.051,0.33]	[-0.023,0.274]
x Treatment	-0.053	-0.022
	[-0.184,0.105]	[-0.122,0.08]
Log material investment	0.145	0.251
	[-0.757,0.984]	[0.006,0.534]
x Treatment	0.032	-0.061
	[-0.731,0.894]	[-0.21,0.07]
Log time investment		-0.023
		[-0.404,0.295]
x Treatment		0.073
		[-0.102,0.298]
P-value of F-test of joint significance of interactions with treatment dummy		
Excluding the interaction with the intercept	0.283	0.377
Including the interaction with the intercept	0.284	0.47

Note: Dependent variable in column 1 (2) is the log of cognitive (socio-emotional) skill measured at follow-up $(t+1)$. t refers to baseline/pre-treatment measurement. The specification for cognitive skills corresponds to the specification reported in column 5 of Table 6, and the specification for socio-emotional skills corresponds to that reported in column 1 of Table 7. 95% confidence intervals in brackets based on 1,000 bootstrap replications of the entire estimation process taking into account clustering at the municipality level.

Table E.4: Estimates of the production functions on the sample excluding children who received micronutrient supplementation

	(1) Cognitive skill	(2) Socio-emotional skill
	IV (prices, treatment)	OLS
Intercept	-0.006	-0.007
	[-0.122,0.063]	0.103,0.087]
Treatment		-0.017
		[-0.154,0.127]
Log child's cognitive skill (t)	0.628	0.142
	[0.485,0.802]	[-0.025,0.286]
Log child's socio-emotional skill (t)	0.064	0.606
	[-0.077,0.209]	[0.475,0.795]
Log mother's cognitive skill	-0.123	-0.142
	[-0.64,0.197]	[-0.383,0.088]
Log mother's socio-emotional skill	0.115	-0.007
	[-0.021,0.258]	[-0.14,0.131]
Log number of children	0.073	0.106
	[-0.016,0.163]	[0.028,0.182]
Log material investment	0.453	0.14
	[0.154,0.97]	[0.022,0.299]
Log time investment		0.139
		[-0.236,0.275]

Note: Dependent variable in column 1 (2) is the log of cognitive (socio-emotional) skill measured at follow-up $(t+1)$. t refers to baseline/pre-treatment measurement. The specification for cognitive skills corresponds to the specification reported in column 5 of Table 6, and the specification for socio-emotional skills corresponds to that reported in column 1 of Table 7. The sample excludes children who received the micronutrient supplementation intervention. 95% confidence intervals in brackets based on 1,000 bootstrap replications of the entire estimation process taking into account clustering at the municipality level.

Table E.5: Estimates of the production functions where tester effects have been removed from the measures

	(1) Cognitive skill	(2) Socio-emotional skill
	IV (prices, treatment)	OLS
Intercept	-0.019	-0.011
	[-0.117,0.051]	[-0.069,0.045]
Treatment		0.003
		[-0.084,0.096]
Log child's cognitive skill (t)	0.663	0.063
	[0.558,0.79]	[-0.022,0.153]
Log child's socio-emotional skill (t)	-0.053	0.478
	[-0.158,0.039]	[0.392,0.617]
Log mother's cognitive skill	-0.143	0.014
	[-0.659,0.241]	[-0.128,0.137]
Log mother's socio-emotional skill	-0.039	0.057
	[-0.187,0.084]	[-0.024,0.129]
Log number of children	0.093	0.072
	[0.018,0.183]	[0.023,0.121]
Log material investment	0.617	0.11
	[0.153,1.236]	[0.037,0.206]
Log time investment		0.054
		[-0.127,0.134]

Note: Dependent variable in column 1 (2) is the log of cognitive (socio-emotional) skill measured at follow-up $(t+1)$. t refers to baseline/pre-treatment measurement. The specification for cognitive skills corresponds to the specification reported in column 5 of Table 6, and the specification for socio-emotional skills corresponds to that reported in column 1 of Table 7. In this specification, the latent factors have been estimated using measures from which we first removed tested fixed effects. 95% confidence intervals in brackets based on 1,000 bootstrap replications of the entire estimation process taking into account clustering at the municipality level.

Table E.6: Estimates of the investment equations with stratum fixed effects

	Material investment	Time investment
Intercept	0.181	0.106
	[0.008,0.357]	[-0.014,0.232]
Treatment	0.208	0.306
	[0.036,0.371]	[0.142,0.46]
Log child's cognitive skill (t)	0.067	0.042
	[-0.025,0.164]	[-0.07,0.154]
Log child's socio-emotional skill (t)	-0.004	0.054
	[-0.109,0.119]	[-0.056,0.184]
Log mother's cognitive skill	0.799	0.378
	[0.632,0.991]	[0.187,0.531]
Log mother's socio-emotional skill	0.083	0.031
	[0.011,0.149]	[-0.048,0.119]
Log number of children	-0.128	-0.125
	[-0.179,-0.077]	[-0.184,-0.069]
Log toy price	-0.064	0.000
	[-0.135,0.005]	[-0.064,0.061]
Log food price	0.144	0.099
	[0.044,0.245]	[0.014,0.197]
Maternal childhood exposure to conflict	-0.003	-0.068
	[-0.083,0.076]	[-0.123,-0.009]
Central region	-0.224	-0.015
	[-0.455,0.01]	[-0.199,0.159]
Oriental region	-0.359	-0.298
	[-0.603,-0.124]	[-0.539,-0.073]
Rank test	0.071	
Cragg-Donald test	0.065	
Test of joint significance - F-statistic (p-values):		
Toy price, food price, conflict	12.15 (0.014)	13.43 (0.007)
Toy price, food price, conflict, treatment	23.75 (0.000)	28.16 (0.000)
Toy price, food price	12.15 (0.005)	4.44 (0.106)
Toy price, food price, treatment	23.69 (0.000)	23.28 (0.000)
Conflict, treat	5.74 (0.050)	18.12 (0.000)

Note: Investment equations including stratum fixed effects, where Zona Cafetera is the reference region. Dependent variable in column 1 (2) is the log of material (time) investment measured at follow-up ($t+1$). t refers to baseline/pre-treatment measurement. 95% confidence intervals in brackets based on 1,000 bootstrap replications of the entire estimation process taking into account clustering at the municipality level.

Table E.7: Cognitive and socio-emotional production functions with stratum fixed effects

	Cognitive skill		Socio-emotional skill	
	OLS	IV	OLS	IV
Instruments		*Prices and conflict*		*Prices and conflict*
Intercept	-0.038	-0.066	0.111	0.136
	[-0.143,0.069]	[-0.203,0.059]	[0.003,0.219]	[-0.01,0.305]
Treatment	0.078	0.12	-0.007	-0.122
	[-0.024,0.184]	[-0.067,0.289]	[-0.116,0.094]	[-0.309,0.137]
Log child's cognitive skill (t)	0.691	0.689	0.055	0.046
	[0.609,0.788]	[0.595,0.803]	[-0.043,0.158]	[-0.091,0.172]
Log child's socio-emotional skill (t)	-0.021	0.005	0.533	0.498
	[-0.118,0.066]	[-0.116,0.11]	[0.413,0.672]	[0.377,0.676]
Log mother's cognitive skill	0.191	-0.008	-0.031	0.048
	[0.059,0.345]	[-0.313,0.346]	[-0.183,0.105]	[-0.315,0.498]
Log mother's socio-emotional skill	0.102	0.081	0.042	0.051
	[0.032,0.175]	[-0.006,0.169]	[-0.049,0.125]	[-0.048,0.153]
Log number of children	0.042	0.038	0.097	0.12
	[-0.008,0.091]	[-0.036,0.125]	[0.045,0.151]	[0.017,0.211]
Log material investment	0.082	0.569	0.142	-0.38
	[0.003,0.156]	[0.006,0.922]	[0.05,0.239]	[-0.868,0.264]
Log time investment	0.065	-0.423	0.11	0.851
	[-0.022,0.159]	[-0.913,0.26]	[-0.009,0.212]	[0.014,1.41]
Central region	-0.113	-0.006	-0.174	-0.316
	[-0.262,0.034]	[-0.211,0.186]	[-0.322,-0.033]	[-0.527,-0.07]
Oriental region	0.172	0.179	-0.183	-0.118
	[0.026,0.327]	[-0.009,0.396]	[-0.332,-0.046]	[-0.406,0.052]

Note: Production functions including stratum fixed effects, where Zona Cafetera is the reference region. Dependent variable in column 1 and 2 (3 and 4) is the log of cognitive (socio-emotional) skill of the child at follow up ($t+1$). t refers to baseline/pre-treatment measurement. 95% confidence intervals in brackets based on 1,000 bootstrap replications of the entire estimation process taking into account clustering at the municipality level.

Table E.8: Estimates of factor loadings in the control group

	Latent factor							
Measures	θ^C_{t+1}	θ^C_t	θ^S_{t+1}	θ^S_t	I^M_{t+1}	I^T_{t+1}	P^C	P^S
Bayley: cognitive	1	0	0	0	0	0	0	0
Bayley: receptive language	1.005	0	0	0	0	0	0	0
Bayley: expressive language	1.014	0	0	0	0	0	0	0
Bayley: fine motor	0.852	0	0	0	0	0	0	0
MacArthur: words the child can say	0.827	0	0	0	0	0	0	0
MacArthur: complex phrases the child can say	0.801	0	0	0	0	0	0	0
Bayley: cognitive	0	1	0	0	0	0	0	0
Bayley: receptive language	0	1.023	0	0	0	0	0	0
Bayley: expressive language	0	1.046	0	0	0	0	0	0
Bayley: fine motor	0	0.889	0	0	0	0	0	0
MacArthur: words the child can say	0	0.689	0	0	0	0	0	0
ICQ: difficult (-)	0	0	1	0	0	0	0	0
ICQ: unsociable (-)	0	0	0.532	0	0	0	0	0
ICQ: unstoppable (-)	0	0	0.807	0	0	0	0	0
ECBQ: inhibitory control	0	0	1.025	0	0	0	0	0
ECBQ: attentional focusing	0	0	0.450	0	0	0	0	0
ICQ: difficult (-)	0	0	0	1	0	0	0	0
ICQ: unsociable (-)	0	0	0	0.460	0	0	0	0
ICQ: unadaptable (-)	0	0	0	0.556	0	0	0	0
ICQ: unstoppable (-)	0	0	0	0.433	0	0	0	0
FCI: no. of different types of play materials	0	0	0	0	1	0	0	0
FCI: no. of books to pain and draw	0	0	0	0	0.717	0	0	0
FCI: no. of toys bought	0	0	0	0	0.382	0	0	0
FCI: no. of toys to learn movement	0	0	0	0	0.561	0	0	0
FCI: no. of toys to learn shape	0	0	0	0	0.740	0	0	0
FCI: no. of different types of play activities in last 3 days	0	0	0	0	0	1	0	0
FCI: no. of times told a story to child in last 3 days	0	0	0	0	0	1.183	0	0
FCI: no. of times read to child in last 3 days	0	0	0	0	0	1.016	0	0
FCI: no. of times played with toys in the last 3 days	0	0	0	0	0	0.716	0	0
FCI: no. of times named things to child in last 3 days	0	0	0	0	0	0.823	0	0
Mothers' years of education	0	0	0	0	0	0	1	0
Mother's Raven's score (IQ)	0	0	0	0	0	0	1.032	0
Mother's vocabulary	0	0	0	0	0	0	1.198	0
FCI: no. of books for adults in the home	0	0	0	0	0	0	0.819	0
FCI: no. of magazines and newspapers in the home	0	0	0	0	0	0	0.601	0
CESD: did you feel depressed? (-)	0	0	0	0	0	0	0	1
CESD: are you bothered by what usually don't? (-)	0	0	0	0	0	0	0	0.682
CESD: did you have trouble keep mind on doing? (-)	0	0	0	0	0	0	0	0.775
CESD: did you feel everything you did was an effort? (-)	0	0	0	0	0	0	0	0.718
CESD: did you feel fearful? (-)	0	0	0	0	0	0	0	0.737
CESD: was your sleep restless? (-)	0	0	0	0	0	0	0	0.655
CESD: did you feel happy?	0	0	0	0	0	0	0	0.687
CESD: how often did you feel lonely last week? (-)	0	0	0	0	0	0	0	0.764
CESD: did you feel you couldn't get going? (-)	0	0	0	0	0	0	0	0.749

Table E.9: Estimates of factor loadings in the treated group

	Latent factor							
Measures	θ^C_{t+1}	θ^C_t	θ^S_{t+1}	θ^S_t	I^M_{t+1}	I^T_{t+1}	P^C	P^S
Bayley: cognitive	1	0	0	0	0	0	0	0
Bayley: receptive language	0.885	0	0	0	0	0	0	0
Bayley: expressive language	1.056	0	0	0	0	0	0	0
Bayley: fine motor	0.778	0	0	0	0	0	0	0
MacArthur: words the child can say	0.792	0	0	0	0	0	0	0
MacArthur: complex phrases the child can say	0.620	0	0	0	0	0	0	0
Bayley: cognitive	0	1	0	0	0	0	0	0
Bayley: receptive language	0	1.023	0	0	0	0	0	0
Bayley: expressive language	0	1.046	0	0	0	0	0	0
Bayley: fine motor	0	0.889	0	0	0	0	0	0
MacArthur: words the child can say	0	0.689	0	0	0	0	0	0
ICQ: difficult (-)	0	0	1	0	0	0	0	0
ICQ: unsociable (-)	0	0	0.413	0	0	0	0	0
ICQ: unstoppable (-)	0	0	0.893	0	0	0	0	0
ECBQ: inhibitory control	0	0	0.936	0	0	0	0	0
ECBQ: attentional focusing	0	0	0.452	0	0	0	0	0
ICQ: difficult (-)	0	0	0	1	0	0	0	0
ICQ: unsociable (-)	0	0	0	0.460	0	0	0	0
ICQ: unadaptable (-)	0	0	0	0.556	0	0	0	0
ICQ: unstoppable (-)	0	0	0	0.433	0	0	0	0
FCI: no. of different types of play materials	0	0	0	0	1	0	0	0
FCI: no. of books to pain and draw	0	0	0	0	0.130	0	0	0
FCI: no. of toys bought	0	0	0	0	0.765	0	0	0
FCI: no. of toys to learn movement	0	0	0	0	0.721	0	0	0
FCI: no. of toys to learn shape	0	0	0	0	0.718	0	0	0
FCI: no. of different types of play activities in last 3 days	0	0	0	0	0	1	0	0
FCI: no. of times told a story to child in last 3 days	0	0	0	0	0	0.670	0	0
FCI: no. of times read to child in last 3 days	0	0	0	0	0	0.690	0	0
FCI: no. of times played with toys in the last 3 days	0	0	0	0	0	0.613	0	0
FCI: no. of times named things to child in last 3 days	0	0	0	0	0	0.625	0	0
Mothers' years of education	0	0	0	0	0	0	1	0
Mother's Raven's score (IQ)	0	0	0	0	0	0	1.032	0
Mother's vocabulary	0	0	0	0	0	0	1.198	0
FCI: no. of books for adults in the home	0	0	0	0	0	0	0.819	0
FCI: no. of magazines and newspapers in the home	0	0	0	0	0	0	0.601	0
CESD: did you feel depressed? (-)	0	0	0	0	0	0	0	1
CESD: are you bothered by what usually don't? (-)	0	0	0	0	0	0	0	0.682
CESD: did you have trouble keep mind on doing? (-)	0	0	0	0	0	0	0	0.775
CESD: did you feel everything you did was an effort? (-)	0	0	0	0	0	0	0	0.718
CESD: did you feel fearful? (-)	0	0	0	0	0	0	0	0.737
CESD: was your sleep restless? (-)	0	0	0	0	0	0	0	0.655
CESD: did you feel happy?	0	0	0	0	0	0	0	0.687
CESD: how often did you feel lonely last week? (-)	0	0	0	0	0	0	0	0.764
CESD: did you feel you couldn't get going? (-)	0	0	0	0	0	0	0	0.749

Table E.10: Production functions where the measurement system allows for different factor loadings and intercepts between the control and treatment groups

	(1) Cognitive skill	(2) Socio-emotional skill
	IV (prices, treatment)	OLS
Intercept	-0.007	0.005
	[-0.075,0.054]	[-0.065,0.068]
Treatment		-0.009
		[-0.118,0.099]
Log child's cognitive skill (t)	0.633	0.119
	[0.528,0.75]	[0.028,0.209]
Log child's socio-emotional skill (t)	0.012	0.469
	[-0.086,0.116]	[0.372,0.611]
Log mother's cognitive skill	-0.104	-0.083
	[-0.495,0.201]	[-0.222,0.042]
Log child's socio-emotional skill	0.073	0.041
	[-0.024,0.151]	[-0.054,0.125]
Log number of chidlren	0.086	0.168
	[0.017,0.181]	[0.074,0.277]
Log material investment	0.517	0.11
	[0.16,1.03]	[0.003,0.198]
Log time investment		0.104
		[0.051,0.159]

Note: Dependent variable in column 1 (2) is the log of cognitive (socio-emotional) skill measured at follow-up $(t+1)$. t refers to baseline/pre-treatment measurement. The specification for cognitive skills corresponds to the specification reported in column 5 of Table 6, and the specification for socio-emotional skills corresponds to that reported in column 1 of Table 7. In this specification, the latent factors have been estimated using a measurement system that allows for different factor loadings and intercepts for all measures of child skills and investments at follow up. 95% Confidence intervals in brackets based on 1,000 bootstrap replications of the entire estimation process taking into account clustering at the municipality level.

F Monte Carlo to evaluate the presence of weak instruments bias

In this section, we provide details on the Monte Carlo simulation used to assess whether with our data and first stage we obtain bias due to weak instruments. To run the Monte Carlo we choose parameters to equal the estimated parameter estimates. To replicate the structure of our data we need to choose appropriately the covariance structure.

Define by e_n an n-vector of ones (the number of individuals in a community) and by I_n the n dimensional identity matrix.[†] The covariance matrix between outcome cognition and the two investments is denoted by Σ_d. The off diagonal blocks, reflecting spatial covariance within the community, are denoted by Σ_{off}. Then the covariance matrix for a community i with n individual observations can be written as

$$\Sigma_i = I_n \otimes \Sigma_d + (e_n \times e_n' - I_n) \otimes \Sigma_{off}$$

The errors for outcome cognition, time investments and material investments for all individuals in a community (cluster) is a $3n \times 1$ vector u_{3n} and is distributed as:

$$u_{3n} \sim N(0, \Sigma_i) \tag{4}$$

One difficulty was to decide on the value of the covariance matrix of the residuals from which to draw the errors in the investment equations and the production function. Given parameter values for the first stage (chosen to equal the ones we estimate in the real data), the sample size and the cluster structure, this drives the precision of estimation and of course the strength of the first stage. The most conservative approach is to use the estimated residuals from our model to compute Σ_d and Σ_{off}. The reason this is conservative is because the errors will contain an additional component corresponding to the estimation error of the

[†]In practice the number of individuals varies across communities, but we simplified the notation by using n instead of n_i. The Monte Carlo simulation uses the exact same structure of the data as in reality and the actual values of the exogenous variables, including prices and cognition.

factor scores, which will now be taken to be part of the error structure in the production function and the investment equations. The results using this approach are given under Covariance B below. The first stage is weaker than in the actual data with this covariance. In another approach we computed a covariance matrix after removing community specific means from the residuals for Σ_d. We then set the diagonal elements of Σ_{off} to 0.04, having tried a number of values so as to achieve a first stage closer in precision to our actual first stage. The result was that the estimation precision of the price coefficients in the material investment equation is a bit higher than in reality, but the precision of the conflict variable in the time investment equation is closer to the actual one. In both cases the rank test and the Cragg-Donald test are considerably weaker than in our actual data. The two covariances we use are given in Table F.1.

Table F.1: Covariances of the error structures used in the Monte Carlo simulations

Covariance A

	Within person Σ_d			Spatial covariance Σ_{off}		
	Material investment	Time investment	Cognitive skill	Material investment	Time investment	Cognitive skill
Material investment	0.623	0.202	-0.254	0.040	0	0
Time investment	0.202	0.561	0.033	0	0.040	0
Cognitive skill	-0.254	0.033	0.565	0	0	0.040

Covariance B

	Within person Σ_d			Spatial covariance Σ_{off}		
	Material investment	Time investment	Cognitive skill	Material investment	Time investment	Cognitive skill
Material investment	0.823	0.351	-0.299	0.147	0.132	-0.023
Time investment	0.351	0.815	0.032	0.132	0.207	-0.003
Cognitive skill	-0.299	0.032	0.661	-0.023	-0.003	0.048

The simulated data is generated as follows: we draw errors for all individuals jointly in

each community based on equation (4). This respects the actual cluster structure. Using the actual treatment status, prices and conflict values as well as the factor scores estimated for the child's and mother's baseline cognition in the actual data, we generate material and time investments for each household in our data. Given the generated investments and the exogenous variables just described, we generate the cognitive outcome for each child. For the production function we use as true value the estimated coefficient on the material investment (0.594) in column 2 of Table 6 and 0 for the time investment, which is what we interpret the results to imply. In other experiments (not reported here) we also used a positive coefficient for time investment as well, and our conclusions on bias do not change. Once we generate the latent factors for the child's cognition and the two investments, we then generate the equivalent of the original measurements, for the baseline factors (mother and child cognition), the cognitive outcome of the child in the follow-up period and the two investments themselves. To do this we use the measurement system in reverse (from latent factor to measurement) based on the factor loadings and variance of the measurement errors (assumed normal) that we estimated in the real data. This provides a simulated data set of a similar structure to the one we observe. We then estimate the model on this data. We do not use knowledge of the fact that the distribution is normal and we base our estimates on a distribution free approach, which can be done because the model is linear. We employ 1000 replications.[‡]

In Table F.2 we show the standard errors from the estimated model and compare them to the standard deviations of the Monte Carlo simulations. In the top panel we show the ones for the instruments in the material and time investments. These show that the simulation with Covariance B gets the standard deviations for the price coefficients very similar to the ones in the estimated model, but that the standard deviation of the conflict variable is too high, implying that the first stage under Covariance B is weaker than the one in the real data

[‡]The estimated production function is a bit more precise because we use Maximum likelihood to estimate the distribution of the factors as in Attanasio, Meghir, and Nix (2019). The Monte Carlo estimates the distribution of factors non-parametrically and hence is more conservative. In practice, this makes almost no difference to the estimates but slightly improves precision.

Table F.2: Standard errors of key parameters estimated in the true data and in simulated data

Panel A - Standard errors of the estimates of the coefficients on instruments in the material and time investment functions

	Estimated on true data		Estimated on simulated data Covariance A		Covariance B	
	Material	Time	Material	Time	Material	Time
Toy price	0.036	0.031	0.028	0.03	0.038	0.044
Food price	0.043	0.038	0.033	0.034	0.044	0.051
Conflict	0.037	0.028	0.028	0.031	0.039	0.044

Panel B - Standard errors of IV estimates of the coefficients on on material and time investment in the production function for cognitive skill

	Estimated on true data	Estimated on simulated data Covariance A	Covariance B
Material investment	0.307	0.289	0.406
Time investment	0.391	0.304	0.472

(as also implied by comparing the rank tests). Under covariance A the price coefficients have a standard deviation slightly lower than the standard errors in the estimated model but this time the coefficient on the conflict instrument has a very similar standard deviation to the estimated standard error in the time investment equation, where it is significant in reality. Importantly the standard deviations of the investment coefficients in the Monte Carlo are similar to those in the estimated model.

In Tables F.3 and F.4 we show the corresponding Monte Carlo simulations for the three equation system of material and time investments and cognitive skills. The coefficients of the production function in the Monte Carlo average out very close to the true values when we allow for endogeneity (as we should), while the model where we treat investments as exogenous show substantial bias. These results taken together imply that there is no evidence that the instruments are weak. More discussion and interpretation of the results is provided in the main text.

Table F.3: Estimates of parameters of the investment functions on simulated data

	True values		Estimates on simulated data			
			Covariance A		Covariance B	
	Material	Time	Material	Time	Material	Time
Intercept	-0.015	0.001	0.001	0.001	0.001	0.001
			(0.000)	(0.000)	(0.001)	(0.001)
			[0.001,0.002]	[0,0.002]	[0,0.002]	[0,0.002]
Treatment	0.209	0.318	0.189	0.327	0.173	0.292
			(0.062)	(0.065)	(0.083)	(0.098)
			[0.069,0.315]	[0.198,0.456]	[0.014,0.338]	[0.108,0.492]
Log child's cognitive skill (t)	0.130	0.068	0.13	0.075	0.121	0.068
			(0.039)	(0.039)	(0.043)	(0.048)
			[0.053,0.206]	[0,0.15]	[0.033,0.202]	[-0.035,0.16]
Log mother's cognitive skill	0.748	0.349	0.746	0.388	0.689	0.347
			(0.047)	(0.044)	(0.050)	(0.049)
			[0.654,0.835]	[0.305,0.472]	[0.593,0.787]	[0.259,0.447]
Log toy price	-0.096	-0.020	-0.091	-0.019	-0.083	-0.02
			(0.028)	(0.030)	(0.038)	(0.044)
			[-0.147,-0.036]	[-0.079,0.039]	[-0.158,-0.009]	[-0.113,0.064]
Log food price	0.091	0.042	0.089	0.042	0.081	0.038
			(0.033)	(0.034)	(0.044)	(0.051)
			[0.024,0.154]	[-0.025,0.108]	[-0.008,0.167]	[-0.061,0.133]
Maternal exposure to conflict	-0.009	-0.089	-0.01	-0.094	-0.008	-0.084
			(0.028)	(0.031)	(0.039)	(0.044)
			[-0.067,0.042]	[-0.156,-0.035]	[-0.085,0.068]	[-0.17,0.001]
Rank test (p-value)			0.042		0.029	
Cragg-Donald test (p-value)			0.06		0.021	

Note: Dependent variable in column 1, 3 and 5 (2, 4 and 6) is the log of material (time) investment measured at follow-up ($t+1$) simulated using either Covariance A or Covariance B. t refers to baseline/pre-treatment measurement. 95% confidence intervals in brackets based on 1,000 replications of the entire estimation procedure taking into account clustering at the municipality level.

Table F.4: Estimates of parameters of the investment functions on simulated data

	True values	Estimates on simulated data			
		Covariance A		Covariance B	
		OLS	IV	OLS	IV
Intercept	-0.019	0.000	0.000	0	0
		(0.000)	(0.001)	(0.000)	(0.001)
		[0,0]	[-0.001,0.001]	[0,0]	[-0.001,0.001]
Treatment	0.049	0.082	0.039	0.072	0.039
		(0.053)	(0.114)	(0.051)	(0.146)
		[-0.02,0.184]	[-0.188,0.275]	[-0.025,0.174]	[-0.231,0.315]
Log child's cognitive skill (t)	0.648	0.626	0.594	0.603	0.574
		(0.035)	(0.054)	(0.035)	(0.063)
		[0.56,0.696]	[0.486,0.697]	[0.539,0.671]	[0.452,0.68]
Log mother's cognitive skill	-0.075	0.174	-0.071	0.156	-0.06
		(0.041)	(0.207)	(0.040)	(0.277)
		[0.093,0.253]	[-0.511,0.302]	[0.076,0.233]	[-0.553,0.367]
Log of material investment	0.594	0.11	0.531	0.129	0.546
		(0.033)	(0.289)	(0.033)	(0.406)
		[0.044,0.173]	[-0.005,1.119]	[0.067,0.192]	[-0.079,1.349]
Log of time investment	0	0.194	0.016	0.213	0.004
		(0.034)	(0.304)	(0.033)	(0.472)
		[0.128,0.261]	[-0.691,0.567]	[0.15,0.276]	[-0.892,0.696]

Note: Dependent variable is the log of cognitive skill measured at follow-up (t+1) simulated using either Covariance A or Covariance B. t refers to baseline/pre-treatment measurement. 95% confidence intervals in brackets based on 1,000 replications of the entire estimation procedure taking into account clustering at the municipality level.

References

Attanasio, O., C. Fernandez, E. Fitzsimons, S. Grantham-McGregor, C. Meghir, and M. Rubio-Codina (2014). Using the infrastucture of a Conditional Cash Transfer programme to deliver a scalable integrated early child development programme in Colombia: A cluster randomised controlled trial. *British Medical Journal 349*, g5785.

Attanasio, O., C. Meghir, and E. Nix (2019). Investments in children and the development of cognition and health in india. NBER Working Paper No. 21740.

Bates, J., C. Freeland, and M. Lounsbury (1979). Measurement of infant difficultness. *Child Development 50*, 794–803.

Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). San Antonio, TX: Harcourt Assessment.

Caldwell, B. and R. Bradley (2001). *HOME Inventory and Administration Manual* (Third Edition ed.). Little Rock, AK: University of Arkansas for Medical Sciences.

Cattell, R. (1966). The scree test for the number of factors. *Multivariate Behavioral Research 1*, 629–637.

Dunn, L., E. Padilla, and D. Lugo (1986). *Manual del Examinador para el Test de Vocabulario en Imenes Peabody (Peabody Picture Vocabulary Test), Adaptaci Hispanoamericana (Hispanic-American Adaption)*. Minneapolis, MS: Pearson Assessments.

Fernald, L., P. Kariger, P. Engle, and A. Raikes (2009). *Examining child development in low-income countries: A toolkit for the assessment of children in the first five years of life*. Washington DC: The World Bank.

Gorsuch, R. (1983). *Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gorsuch, R. (2003). Factor analysis. In *Handbook of psychology: Research methods in psychology*, Volume 2, pp. 143–164. Hoboken, NJ: John Wiely & Sons, Inc.

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

Hamadani, J., F. Tofail, A. Hilaly, S. Huda, P. Engel, and S. Grantham-McGregor (2010). Use of Family Care Indicators and their relationship with child development in bangladesh. *Journal of Health Population and Nutrition 28*(1), 23–33.

Horn, J. (1965). A rationale and a test for the number of factors in factor analysis. *Psychometrika 30*, 179–185.

Jackson-Maldonado, D. (2011). Spanish cdi-iii development and preliminary results. Presentation at the International Congress for the Study of Child Language, Montreal.

Jackson-Maldonado, D. and B. Conboy (2011). Preliminary norms and validity for the Spanish CDI-III. American Speech-Language-Hearing Association (ASHA) Annual Conference. San Diego, CA.

Jackson-Maldonado, D., V. Marchman, and L. Fernald (2012). Short form versions of the Spanish MacArthur-Bates Communicative Development Inventories. In *Applied Psycholinguistics*. Cambridge: Cambridge University Press.

Jackson-Maldonado, D., D. Thal, V. Marchman, T. Newton, L. Fenson, and B. Conboy (2003). *MacArthur Inventarios del Desarrollo de Habilidades Comunicativas: User's guide and technical manual*. Baltimore, MD: Brookes.

Kaiser, H. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement 20*, 141–151.

Kariger, P., E. Frongillo, P. Engle, P. Britto, S. Sywulka, and P. Menon (2012). Indicators of family care for development for use in multicountry surveys. *J Heal Popul Nutr 30*, 472–486.

Putnam, S. P., M. Gartstein, and M. K. Rothbart (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior questionnaire. *Infant Behavior and Development 29*(3), 386–401.

Putnam, S. P., J. Jacobs, M. Garstein, and M. Rothbart (2010). Development and assessment of short and very short forms of the Early Childhood Behavior Questionnaire. International Conference on Infant Studies. Baltimore, MD.

Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement 1*(3), 385–401.

Raven, J. (1981). Manual for Raven's progressive matrices and vocabulary scales. In *Research Supplement No.1: The 1979 British Standardisation of the Standard Progressive Matrices and Mill Hill Vocabulary Scales, Together With Comparative Data From Earlier Studies in the UK, US, Canada, Germany and Ireland*. San Antonio, TX: Harcourt Assessment.

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

Velicer, W. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika 41*, 32l–327.