ESTIMATING THE PRODUCTION FUNCTION FOR HUMAN CAPITAL:
RESULTS FROM A RANDOMIZED CONTROL TRIAL IN COLOMBIA

Orazio Attanasio
Sarah Cattan
Emla Fitzsimons
Costas Meghir
Marta Rubio-Codina

Estimating the Production Function for Human Capital: Results from a Randomized Control
Trial in Colombia
Orazio Attanasio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina
NBER Working Paper No. 20965
February 2015
JEL No. H31,I24,I25,I28,I3,J38,O1,O15,O54

ABSTRACT

We examine the channels through which a randomized early childhood intervention in Colombia
led to significant gains in cognitive and socio-emotional skills among a sample of disadvantaged
children aged 12 to 24 months at baseline. We estimate the determinants of parents' material and
time investments in these children and evaluate the impact of the treatment on such investments.
We then estimate the production functions for cognitive and socio-emotional skills. The effects of
the program can be explained by increases in parental investments, emphasizing the importance
of parenting interventions at an early age.

Orazio Attanasio
Department of Economics
University College London
Gower Street
London WC1E 6BT
United Kingdom
and NBER
o.attanasio@ucl.ac.uk

Sarah Cattan
Institute for Fiscal Studies
7 Ridgmount Street
London, WC1E 7AE
United Kingdom
sarah_c@ifs.org.uk

Emla Fitzsimons
Institute for Fiscal Studies
7 Ridgmount St
London, WC1E 7AE
United Kingdom
e.fitzsimons@ioe.ac.uk

Costas Meghir
Department of Economics
Yale University
37 Hillhouse Avenue
New Haven, CT 06511
and IZA
and also NBER
c.meghir@yale.edu

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

An online appendix is available at http://www.nber.org/data-appendix/w20965

1 Introduction

The first five years of life lay the basis for lifelong outcomes (Almond and Currie, 2011). Due to rapid brain development and its malleability during the early years (Knudsen, 2004; Knudsen et al., 2006), investments during this period play a crucial role in the process of human capital accumulation. At this time however, many children are exposed to risk factors such as poverty, malnutrition and non-stimulating home environments preventing them from reaching their full potential, particularly in developing countries (Black et al., 2016; Grantham-McGregor et al., 2007; Lu et al., 2016). Thus children from poor backgrounds accumulate developmental deficits from a very early age (Lancet, 2016; Rubio-Codina et al., 2015). These factors are likely to play an important role in the intergenerational transmission of poverty.

There is increasing evidence that early childhood interventions can help overcome these detrimental factors and have positive effects on children's development in both the short- and long-term. Examples include the Jamaica study (Grantham-McGregor et al. (1991), Walker et al. (2011) and Gertler et al. (2014)), the Perry Preschool program (Heckman et al., 2010) and the Abecedarian experiment (Campbell and Ramey (1994), Campbell et al. (2014)). In Attanasio et al. (2014), we present the impacts of an 18-month long early childhood intervention in Colombia targeted at disadvantaged children aged 12-24 months old at baseline and evaluated by a randomized controlled trial. The intervention was based on the Jamaican model of psychosocial stimulation via weekly home visits based on the curriculum now known as "Reach-up and Learn", and also offered micronutrient supplementation. However, unlike the Jamaican program, it was designed to be scalable by training local women involved in the implementation of a large welfare program to administer the weekly home visits.

Attanasio et al. (2014) shows that stimulation led to highly significant improvements in cognition and language development measured immediately following end of

the intervention[1], and that micronutrient supplementation did not affect any outcome observed in the data. The paper also reported impacts on two raw measures of the home learning environment (measured by the Family Care Indicators). However, it did not put these results together into a model of cognitive and socio-emotional skill production that would allow assessing the (joint) role of parental investments and of the intervention in shaping child development in the earliest of years.

Building on these results, the main aim of this paper is to understand *how* the stimulation component of the intervention led to improvements in child development, described by both cognitive and socioemotional skills. For example, it could have led parents to make greater material and time investments in their children. But it could also have changed the production function for child skills, through the direct effect of the home visits as a new input or by changing the effectiveness of parental inputs. In what follows, we build a model of parental investments, taken as endogenous, and child skill formation to tease out the relative importance of these different mechanisms, a crucial step to better focus and increase the sustainability of interventions in the future. In so doing we also provide some of the first estimated models of parental investments and human capital production functions at such an early age in a context of poverty in a developing country.

We start by estimating the determinants of parental investments and assessing how the intervention changed parental choices. Indeed the way parents respond to such programs, which can be seen as a type of in-kind transfer, is an open question: the intervention could lead parents to reinforce their engagement with the child or instead crowd-out their investments. Gelber and Isen (2010), for example, provide evidence that the US early childhood program Head Start led to an increase in parental involvement, thus crowding-in household resources. In our treatment of the question here, we

---

[1]Using the Bayley (2006) scale of infant development, cognition improved by 26% of a Standard Deviation (SD) (p-value 0.002) and receptive language by 22% of a SD (p-value 0.032). These reported p-values are adjusted for testing 12 hypotheses.

exploit the experimental variation induced by the RCT and distinguish between material investments (e.g. books and toys around the house) and quality time investments (e.g. time spent by an adult in the household on education activities with the child).[2]

We then estimate production functions for child cognitive and socio-emotional skills. The main inputs we specify are baseline child skills, maternal skills, and material and quality time investments, which we treat as endogenous. Within this framework, we quantify by how much changes in parental investments contributed to improving child outcomes in the treatment group. We also test whether the intervention shifted the production function or otherwise changed its parameters, which, as discussed above, could reflect the direct effect of the stimulation provided by the home visitors or a change in the productivity of inputs.

The two waves of data we use were collected just before and just after the intervention and contain rich measures of child development, maternal skills and parental investments. Importantly, we collect information on materials and activities that have an educational aspect, thus enabling a clear interpretation of parental behavior as investments in their children. To our knowledge, our sample is one of the largest ever collected with this type of data in the literature evaluating stimulation programs. Even with such rich data however, estimating the parameters governing the skill formation process remains challenging for two reasons. First, inputs and outputs are likely to be measured with error. Second, inputs, especially investments, can be endogenous, if parental decisions respond to shocks or inputs that are unobserved to the econometrician. To deal with the measurement error issue, we use dynamic latent factor models as in Cunha and Heckman (2008) and Cunha, Heckman, and Schennach (2010). To deal with the possible econometric endogeneity of investments, we use an instrumental variable strategy, adapted to the latent factor structure of the model.

---

[2]See DelBoca et al. (2014) for a structural model of household choices and child development based on the PSID Child Development Supplement data and also including time and resource investments.

The estimates of the investment functions reveal important information about some of the drivers of developmental inequality: children with better initial cognitive skills receive more investments and, crucially, mothers with higher skill levels invest more in their children *given* the child's skills. In line with the existing literature, we find that a child's current stock of skills fosters the development of future skills, although we do find mean reversion.[3] Second, parental investments and in particular our measure of material investments are an important determinant of future cognitive and socio-emotional skills. This becomes even more evident once we control for the endogeneity of such investments in line with results from Cunha, Heckman, and Schennach (2010) and Attanasio et al. (2017) in very different contexts.[4]

With respect to the mechanisms underlying the impacts, we find that the intervention significantly increased parental investments among treated families compared to non-treated ones. This increase is the dominant reason underlying the observed impacts. The direct effect of the intervention, instead, is both small and insignificant. These two findings mean that the gains in cognitive and socio-emotional skills among children who received the intervention are mainly explained by changes in parental investments and imply that having the home visitor merely interact with the child for an hour a week, without trying to strengthen parenting practices, would have been unlikely to benefit children.

Beyond revealing important aspects of the process of human capital accumulation

---

[3]These features of the technology of skill formation are often referred to as *self-productivity* and *cross-productivity* (Cunha et al., 2006).

[4]The former use the children of the National Longitudinal Survey of Youth 1979, a longitudinal panel following the children of a representative sample of women born between 1956 and 1964 in the US. The latter use the Young Lives Survey for India, a longitudinal survey following the lives of children in two age-groups: a Younger Cohort of 2,000 children who were aged between 6 and 18 months when Round 1 of the survey was carried out in 2002, and an Older Cohort of 1,000 children then aged between 7.5 and 8.5 years. The survey was carried out again in late 2006 and in 2009 (when the younger children were about 8, the same age as the Older Cohort when the research started in 2002). See also Helmers and Patnam (2011) for the estimation of a linear production function in India. Finally, and also in line with the existing literature, we find that current skills and parental investments are complementary in the production of future skills, meaning that returns to investments are higher for children with better initial conditions.

and parental investments at a very early age, the importance of our results lie in two key findings. First, parents reinforce interventions by investing more not less - there is crowding in of material and time resources. Second, the intervention works because of the increase in parental investments and there is little evidence of a direct effect. If the intervention did not induce parents to invest more, it would have had no discernible effect. This, together with the mean reversion result that governs the longer-term impact of the intervention, emphasizes the key importance of improving parenting practices for the success and longer term sustainability of early childhood interventions.

Along with Heckman, Pinto, and Savelyev (2013) and a few other papers (Attanasio, Meghir, and Santiago, 2012; Duflo, Hanna, and Ryan, 2012; Todd and Wolpin, 2006), our paper illustrates how data from randomized trials can be profitably combined with behavioral models to go beyond the estimation of experimentally induced treatment effects and interpret the mechanisms underlying them. While there is a large literature evaluating the impact of early childhood interventions on child development, our paper innovates by complementing the information obtained from the RCT of a specific intervention with a model of skill formation and parental investment in order to understand the mechanisms behind the observed impacts.

In this sense, our paper shares the motivation of Heckman, Pinto, and Savelyev (2013) who document the channels through which the Perry Pre-School Program produced gains in adult outcomes. But our focus and methodology are different: Heckman, Pinto, and Savelyev (2013) perform a mediation analysis that decomposes linearly the treatment effects on adult outcomes into components attributable to early changes in different personality traits. Instead, we use a model in which parents make investment choices and human capital accumulates according to a production function, so as to interpret and explain the impacts induced by a successful intervention.

The focus of our intervention is also different. Unlike the Jamaican intervention,

which targeted malnourished children, and the Perry Preschool Program, which targeted children with specifically low cognition, we target a broader population. Our subjects are drawn from the beneficiaries of the Colombian Conditional Cash Transfer (CCT) program *Familias en Acción*, which covers the poorest 20% of the population.[5] In this sense, our program has the potential to serve as a model for early childhood policy that could be broadly implemented alongside CCT programs or other welfare programs targeting poor families.

The paper proceeds as follows. Section 2 provides some background on the intervention. Section 3 describes the data and the factor model approach we take to extract error-free measures of children's skills, parental skills and investments. Section 4 discusses the short-term impacts of the intervention and some suggestive evidence of its underlying mechanisms. Section 5 presents our theoretical framework and its empirical implementation. Section 6 presents the estimates of the model and discusses their implications for our understanding of the intervention. Section 7 concludes.

2   Background on the intervention and its evaluation

The early childhood program analyzed in this paper was targeted at children aged between 12 and 24 months living in families receiving the Colombian CCT program, which targets the poorest 20% of households in the country. The intervention lasted 18 months, starting in early 2010. Appendix A contains a detailed description of the program's design, implementation and delivery. Here we summarize the key aspects.

The program was implemented in semi-urban municipalities in three regions of central Colombia, covering an area around the size of California. It had two components: psychosocial stimulation and micronutrient supplementation. The stimulation curriculum was based on the Jamaican home visiting model, which obtained positive

---

[5]See Attanasio et al. (2010) for a description and evaluation of that program.

short- and long-term effects (Grantham-McGregor et al. (1991), Walker et al. (2006, 2011) and Gertler et al. (2014)). The protocols designed by Grantham-McGregor et al. (1991) for Jamaica were adapted to be culturally appropriate for Colombia. The aims of the home visits were to improve the quality of maternal-child interactions and to assist mothers to participate in developmentally-appropriate learning activities, centered around daily routines and using household resources as learning tools.

We implement two key innovations vis-a-vis the Jamaican intervention with scalability and sustainability in mind. Indeed our program was specifically designed to go beyond the earlier small scale and tightly supervised efficacy trials. The first was that the intervention was implemented on a much larger scale than in Jamaica, covering a large part of the country and obtaining much larger sample sizes. The second was that the intervention was designed to be delivered by women drawn from the local community, with no specific prior professional experience.

To this end, home visitors were drawn from a network of local women, created by the administrative set-up of the CCT program. Every 50-60 beneficiaries elect a representative who is in charge of organizing social activities and acts as mediators between them and the program administrators. These women, known as *Madre Líderes* (MLs), are beneficiaries of the program themselves. Given they are selected by their peers, one can deduce that they enjoy the trust of the community and are probably more entrepreneurial and proactive than the average beneficiary. In terms of specific characteristics they are on average about 10 years older (37) and have about one more year of education (8.5) than the subject mothers. Their score on a vocabulary test is slightly higher than that of the mothers.[6] Finally, as mentioned in the introduction, another distinct feature of our intervention is that we targeted a more general poor population, namely the beneficiaries of the CCT program, which in Colombia is offered

---

[6]To measure vocabulary, we use the Peabody Picture Vocabulary Test (PPVT). The scores of the MLs was 28.2 versus 26.9 for the subject mothers in our sample. The difference has a p-value of 0.061.

to the 20% poorest segment of the population, as compared to the extreme disadvantage of the malnourished population targeted by the Jamaican experiment.

The intervention was evaluated through a cluster randomized controlled trial involving the random allocation of 96 municipalities across central Colombia. After first stratifying into three large regions, 32 municipalities in each were randomly assigned to one of 4 groups: (i) psychosocial stimulation, (ii) micronutrient supplementation, (iii) both, and (iv) control. In each municipality, 3 MLs were selected and the children aged 12-24 months of the beneficiary households represented by each of these MLs were recruited to the study. There was a total of 1,429 children living in 96 towns in central Colombia. Possibly because the MLs are such trusted figures in their communities, compliance was high and the average number of home visits made was 63, which is 81% of those scheduled. The attrition rate between baseline and follow-up was around 10% across treatment arms, and the difference in loss among the groups was not statistically significant.[7]

As reported in Attanasio et al. (2014), there was no significant impact of micronutrient supplementation on any child developmental outcomes. In this paper, therefore, we focus on the psychosocial stimulation arm of the program and we refer to the "treated" group as those children who received the stimulation component of the intervention (groups i and iii) and to the "control" group as those children who did not (groups ii and iv).

Individuals randomized into our intervention were all eligible for and receiving subsidies from the CCT program. On average, households had been part of the CCT program for 21 months at baseline. This feature is common between treatment and control communities, but it is true that the context in which our program was implemented and in particular the existence of the CCT may be a factor in how effective

---

[7]As we explain in Section 3.1, our data at baseline and at follow-up come from a household survey and from direct assessments administered to children in a community centre. The attrition rate for the household survey was 6.9%. The attrition rate for the direct assessments was 10.7%.

the program was. This, of course, is related to the more general issue of extrapolating the effects of the program to other contexts outside the support of the data. Nevertheless, CCT programs are quite common in low-and-middle income countries and consequently the context is directly relevant to many other countries besides Colombia.

Finally, a frequently asked question is whether the intervention is just "teaching to the test" without leading to genuine advances in cognition. First, implementation of the curriculum has been shown to have long-run effects on cognition (Walker et al., 2005, 2011) and labour market outcomes (Gertler et al., 2014). This in itself is evidence that it can induce deep changes in achievement rather than just teach children to remember a few activities and perform better on a test. More generally, the intervention curriculum emphasizes cognitive, language and socio-emotional development through play and the promotion of mother-child interactions. While some of the play activities specifically address the type of cognitive and fine motor skills (building towers with blocks, tracing lines) and concepts (shapes, sizes, colors) that are assessed in developmental tests, the focus is on learning through play in a supportive and stimulating environment. Activities are introduced progressively and in developmental order to facilitate scaffolding - i.e. increasing or decreasing the challenge based on the child's performance - and there is a strong emphasis on praising attempts and not only successes. The approach is aimed at promoting attention to task, perseverance and self-esteem; similarly, there is a strong focus on labelling the environment and looking at picture books together, which are activities that enrich vocabulary and promote bonding, attention (i.e. following a story) and other cognitive abilities (i.e. linking concepts, understanding cause and effect relations). All of these skills are associated with improved school readiness, school attainment and other outcomes linked to socio-economic success in life.

3 Data and measurement system

In this section, we describe the data we use, which was collected around the evaluation of the parenting intervention mentioned above. We then discuss an effective way of extracting the relevant information from such rich data with a measurement system that explicitly takes into account the relationship between relevant factors and available measures and the presence of measurement error.

3.1 Data

The main data we use in this paper comes from two rounds of data collection: before the intervention started (baseline) and just after it ended 18 months later (follow-up). In each round, information was collected in two ways: via a household survey in the home and via tests directly administered to children in a community centre. At the end of the paper, we also briefly discuss results from a second follow up (FU2), two years after the end of the intervention, although we do not use those data here.

The household surveys contain information on an extensive set of socio-economic and demographic characteristics, alongside a wealth of information around parenting, parental characteristics and maternal skills, including mothers' years of education, verbal ability, IQ, depressive symptoms and knowledge of child development. Among others, we administered among mothers the Raven's progressive matrices to test for IQ and the CES-D 10-item scale to assess depressive symptoms.

To measure children's developmental outcomes, we collected data based both on maternal reports and on direct assessment of the child. The measures of child development that we collected in the home setting via maternal report include: language development (that is, the number of words and complex sentences the child can say) using the vocabulary checklists in the Spanish Short-Forms of the MacArthur-Bates Communicative Development Inventories I and II (MacArthur); child temperament

using Bates' Infant Characteristics Questionnaire (ICQ); and the attentional focusing and inhibitory control scales of the short versions of the Early Children's Behavior Questionnaire (ECBQ). All of these were measured using age-appropriate items pre- and post-intervention, with the exception of the ECBQ which was administered at follow-up only. In addition to these assessments via maternal reports, trained psychologists administered the Bayley Scales of Infant and Toddler Development III (Bayley) in community centres.[8] These direct assessments of the child took place over an average period of 1.5 hours and were aimed at measuring children's cognitive, language and motor development in depth.

The household surveys also contain detailed information on parental investments. We used a slightly modified version of the UNICEF Family Care Indicators (FCI) (Frongillo, Sywulka, and Kariger, 2003) which is based both on interviewer observations and maternal reports of the home environment. Specifically, this instrument includes interviewer observations of the types and numbers of play materials around the home and maternal reports of the types and frequency of play activities performed by the primary caregiver or any other adult older than 15 with the child in the last 3 days.

Examples of play materials include toys designed for learning shapes, toys that induce physical movement, coloring books, and picture books. Examples of play activities include reading or looking at picture books together, telling stories, and labelling items in the home. Importantly therefore, the instrument affords the possibility of distinguishing between parental investments in 'materials' and in 'quality time', which are likely to have different costs and perhaps different (but possibly complementary) effects on child development.

As we want to assess the role of parental investment and distinguish it from the direct role that the intervention might play on child development, in measuring materials

---

[8]See Jackson-Maldonado et al. (2012) for MacArthur-Bates scales, Bates et al. (1979) for the ICQ and Putnam et al. (2006) for ECBQ and Bayley (2006).

we instructed the data collectors to gather separate information about those materials (such as certain books and toys) that were directly provided by the intervention. When estimating the distribution of the factors measuring the two different types of parental investments, we use the measures that are not linked directly to the intervention.

Finally, in addition to the survey data collected around the evaluation of the intervention, we also use additional data sources to obtain information on municipality-level variables that we use as instruments. In particular, we use data on prices (of toys and food) and on maternal childhood exposure to violence. Appendix B provides details on all the measures of child development, maternal skills and parental investments collected as part of the survey and describes the auxiliary data sources we use to construct our instruments.

Appendix Table A.1 reports the baseline characteristics of children, their mothers and their households. At baseline, the children are on average aged 18 months. About 10% of them were born premature and 14% of them were stunted. On average, their mothers are 26 years old, have about 7.5 years of education and two-thirds of them are either married or cohabiting. There were no compromises to the randomization protocol and hence there is no reason to believe there is any bias. Most baseline characteristics are very well balanced including the baseline skills of the children. Although the mean of a few characteristics is significantly different between treated and controls when tested individually (specifically among CESD scale items), none of these differences are significant at all when we allow for multiple hypothesis testing using the Romano and Wolf (2005) procedure.

3.2 Factor Models and the measurement system

Our main aim is to interpret the experimental results within the context of a model of parental investments and human capital production functions. To fix ideas, suppose

we wish to estimate a production function for child skills:

$$\theta_{t+1} = f_{t+1}(\theta_t, I_{t+1}, P_t, X_t, \eta_{t+1}) \tag{1}$$

where θ_t and θ_{t+1} are vectors of the child's skills at t and $t+1$ respectively, I_{t+1} are parental investments that occur between the realizations of θ_t and θ_{t+1}, P_t are maternal skills measured at baseline, X_t is a vector of baseline household characteristics, such as household composition and η_{t+1} are random shocks to child development. The production function allows us to understand the pathways through which the experiment might affect outcomes: changes in parental investments and/or changes in the production function $f_{t+1}(\cdot)$, reflecting, for example, better use of parental inputs.[9]

As Cunha and Heckman (2008) explain, an important obstacle to estimating such a function is that the skills and investments are inherently unobservable. The various measures described in Section 3.1 can be viewed as error ridden indicators for these underlying latent factors. Using any one set of these measures in place of the latent factors could lead to severely biased results, whether the model is linear or not. We thus follow the approach of Cunha and Heckman (2008) and Cunha et al. (2010)[10] and develop a measurement system linking the observed measures to latent factors and estimate the distribution of such factors.

Suppose we have $\mathcal{M}_{kt}^{\theta}$ measures of child's skill θ_t^k of type k (e.g. cognitive or socio-emotional skills) in period t. Moreover, we also have $\mathcal{M}_k^P$ measures of maternal skills P^k of type k. Finally, we have $\mathcal{M}_{\tau t}^I$ measures of parental investments I_t^τ of type τ (e.g.

---

[9]We use maternal skills as measured at baseline. However, we find no evidence of a treatment impact on any measures of cognitive skills or socio-emotional skills of the mother (the main primary caregiver in most households in our sample). This is in line with psychological evidence indicating that cognition (as measured by IQ) is rank stable by the age of 10 (Almlund et al., 2011). While it is more plausible that the intervention could have changed maternal socio-emotional skills, we find no such evidence. Had these maternal measures changed they could have been an additional channel of impact.

[10]More broadly this approach relates to the identification and estimation of nonlinear models with classical measurement error (Schennach, 2004, 2007).

time or material investments) made between $t-1$ and t. We denote m_{kjt}^{θ} the j-th measure of child's skill of type k at t, m_{kj}^{P} the j-th measure of maternal skill of type k, and $m_{\tau jt}^{I}$ the j-th measure of parental investment of type τ at t. As we estimate a different joint distribution of latent factors for the control and treated groups, in what follows we index the measures and latent factors by the treatment subscript d, where $d=0$ refers to the control group (no home visits) and $d=1$ refers to the treatment group (some home visits).

As is common in the psychometric literature, we assume a dedicated measurement system, that is one in which each measure only proxies one factor (Gorsuch, 1983, 2003). Although it is not necessary for identification, we maintain this assumption because it makes the interpretation of the latent factors more transparent and we find clear support for such a system in the data (see Appendix C). Assuming each measure is additively separable in the (log) of the latent factor it proxies,[11] we write the following system of equations mapping the j-th measure observed at some date t to the k-th latent (unobserved) factor for that date:

$$m_{kjdt}^{\theta} = \mu_{kjt}^{\theta} + \alpha_{kjt}^{\theta} \ln \theta_{dt}^{k} + \epsilon_{kjt}^{\theta} \tag{2}$$

$$m_{kjd}^{P} = \mu_{kj}^{P} + \alpha_{kj}^{P} \ln P_{d}^{k} + \epsilon_{kj}^{P} \tag{3}$$

$$m_{\tau jdt}^{I} = \mu_{\tau jt}^{I} + \alpha_{\tau jt}^{I} \ln I_{dt}^{\tau} + \epsilon_{\tau jt}^{I} \tag{4}$$

where the terms μ_{kjt}^{θ}, μ_{kj}^{P} and $\mu_{\tau jt}^{I}$ are intercepts, the terms α_{kjt}^{θ}, α_{kj}^{P} and $\alpha_{\tau jt}^{I}$ are factor loadings, and the terms ϵ_{kjt}^{θ}, ϵ_{kj}^{P} and $\epsilon_{\tau jt}^{I}$ are mean zero measurement error terms which are assumed independent of the latent factors and of each other.[12]

---

[11]The measurement equations are specified in terms of the log latent factors. This ensures that the factors themselves only take positive values as required by the model.

[12]The assumption that the errors are independent of each other can be relaxed somewhat. Some of the child cognitive outcomes, for example, are based on child level observations and are collected by a trained psychologist in community centers, while others are based on maternal reports and are collected in the home (on a different day) by a different interviewer. However, it is certainly possible that measurement errors are correlated, even in this case from say child behavior, the implications of

An assumption we have made in writing the system above is that the measurement system is invariant between treated and controls. This implies that any differences in the distribution of observed measures between the control and treated groups result from differences in the distribution of the latent factors only. As we show in Appendix D, none of our results are sensitive to whether we allow for the measurement treatment to be affected by the treatment or not.

Because the latent factors are unobserved, identification requires normalizations to set their scale and location (Anderson and Rubin, 1956). We set the scale of the factors by setting the factor loading on one of the measures (say the first) of each latent factor to 1, that is: $\alpha^{\theta}_{k1t} = \alpha^{P}_{k1} = \alpha^{I}_{\tau 1t} = 1$, $\forall t, \tau = \{M, T\}$ and $k = \{C, S\}$. When it comes to the child's skills, we normalize the factor loading on the same measures at baseline and follow-up.[13] We set the location of all the factors by fixing the mean of the latent factors in logs to 0 for the control group; the difference between the treatment group's location and that of the control group (which is set to zero) is taken to be the average effect of the treatment.

With the assumptions and normalizations already made and based on the Kotlarski theorem and further extensions, Cunha et al. (2010) show that both the distribution of measurement errors and the latent factor distribution are non-parametrically identified so long as we have at least three measures with nonzero factor loadings corresponding to each latent factor.[14] While these assumptions are sufficient for identification, some of them could be relaxed as shown in Cunha et al. (2010).[15] For instance, the same measure could be allowed to load on several factors, provided there are some dedicated measures. It would also be possible to allow measurement error to be correlated across

---

which should be studied in future research.

[13]For cognitive skills, we define the scale based on the Bayley cognitive score both at baseline and follow-up. For socio-emotional skills, we normalize the factor loadings on the item measuring difficulty in child's temperament in the ICQ.

[14]See also Schennach (2004), Schennach (2007), Hu and Schennach (2008), Carneiro, Hansen, and Heckman (2003), Heckman, Pinto, and Savelyev (2013) and Cunha and Heckman (2008).

[15]See also Carneiro et al. (2003) and Cunha and Heckman (2008)

measures of the same factor, as long as there is one measure whose error is independent from those of other measures of the same factor.

A question of practical importance relates to the scale of the latent factors and what they actually mean for measures of interest such as earnings. This is the issue of anchoring discussed in Cunha et al. (2010) who provide a theoretical treatment.[16] In our paper, we normalize the cognitive factor in both periods to the Bayley cognitive scale. This has a cardinal interpretation (the number of tasks completed correctly), and the same test is applied across different ages (up until 42 months), allowing for comparability. For socio-emotional skills we also normalize to the same ICQ item (whether the child is difficult) in both periods. The lack of long-term longitudinal data prevents us from converting these units to future earnings or other adult outcomes of interest.

3.3   Specification of the measurement system

To implement the measurement system above, we first perform an exploratory factor analysis, reported in Appendix C, to identify in a preliminary step the relevant measures and their allocation to factors. We then allocate measures observed in the data to particular factors, as is shown in Table 1. The factor loading on the first measure is normalized to one and thus this measure defines the scale of the latent factor.

As reflected in the table, we did not necessarily use the same set of measures of the child's skill at baseline and at follow-up, the main reason being that we only included age-appropriate items that provide relevant information about the latent skill. For example, the MacArthur item measuring the number of complex phrases a child can

---

[16]Cunha et al. (2010) provide a general theoretical treatment of anchoring and in their main empirical results they anchor the measure of skills measured at the oldest age to years of education. They then assume that the same anchoring scale applies to measures of cognition and socio-emotional skill measured at earlier ages. Nielsen (2015) discusses using ordinal tests scores to measure achievement gaps, and Agostinelli and Wiswall (2016) discuss how rescaling in a multiperiod production function can lead to biases in the estimation of the substitution elasticity.

Table 1: Measurement system

Latent factor	Measurement	Survey	% Signal	
			Controls	Treated
Child's cognitive skills at FU (θ_{t+1}^C)	Bayley: cognitive	FU	78%	79%
	Bayley: receptive language	FU	75%	76%
	Bayley: expressive language	FU	78%	79%
	Bayley: fine motor	FU	59%	61%
	MacArthur: words the child can say	FU	64%	65%
	MacArthur: complex phrases the child can say	FU	52%	54%
Child's cognitive skills at BA (θ_t^C)	Bayley: cognitive	BA	70%	70%
	Bayley: receptive language	BA	73%	72%
	Bayley: expressive language	BA	75%	74%
	Bayley: fine motor	BA	60%	59%
	MacArthur: words the child can say	BA	45%	44%
Child's socio-emotional skills at FU (θ_{t+1}^S)	ICQ: difficult (-)	FU	74%	71%
	ICQ: unsociable (-)	FU	33%	30%
	ICQ: unstoppable (-)	FU	59%	55%
	ECBQ: inhibitory control	FU	73%	69%
	ECBQ: attentional focusing	FU	27%	24%
Child's socio-emotional skills at BA (θ_t^S)	ICQ: difficult (-)	BA	68%	71%
	ICQ: unsociable (-)	BA	28%	31%
	ICQ: unadaptable (-)	BA	35%	38%
	ICQ: unstoppable (-)	BA	22%	25%
Material investment at FU (I_t^M)	FCI: no. of different types of play materials	FU	94%	97%
	FCI: no. of coloring and drawing books	FU	17%	29%
	FCI: no. of toys to learn movement	FU	61%	76%
	FCI: no. of toys to learn shapes	FU	69%	82%
	FCI: no. of shop-bought toys	FU	61%	76%
Time investment at FU (I_t^T)	FCI: no. of different types of play activities in last 3 days	FU	87%	93%
	FCI: no. of times told a story to child in last 3 days	FU	66%	81%
	FCI: no. of times read to child in last 3 days	FU	73%	85%
	FCI: no. of times played with toys in the last 3 days	FU	55%	72%
	FCI: no. of times named things to child in last 3 days	FU	56%	73%
Mother's cognitive skills at BA (P^C)	Mothers' years of education	FU	54%	50%
	Mother's Raven's score (IQ)	BA	54%	51%
	Mother's vocabulary	FU2	65%	62%
	FCI: no. of books for adults in the home	BA	39%	36%
	FCI: no. of magazines and newspapers in the home	BA	20%	19%
Mother's socio-emotional skills at BA (P^S)	CESD: did you feel depressed? (-)	BA	70%	73%
	CESD: are you bothered by what usually don't? (-)	BA	42%	45%
	CESD: did you have trouble keep mind on doing? (-)	BA	49%	52%
	CESD: did you feel everything you did was an effort? (-)	BA	45%	49%
	CESD: did you feel fearful? (-)	BA	47%	51%
	CESD: was your sleep restless? (-)	BA	34%	38%
	CESD: did you feel happy?	BA	46%	50%
	CESD: how often did you feel lonely last week? (-)	BA	51%	55%
	CESD: did you feel you couldn't get going? (-)	BA	48%	52%

Note: This table shows the measures allowed to load on each latent factor, as well as the fraction of the variance in each measure that is explained by the variance in signal, for the control and treatment groups separately. "BA" refers to Baseline, "FU" refers to the first-follow-up survey and "FU2" refers to the second follow-up survey collected 2 years after the intervention ended. The symbol (-) indicates that the scoring on these measures was reversed so that a higher score on the corresponding latent factor means a higher level of skill.

say is too advanced for children at 1-2 years old and hence was only administered at follow-up when children were between 2.5 to 3.5 years old. Similarly, with respect to socio-emotional skills, the ECBQ is designed to measure temperament among children aged 3-7 and therefore was only administered at follow-up.[17] However, in both rounds, we use the same measure to normalize the child's baseline cognitive and socio-emotional skills.

In our model we use mother's skills to control for parental background. During the data collection process, we had to focus only on the mother's skills (who is almost always the principal caregiver and often a single mother) because of resource constraints and in order to keep interview times at a reasonable level. In so doing, it is possible that we miss the influence of the father; however, we expect to be capturing at least some of that by conditioning on the baseline skills of the child. We use baseline measures to extract two factors measuring the mother's cognitive and socio-emotional skills, with the exception of the vocabulary test, which was administered at follow up and the Raven's score which was administered at a later round of data collection (2 years after the end of the intervention). In both cases we checked and the intervention had no impact on the scores.

The parameters of the measurement system for treatment and control are estimated together with the latent factor distributions as described above. To do so, we use the estimation method described in Attanasio et al. (2017), which approximates the joint distribution of the latent factors by a mixture of normals (as in Cunha et al. (2010)) and the measurement error distribution by a normal distribution. We report estimates of the factor loadings and distribution of measurement errors in Appendix C. To assess the extent of information relative to measurement error contained in each of the measures, we compute the signal-to-noise ratio measuring the fraction of the variance of each

---

[17]The ICQ is in principle designed for children up to 2 years old. We administered the same questions of the ICQ at baseline and follow-up after consultation with the developer of the test.

measure driven by signal. For example, for the j-th measure of child's skills of type k, this ratio is defined as:

$$s_j^{\ln \theta^k} = \frac{(\alpha_j^k)^2 \, Var(\ln \theta^k)}{(\alpha_j^k)^2 \, Var(\ln \theta^k) + Var(\epsilon_j^k)}$$

where we have assumed that the j-th measure of latent factor θ^k can be written, simplifying notation, as:

$$m_j^\theta = \mu_j^k + \alpha_j^k \ln \theta^k + \epsilon_j^k$$

The last two columns of Table 1 report the signal-to-noise ratio for each of the measures used in the analysis for the control and treated groups separately. These numbers can be different because the joint distribution of latent factors is allowed to be different between the two groups. Clearly, there is much variation in the amount of information contained in each measure of the same factor. For example, 78% of the variance in the *Bayley: Cognitive* item is due to signal, whereas only 51% of the variance in the *Mac Arthur: Complex Phrases* item is due to signal. Overall, most measures are far from having 100% of their variance accounted for by signal, which illustrates the usefulness of the latent factor approach in modeling human capital accumulation and parental investments: without such an approach, one would risk to obtain severely attenuated coefficients, masking the importance of investments and background variables on child development.

4 Short-term impacts on child outcomes and parental investments

In this section, we document the impacts of the intervention on child's cognitive and socio-emotional development as well as parental investments, observed at first follow-up, just after the 18 month-long intervention ended. Impacts on the latter provide the

basic input to perform the mediation analysis discussed below to uncover the mechanisms behind the observed impacts on children outcomes. We focus on the impact of the psychosocial stimulation component of our intervention because there were no significant impact of micronutrient supplementation on any child developmental outcomes (Attanasio et al., 2014).[18]

4.1 Impacts on child development

Each panel of Table 2 reports the estimated impacts of receiving the home visits on one of four sets of outcomes: (i) cognitive development; (ii) socio-emotional development; (iii) parental investment in play materials; (iv) parental investment in play activities. Impacts on the Bayley outcomes outcomes and two of the FCI measures were previously reported in Attanasio et al. (2014) and are repeated here for completeness.[19] In addition to the impact on each measure, we also report the impact on the mean of the corresponding log latent factor. The results in the first panel imply an increase of 0.25 of a standard deviation (SD) in cognitive development and an increase of 0.175 SD in receptive language, assessed using the Bayley.[20] The cognitive factor summarizing all these effects shows a substantial and significant increase of 11% (0.11 log points) amongst the treated group relative to the control group. The second panel of the table also shows that the intervention led to an overall improvement in socio-emotional development (p-value<0.05).

In Figure 1, we plot the estimated densities of some of the factors for the control and treated groups and perform a Kolmogorov-Smirnov (K-S) test of the hypothesis that

---

[18]If we explicitly control for the fact that half of the stimulation group also received micronutrient supplementation, the impact on cognition and receptive language remains virtually the same, with a very small increase in the point estimates we report below (see Appendix Table D.1).

[19]Here we include among the control group those who obtained micronutrient supplementation, since that intervention had no effect on any developmental measure.

[20]These treatment effects are slightly different from those reported in Attanasio et al. (2014) because in this paper we estimate the impact of psychosocial stimulation by pooling the two groups that received it and the two groups that did not, while Attanasio et al. (2014) estimates the impact of each of the four arms of the intervention separately.

Figure 1: Kernel densities of latent factors

(a) Children's cognitive skills, baseline

(b) Children's socio-emotional skills, baseline

(c) Children's cognitive skills, follow-up

(d) Children's socio-emotional skills, follow-up

(e) Material investments - follow-up

(f) Time investments - follow-up
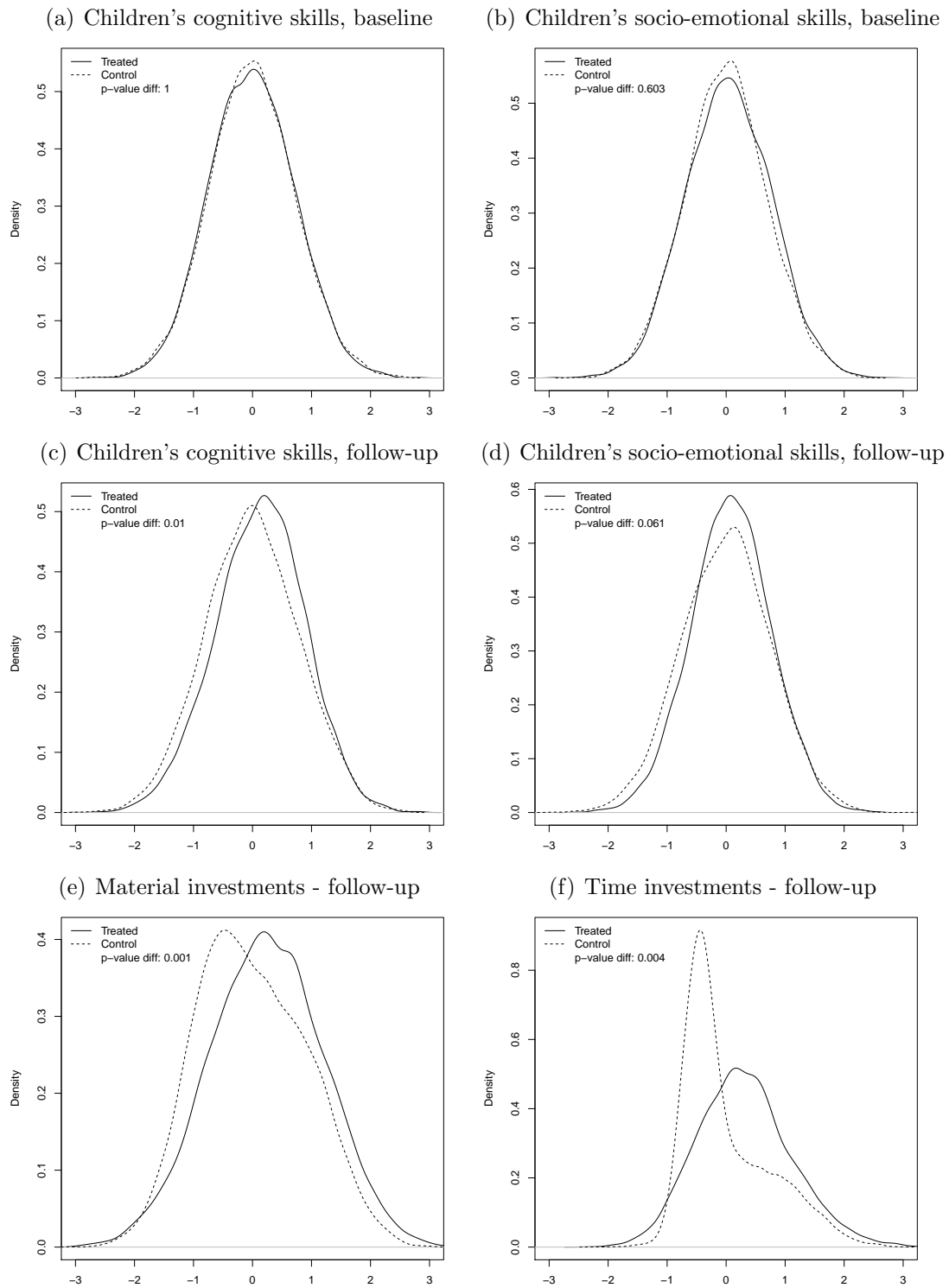
Note: These kernel densities are constructed using 10,000 draws from the estimated joint distribution of latent factors for the control group and for the treated group. For each factor, we perform a Kolmogorov-Smirnov test using the bootstrap and accounting for the entire estimation procedure. p-values reported in each panel.

Table 2: Treatment impacts on raw measures and factors

	Treatment effect		
	Point estimate	Standard error	Sample size
A - Child's cognitive skills at follow-up			
Bayley: cognitive	0.250	(0.063)	1,264
Bayley: receptive language	0.175	(0.063)	1,264
Bayley: expressive language	0.032	(0.062)	1,263
Bayley: fine motor	0.072	(0.060)	1,262
MacArthur: words the child can say	0.092	(0.064)	1,322
MacArthur: complex phrases the child can say	0.058	(0.054)	1,322
Cognitive factor	0.115	(0.051)	
B - Child's socio-emotional skills at follow-up			
ICQ: difficult (-)	-0.041	(0.054)	1,326
ICQ: unsociable (-)	-0.075	(0.045)	1,326
ICQ: unstoppable (-)	-0.032	(0.054)	1,326
ECBQ: inhibitory control	-0.003	(0.058)	1,323
ECBQ: attentional focusing	0.070	(0.049)	1,323
Socio-emotional factor	0.087	(0.044)	
C - Material investment at follow-up			
FCI: no. of different types of play materials	0.215	(0.064)	1,326
FCI: no. of coloring and drawing books	-0.133	(0.056)	1,326
FCI: no. of toys to learn movement	-0.048	(0.065)	1,326
FCI: no. of toys to learn shapes	0.416	(0.088)	1,326
FCI: no. of shop-bought toys	0.024	(0.061)	1,326
Material investment factor	0.227	(0.069)	
D - Time investment at follow-up			
FCI: no. of different types of play activities in last 3 days	0.277	(0.050)	1,326
FCI: no. of times told a story to child in last 3 days	0.138	(0.060)	1,326
FCI: no. of times read to child in last 3 days	0.362	(0.062)	1,326
FCI: no. of times played with toys in last 3 days	0.175	(0.056)	1,326
FCI: no. of times named things to child in last 3 days	0.137	(0.048)	1,326
Time investment factor	0.302	(0.068)	

Note: All scores have been internally standardized non-parametrically for age and are expressed in standard deviation units (see Appendix B for details about the measures and the standardization procedure). The effects relating to the *latent factors* are in log points. 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 for whether the child received any psychosocial stimulation and controlling for child's sex; tester effects and baseline level of the outcome.

the corresponding CDFs are equal to each other (the p-values of the tests are reported

in the figure and have been derived using the bootstrap).[21] The first two panels show the distribution, in treatment and control villages, of cognitive and socio-emotional skills at baseline. The two densities overlap each other and the K-S test cannot reject that they are equal to each other, thus confirming that our sample is balanced. The following two panels depict the distribution of cognitive and socio-emotional factors at follow-up. In the case of the cognitive factor, we see that the shift in the mean reported in Table 2 reflects a shift in the entire distribution. For the socio-emotional factor, however, the shift occurs mainly for children below the median.

4.2   Suggestive evidence on the mechanisms behind the impacts

In the last two panels of Figure 1, we notice a strong shift to the right of the distributions of both the material and time investment factors. For either type of investments, the K-S strongly rejects the equality of the corresponding densities between control and treated groups. The bottom two panels of Table 2 focus on the mean impacts of the stimulation intervention on parental investments and indicate substantial impacts on several individual items of the Family Care Indicator (FCI), as well as on the two latent factors measuring investments. Panel D of Table 2 shows that all types of time activities increase, but among play materials the increase is not uniform (Panel C). Specifically, there is an increase in most toys but a reduction in coloring books, which may reflect crowding out specially because the home visitors intentionally left picture books behind. The overall material investment factor registers a highly significant increase however.

As mentioned above, the measure of materials relates specifically to items provided by the parents. Although the home visitors were supposed to take away all intervention toys (with the exception of picture books) upon the completion of the last home visit, some were left behind at the end of the intervention as is evident from the summary

---

[21]The estimation method used for this purpose is based on Attanasio et al. (2017)

statistics reported in Appendix Table D.3.[22] However, since we are able to separately measure parental contributions from intervention materials, this does not pose any problem for our analysis, and whenever we refer to material investments we exclude intervention play materials and only keep items provided by the parents. The impact of the intervention materials, if any, will be captured by the treatment dummy; we return to this point when interpreting our results.

As we show above, the measures of parental inputs relating to materials and quality time both increased. As argued by DelBoca, Flinn, and Wiswall (2014) it may be important distinguish these inputs because they can have different impacts on child development. On the other hand the inputs are clearly related and the two factors are correlated (0.64). Thus one could imagine a more parsimonious approach where they are combined into one investment factor. However, another good reason for keeping them separate relates to the way they are measured. Materials are actually observed and enumerated by the surveyors; quality time items are measured by maternal self-reports, which may make them noisier measures of parental investment. This may also make them more likely to be subject to intervention bias due to the fact that mothers in the treatment group may exaggerate the extent to which they engage in developmental activities with the children. Without this implying that there is no information in the measures of time investments, these may be less reliable measures of parental behavior than material investments, which are directly observed and also likely to reflect actual effort by the parents. We will return to this important issue when interpreting the results.

---

[22]The median number of days between the end of the intervention and data collection was 10. Almost all interviews were completed within 40 days. Very few households were interviewed a few days before the end of the intervention. Omitting these households from the analysis leaves the results completely unchanged.

5 The accumulation of human capital in the early years and the role of the intervention

To better understand the determinants of early childhood development and to explore the way that the intervention affected outcomes we now specify a model of parental investments and child skill formation, where skills take two dimensions, namely cognitive and socio-emotional skills. We use such a model to inform the mediation analysis aimed at explaining the channels through which the intervention generated the impacts documented in Section 4. An important element of the model is that parents can choose to invest in play materials and quality time.

We refer to the baseline period as t, when children were between 12 to 24 months old, and to the post-intervention period as $t + 1$, when children were between 30 to 42 months old. Child skills at $t+1$ are assumed to be a function of the vector of child skills at t, maternal skills at t, parental investments in the intervening period and random shocks. However, rather than modeling investment choices resulting from the dynamic optimization of a household problem as in DelBoca, Flinn, and Wiswall (2014), we estimate a pair of reduced form investment equations, which can be interpreted as an approximation to those derived (numerically) in a full structural model. By not imposing all the restrictions from a specific structural model we do not have to take a stance on whether parents know the process of child development reflected in the structure of the production function.

The model we use allows us to characterize the process of early child development and provides a framework to understand the mechanisms that generated the intervention's impacts. One mechanism through which the intervention may have operated is by changing the production function itself. On the one hand, the stimulation provided during the home visits may be a new input in the development of the child, and this

would be captured by a shift in Total Factor Productivity (TFP) parameter or other parameters of the production function. On the other hand, parents, now guided by the intervention, may use time and resources in a more effective way. This interpretation implies that, despite the richness of our data, some aspects of investment quality may not be captured by our measures and thus get embodied in the estimates of the production function parameters.[23]

A second mechanism through which the intervention could generate impacts on child development could be an increase in parental investments. Indeed, the intervention aims to strengthen child-mother interactions and encourage mothers to engage more with the child by incorporating age-appropriate play activities in the daily routine, introducing new toys constructed with home-made materials and spending time reading, telling stories or singing. However, it is also possible that investments could decline as parents shift their attention and resources elsewhere (for example, to other children) because they perceive the intervention itself as some form of investment either in itself (effectively an in-kind transfer). Such crowding-out of private resources is a standard concern in programs that target children.

Finally, the intervention could also have affected maternal cognitive or, more plausibly, socio-emotional skills. Many of the mothers (37%) were depressed at baseline according to the CESD scale, and it is plausible that the treatment mitigated this. Although we checked for such impacts, we did not detect any differences in our measures of maternal skills (either cognitive or socio-emotional skills) between the control and treatment groups after the intervention; thus this potential change is not a mechanism that contributed to the outcome. In our estimated model we only include baseline

---

[23]We made every effort to collect both time and resource use carefully targeted to the child with an emphasis on items that can drive development. For example, one of our measures is the number of times spent reading with the child in the last 3 days. Yet, it is still a possibility that as a result of the intervention, parents may be more able to select age-appropriate or stimulating stories to read with their child. Our measure of the frequency with which parents read with their child would not pick up this change in the quality of interaction, which would instead be picked up by a shift in the productivity of time investments.

maternal skills.[24]

5.1 The production function for human capital

We consider a production function of human capital that maps initial conditions, parental investment of different types and other factors on two different dimensions of child development. In particular, we assume that the stock of skills of child i in period $t+1$ is determined by the vector of child's baseline cognitive and socio-emotional skills θ_{it} embodying the initial conditions at the time of observation (possibly including any paternal influence), the mother's cognitive and socio-emotional skills denoted by P_{it}^C and P_{it}^S respectively, and the investments I_{it+1} made by the parents between t and $t+1$. We also allow for the effect of a variable η_{it+1}^k that reflects unobserved shocks. As with skills, parental investments I_{it+1} can be a multi-dimensional vector. We denote material investments by I_{it+1}^M and time investments by I_{it+1}^T.

Following our own earlier experimentation, we assume the production function for each of the two skills is Cobb-Douglas, so we can write the technology of formation for skill k as follows:[25]

$$
\begin{aligned}
ln(\theta_{idt+1}^k) =& A_d^k + \gamma_{1d}^k ln(\theta_{it}^C) + \gamma_{2d}^k ln(\theta_{it}^S) + \gamma_{3d}^k ln(P_{it}^C)^+ \gamma_{4d}^k ln(P_{it}^S) \\
& + \gamma_{5d}^k ln(I_{idt+1}^M) + \gamma_{6d}^k ln(I_{idt+1}^T) + \gamma_{7d}^k n_{it} + \eta_{it+1}^k \qquad k \in \{C, S\}
\end{aligned}
\tag{5}
$$

where C and S stand for cognitive and socio-emotional skills respectively. The term n_{it} is the number of children in the household and allows for the possibility that the presence of siblings affects child development because of spillover effects and more broadly because of the learning and socialization that can be achieved by interacting

---

[24]The effect of the intervention on the principal component factor of the CES-D scale items at follow-up is 0.13 of a standard deviation (with a p-value of .12), which, given the way the factor is defined, is indicative of an improvement but too insignificant to rely upon.

[25]Cunha et al. (2010) use a CES, while Cunha and Heckman (2008) use a log linear specification.

with other older children.[26] It is possible, on the other hand, that the presence of siblings dilute attention and resources. As we discuss below, such an effect could be captured by the investment functions. A_d^k is a factor-neutral productivity parameter or TFP and depends on the treatment status of the child (d) to capture the potential direct effect of the home-visitor stimulation during her weekly visit. Finally, all the parameters are specific to a particular skill.

5.2 Parental investments

We model investments as a function of the child and the mother's baseline skills and the number of children in the household.[27] The number of children in the household may dilute both resources and time devoted to the subject child. We also include a vector of variables Z_{it}, which determine investments but do not enter the production function. We discuss them below. The investment equations we estimate are

$$
\begin{aligned}
\ln(I_{idt+1}^\tau) =& \lambda_{0d}^\tau + \lambda_{1d}^\tau \ln(\theta_{it}^C) + \lambda_{2d}^\tau \ln(\theta_{it}^N) + \lambda_{3d}^\tau \ln(P_i^C) + \lambda_{4d}^\tau \ln(P_i^S) \\
& + \lambda_{5d}^\tau \ln(n_{it}) + \lambda_{6d}^\tau Z_{it} + u_{it+1}^\tau, \qquad\qquad \tau = \{M, T\}
\end{aligned}
\tag{6}
$$

As implied by the subscript d all coefficients could change with the treatment, a hypothesis we directly test. The effect of background variables on parental investment, given child initial conditions, is an important potential source of socio-economic gradients in child development. Moreover, the extent to which investments increase with child initial abilities is a reflection of parental beliefs about the heterogeneity of returns to such investments as well as parental taste for redistribution among children.

---

[26]Since our subject children are 12-24 months old at baseline these are almost always older children.
[27]The measure includes the subject child so the minimum is 1.

5.3 Estimation and mediation analysis

Parental investments are an input in the production function. However, they may be endogenous, i.e. it may be that $E(\eta_{it+1}^k|I_{it+1}^\tau) \neq 0$. In particular, parental investments might respond to unobserved, time-varying shocks in order to compensate or reinforce their effects on child development. Consider, for example, the case of a child who is suddenly affected by a negative shock, such as an illness, which is unobserved to the econometrician but perceived by the parents as delaying the child's development. As a result of this shock, parents might decide to invest in their child's development more than they would have otherwise. This parental response would create a negative correlation between parental investments and the unobserved error η_{it+1}^k, biasing downwards the impact of investments. Alternative assumptions about preferences and technologies (or technologies as perceived by the parents) can create different patterns of correlations between shocks and investment and, therefore, introduce different types of biases.

Standard mediation analysis, as in Heckman et al. (2013), takes all inputs as exogenous. It then considers various possible channels through which the intervention could affect outcomes and tries to establish which of them can explain the observed impacts on the outcomes. In our case, this approach would correspond to estimating the production functions by OLS, allowing the intervention to affect outputs (cognition and socio-emotional skills) directly, as well as indirectly through its impact on investments. One could then decompose the overall effect of the intervention into a direct effect, which could be interpreted as an improvement in productivity, and an indirect effect mediated by the increase in investment. Such an approach, however, can lead to misleading conclusions if investments are endogenous: if, for instance, ignoring endogeneity leads to under-estimate the impact of investment on outputs, then the channels through which the intervention can affect outcomes will be potentially misinterpreted.

To deal with the endogeneity of investment we use instrumental variables. We discuss our choice of instruments in the next subsection.

To estimate the model we proceed in three steps. First, as mentioned earlier, we estimate the covariance structure and the factor loadings of the latent factors based on the covariance structure of the observed measures. This step requires no distributional assumptions on the latent factors or the measurement error, but only relies on the restrictions embedded in the measurement system and discussed earlier. For estimation however, we assume that the latent factors are distributed as a mixture of two normal distributions and that measurement error is normally distributed. In the second step, we use the estimates of the measurement system to predict Bartlett factor scores for each individual in the data. In the final and third step, we use these predicted scores as observable data to estimate the parental investment and production function equations.

The third step requires correcting for the measurement error introduced from the fact that we use predicted values of the latent factors instead of the actual ones. To do so, we adapt the correction method described in Heckman et al. (2013) in the context of instrumental variables. Using the predicted factor scores and instruments, we estimate the investment equations and the reduced form of the production functions, where we obtained the latter by substituting material and time investments in the production function with their relevant first stage equations. We correct the investment and reduced form coefficients using the method described in Heckman et al. (2013) and finally recover the structural coefficients of the production function using a minimum distance estimator (Rothenberg, 1971).

We compute 95% confidence intervals and critical values for test statistics using the cluster bootstrap, where the entire estimation procedure is replicated 1000 times. We cluster at the municipality level which was the randomization unit in our experiment.

5.4 Choice of instruments

When investments are treated as endogenous, identification requires instruments that are relevant and can be excluded from the production function, under reasonable assumptions. A potential instrument is the intervention itself, which was randomized. However, the fact that we wish to test whether the intervention affected the production function directly as one of the possible channels through which it operated precludes us from using it as an identifying variable that is excluded from the production function. Moreover, because we have two endogenous variables (material and time investments), we need at least two instruments so the randomization alone would not be enough to identify the model anyway.

Consistent with a standard model of parental investment, we expect material investments to be related to the prices of relevant goods. Specifically, we use as instruments the average log price of toys and the average log price of food items in the municipality of residence. We assume that the variability of prices across communities is unrelated to factors affecting the development of cognition and socio-emotional skills of children.

To provide corroborative evidence of the validity of these exclusion restrictions, in Table 3 we present regressions of the log price of food and the log price of toys on various baseline characteristics, which, reasonably, should not be affected by these prices. The coefficients are jointly insignificant as shown by the p-values at the bottom of the Table. Moreover, the associated coefficients are very small and economically unimportant. Marriage is individually significant and while all other variables are included in the model marriage is not. In other versions of the model we also include the marriage indicator in the production function and none of the conclusions we draw are affected. Indeed, such a variable is completely insignificant in the production functions, further strengthening our conclusions.

Finding instruments for *quality time* is more challenging. This input may reflect

Table 3: Balance test for the instrumental variables

	Log Toy Price	Log Food Price	Conflict[†]
Mother's Cognition	0.020	0.007	-0.005
	(0.009)	(0.007)	(0.003)
Mother's non-cognitive skill	-0.006	-0.014	0.005
	(0.008)	(0.006)	(0.005)
Married	0.013	0.030	0.017
	(0.016)	(0.012)	(0.008)
Wealth Index	-0.003	0.002	0.002
	(0.007)	(0.006)	(0.004)
Terrorism[⋆]	0.014	-0.012	0.005
	(0.014)	(0.013)	(0.009)
Constant	8.036	8.053	0.039
	(0.036)	(0.027)	(0.015)
Observations	1,010	1,023	1,023
R-squared	0.011	0.021	0.017
F-statistic	1.631	2.147	1.438
F-pvalue	0.160	0.0671	0.218

Note: All right hand side variables measured at baseline. [⋆]Terrorist attacks: measured between conception of child and baseline. [†]"Conflict" is the number of conflicts against civil population/1000 population when the mother herself was a child. Asymptotic standard errors in parentheses allowing for clustering at the municipality level (randomization unit).

more the way parents spend time with their children and the type of activities they engage in than the amount of time spent. This intuition is confirmed by results in Table 4, which show that the intervention had no impact on maternal labor supply: the impact of the treatment on both employment and weekly hours is small and insignificant, although it had large effects on time investments. This evidence suggests that the margin of adjustment is not between work and quality time with children, but rather between quality time versus other household production activities or leisure, which are excluded from our measures of time investments. Consequently, measures of the opportunity cost of time, such as village-level female or male wages, are unlikely to have much explanatory power.[28] Below we also show that quality time investment is also unrelated to the relative prices we use as instruments for material investment.

---

[28]In earlier versions we demonstrate that indeed they do not.

Table 4: Impact of the intervention on maternal labor supply

	Employment	Hours of work
Treatment dummy	-0.0373	0.208
	(0.0852)	(1.793)
Demographic controls		
Mother's age	0.193	3.839
	(0.0439)	(0.802)
Mother's age squared	-0.003	-0.0542
	(0.0007)	(0.0130)
Years of education	0.056	0.921
	(0.0121)	(0.214)
No. of children = 2	-0.213	-4.429
	(0.0983)	(1.866)
No. of children = 3	-0.397	-5.905
	(0.124)	(2.382)
No. of children = 4	-0.230	-2.057
	(0.166)	(3.158)
No. of children = 5	-0.664	-13.71
	(0.236)	(3.612)
No. of children > 6	-0.346	-0.169
	(0.244)	(5.107)
Constant	-3.457	-48.39
	(0.659)	(11.91)
Observations	1,210	1,200

Note: Interviewer fixed effects included. Default number of children is one, which is the subject child. Asymptotic standard errors in parentheses allowing for clustering at the municipality level (randomization unit).

To instrument time investments, we instead focus on variables that have the potential to affect the willingness and ability of mothers to engage with their children. Specifically, we exploit the fact that Colombia has a long and well documented history of civil conflict that has affected large parts of the country and, in particular, rural areas. It is well documented that exposure to violence can cause emotional detachment, which can impede or make subsequent interaction with one's own children harder (Betancourt, 2015; Creech and Misca, 2017, for related evidence). This leads us to consider maternal exposure to past conflict as a potential instrument for current quality time investment, where we exploit variation in the intensity of conflict across

municipalities.[29]

As with our price instruments, our identifying assumption is that the variation in conflict across municipalities is orthogonal to unobservable factors affecting the development of the cohort of children we are studying. Although this assumption is not testable, in Table 3 we show that the incidence of conflict is not associated with mothers' baseline characteristics or with household wealth: all coefficients are very small and insignificant with a joint p-value of 0.22. Among these variables we also include the incidence of terrorist activities around the time the child was born: while the conflict itself was mainly over in the sampled communities by the time of our experiment, there were sporadic terrorist attacks during the period of the intervention. We included this variable to check that our measure of maternal childhood exposure to conflict is not related to current violence, which could have a direct impact on the child. The results in Table 3 strongly suggest that they are not. Thus, as is evident from these results, the past incidence of conflict is not associated with baseline characteristics relevant for child development. However, as we show below, mothers' childhood exposure to conflict is a strong determinant of the quality time that they spend with their child.

6   Results

We now report our empirical results. We start with estimates of the investment function before moving on to estimates of the production functions. The latter allow us to investigate what generates the impacts of the intervention on child development, a key question that impinges on the design of such programs. In what follows we report confidence intervals for all parameters. These have been computed using the block bootstrap accounting for the entire estimation procedure, and the cluster structure

---

[29]Specifically, our instrument is defined as the number of conflicts against the civil population divided by population (in thousands) in a given municipality when the mother was a child.

of our data. Whenever we present test statistics we compute the p-values using the bootstrap.

6.1 Estimates of the investment functions

The estimates of the investment equations are reported in Table 5. The first two columns of the table report estimates of the material and time investment equations, where we use toy price, food price and maternal childhood exposure to conflict as exclusion restrictions. The third column reports estimates of the material investment equation, where we only include prices as exclusion restrictions. This first stage will be used to estimate a production function for cognitive skills that does not include time investments as an input (and that will turn out to be our preferred specification).

The first striking result is the impact of treatment on investments: it increases materials by 20% and time by 33%, and both effects are highly significant. The results reported in Table 5 exclude interactions of the treatment parameter with the remaining variables, which were found to be insignificant.[30] These estimates of the impact of the intervention on investments are driven by the experimental design and do not require any of the assumptions necessary for the identification of the production functions.

The fact that the intervention increased the quality time and resources that parents provide to children is important because it shows that parents are willing to reinforce the intervention. While we already showed some evidence of crowding out for individual items, overall the opposite seems to be happening. From a policy perspective this is a major conclusion that should encourage further interventions in such contexts. As we shall see below, this increase in parental investment is the key source of success of the intervention.

---

[30]The estimates where all parameters of the investment functions are allowed to vary with treatment are shown in Web Appendix Table E.1. We test the joint significance of the interaction terms and find that we cannot reject that all the interactions are equal to 0 for both material and time investments: the p-value for the material investment equation is 0.369 and the p-value for the time investment equation is 0.099.

Table 5: Estimates of the material and time investment equations

	Instruments: Prices and Conflict		Instruments: Prices only
	Material investment	Time investment	Material investment
Intercept	-0.016	0.001	-0.016
	[-0.114,0.078]	[-0.089,0.089]	[-0.11,0.078]
Treatment dummy	0.204	0.333	0.203
	[0.038,0.365]	[0.155,0.48]	[0.037,0.362]
Log child's cognitive skill (t)	0.130	0.067	0.133
	[0.016,0.246]	[-0.044,0.18]	[0.017,0.25]
Log child's socio-emotional skill (t)	-0.032	0.021	-0.033
	[-0.133,0.087]	[-0.083,0.145]	[-0.131,0.088]
Log mother's cognitive skill (t)	0.754	0.367	0.755
	[0.582,0.939]	[0.162,0.498]	[0.583,0.943]
Log mother's socio-emotional skill (t)	0.071	0.024	0.071
	[-0.008,0.139]	[-0.06,0.108]	[-0.008,0.139]
Log number of children (t)	-0.129	-0.134	-0.128
	[-0.18,-0.077]	[-0.186,-0.072]	[-0.18,-0.078]
Log toy price	-0.095	-0.020	-0.094
	[-0.168,-0.027]	[-0.085,0.037]	[-0.163,-0.026]
Log food price	0.096	0.042	0.097
	[0.006,0.178]	[-0.026,0.121]	[0.006,0.178]
Maternal childhood exposure to conflict	-0.011	-0.096	
	[-0.08,0.063]	[-0.139,-0.032]	

Rank test[a]		0.011	
Cragg-Donald test		0.019	
Test of joint significance - F-statistic (p-values):			
Toy price, food price, conflict	10.48 (0.028)	13.42 (0.008)	
Toy price, food price, conflict, treatment	22.41 (0.001)	26.42 (0.001)	
Toy price, food price	10.48 (0.013)	1.53 (0.455)	10.57(0.010)
Toy price, food price, treatment	22.41 (0.001)	20.55 (0.000)	22.55(0.000)
Conflict, treatment	5.95 (0.047)	21.79 (0.000)	

Note: Dependent variables are the log of material and time investments. Maternal childhood exposure to conflict is defined as the number of conflicts against civil population/1000 when the mother was a child. t refers to baseline/pre-treatment measurement. [a] This is a test of the null hypothesis that the smallest eigenvalue of the 2×2 matrix $\beta'\beta$ is zero, where β is the 3×2 matrix of coefficients on food price, toy price and conflict in the material and time investment equations. We present this alongside the Cragg-Donald test because in this context it is not clear which is the more powerful. 95% Confidence intervals in square brackets. These as well as the p-values for the rank tests and all other tests are based on 1,000 bootstrap replications of the entire estimation process taking into account of clustering at the municipality level.

Turning now to the other regressors, we find that parents invest more resources in children with a higher baseline level of cognition (elasticity of 0.13) but the child's baseline socio-emotional skills have no impact on either type of investment. The elasticity of both material and time investments with respect to maternal cognition is very high and particularly so for the former; however mother's socio-emotional skills only affect material investments significantly and there the impact is very small. The number of other children in the household at baseline reduce significantly both time and material investments: an extra child reduces both investments by about 13%, which is consistent with a quantity/quality tradeoff among children. Moreover, the results are in line with a model where parents choose investments taking into account complementarity with child cognitive skills.

Importance of instruments. The next set of variables reported in Table 5 explain investments and act as excluded instruments when we estimate the production functions and treat investments as endogenous. These are the prices of toys and food and the level of conflict in the municipality when the mother was a child. The rank test we implement has a p-value of 0.011.[31] As an alternative, we also present the Cragg-Donald form of this test, which has a p-value of 0.019. These establish the strength of the instruments, allowing for the fact that there are two endogenous variables. As we will elaborate in Section 6.4, our Monte Carlo simulations show that these instruments are strong enough to avoid weak instrument bias.

The prices we consider are expected to affect material investment through the household budget constraint. Indeed, the p-value of a test of joint significance of these instruments is 0.013 in the material investment equation. As we would expect, toy price has a negative and significant impact on material investments. Food price, on the other hand, has a significant positive effect on material investment, implying

---

[31]See Robin and Smith (2000) and Blundell et al. (1998).

that play materials and food are substitutes.

Understanding the determinants of quality time is harder, as argued above. Prices have no explanatory power: their joint p-value in the quality time investment equation is 0.46. In results not reported here, we also found that wages did not predict time investments either, supporting our argument that the opportunity cost of quality time is not the time spent in the labor market and that spending quality time with children does not necessarily require monetary resources, but rather perhaps knowledge of child development and a certain willingness to engage with the child. Maternal childhood exposure to conflict, on the other hand, has a strong and significant negative impact on quality time activities (p-value 0), although it has no impact on material investments. The latter result reinforces the idea that this variable is not merely picking up some omitted background factor.

Jointly, these instruments are strong enough to allow us to control for the endogeneity of investments in the production function. In some specifications, we also exclude treatment status from the production functions, thus implicitly using the treatment as an instrument. In this case, the instruments only become stronger.

Finally, in the last column of Table 5 we present a specification for material investment that excludes the conflict variable; the coefficients are almost identical to those in column (1), and the prices are jointly even more significant. We use this investment equation to estimate a model where only material investments enter the production function for cognitive skills.

6.2 The production function for cognitive skills

In Table 6, we report estimates of a Cobb-Douglas production function for cognitive skills. The first column presents results where investments are assumed to be conditionally exogenous; in the remaining columns investments are taken as endogenous.

The production function demonstrates a high level of persistence for cognition;

Table 6: Estimates of the production function for cognitive skills

	(1) OLS	(2)	(3) IV	(4)	(5)
Instruments:		Prices, conflict	Price, conflict, treatment	Prices	Prices and treatment
Intercept	-0.016	-0.02	0.008	-0.006	-0.009
	[-0.094,0.053]	[-0.111,0.085]	[-0.088,0.104]	[-0.092,0.072]	[-0.083,0.057]
Treatment	0.084	0.076		-0.022	
	[-0.025,0.192]	[-0.12,0.391]		[-0.176,0.161]	
Log child's cognitive skill (t)	0.67	0.663	0.634	0.632	0.631
	[0.589,0.77]	[0.544,0.797]	[0.521,0.748]	[0.527,0.747]	[0.533,0.739]
Log child's socio-emotional skill (t)	-0.004	0.004	0.019	0.014	0.02
	[-0.091,0.087]	[-0.09,0.142]	[-0.095,0.142]	[-0.088,0.129]	[-0.081,0.129]
Log mother's cognitive skill (t)	0.217	0.018	-0.179	-0.099	-0.094
	[0.089,0.35]	[-0.435,0.477]	[-0.548,0.198]	[-0.503,0.291]	[-0.45,0.21]
Log mother's socio-emotional skill (t)	0.106	0.089	0.064	0.067	0.074
	[0.031,0.173]	[-0.031,0.162]	[-0.034,0.149]	[-0.021,0.157]	[-0.025,0.154]
Log of number of children (t)	0.041	0.075	0.084	0.09	0.086
	[-0.01,0.092]	[-0.068,0.15]	[0.011,0.163]	[0.004,0.176]	[0.023,0.163]
Log of material investment (t+1)	0.08	0.463	0.799	0.531	0.516
	[0.016,0.157]	[0.033,1.182]	[0.197,1.4]	[0.038,1.015]	[0.195,0.949]
Log of time investment (t+1)	0.04	-0.131	-0.323		
	[-0.051,0.129]	[-1.178,0.3]	[-1.022,0.227]		
Gap in output between treated and control:					
(a) Measured in the data			0.115		
(b) Predicted by the model estimates	0.114	0.127	0.055	0.086	0.105

Note: Dependent variable is the log cognition of the child at follow up. At the top of each of the IV columns we state the exclusion restrictions used. t and $t+1$ denote that a variable was measured at baseline and endline, respectively. 95% Confidence intervals in brackets based on 1,000 bootstrap replications of the entire estimation process taking into account clustering at the municipality level.

however, socio-emotional skills do not affect cognition at this early age, and this result remains unchanged whether we treat investments as endogenous or not. Mothers' cognition and socio-emotional skills have a strong positive effect on cognitive development when we use OLS, but in the remaining columns, when we use IV for the investments, these effects disappear, implying that mothers' skills operate through the initial conditions of the child and through the investment decisions only. Finally, the number of other children in the family improve child cognitive development. This is particularly interesting because the investment equations show that additional children

reduce both material and quality time investments. However the presence of other (for the most part older) children has a direct impact on child development, with an extra child improving cognition by 4%-9% depending on the specification.

We now turn to the estimates of the direct treatment effect and the coefficients on parental investments, as measured by our 'material investment' and our 'quality time investment' factors. When we treat investments as exogenous and use OLS, we find that the impact of material investments on cognitive development is significantly different from zero, but that of time is insignificant. The direct treatment effect is large but very imprecisely estimated, to the extent that it is insignificant. The estimate of the impact of material investments increases dramatically when we treat investment as endogenous. Going from column 1 to column 2, this coefficient increases from 0.08 to 0.46; and although it is now estimated less precisely it remains significant at the 5% level. The change in the size of the coefficient points to parents compensating for negative shocks affecting their children.

The coefficient on time investment, instead, stays insignificant and its point estimate actually turns negative. The estimate of the direct effect of the intervention stays insignificant, although, as we discussed above it is strongly significant in the investment equations. For this reason, in column 3, we force the direct treatment effect to be zero. Despite the now unequivocal strength of the instruments (since we are also using the treatment allocation as an exclusion restriction in column 3), the results do not change: material investments enter with a large and significant coefficient (0.56) and time investments remain completely insignificant.[32]

---

[32]In an earlier version of the paper we also used an indicator for whether the mother is married at baseline as an instrument. Following referees' comments we no longer use it. It is worth mentioning though that while being married had a significant effect in the investment equations it was not at all significant when also included in the production functions. One interpretation is that married couples invest more in their children and this is the channel through which outcomes are affected in our context.

Using material investments alone   Given the measurement issues for quality time, which is based on self-reports and discussed further below,[33] and in light of the results presented thus far, in the next two columns we exclude time investments from the production function and we switch to using *only* prices as instruments so the relevant first stage investment equation is in the third column of Table 5. In one specification we include the treatment dummy reflecting the direct effect of the intervention (column 4) and in the other we exclude it (column 5). The results are essentially the same, though with improved precision, and present a clear message: material investments have a strong and positive impact, while the direct effect of treatment is small and completely insignificant.

Interpreting the impact of the intervention   Based on these estimates, we now consider how the intervention affects outcomes through the lens of the production function. The possible channels include changes in the production function - a direct effect - and changes in parental investment inputs. Changes in the production function could happen for a number of reasons. First, the weekly session of the home visitor with the child, as well as any materials left behind, can be thought of as a new input; second the intervention could lead to a better use of measured inputs by parents or equivalently an improvement in the unmeasured quality of these inputs. These are possible channels through which the intervention could affect outcomes over and above inducing more investments through its emphasis on parenting and the direct involvement of the mother in the home visit.

From the coefficients reported in column 1 of Table 6, which treats investment as exogenous, together with the increase in investment documented in Table 5, we see that increased investment accounts for about 25% of the impact of the intervention. The point estimate of the coefficient on treatment, instead, indicates that the intervention

---

[33]See also section 4.2.

directly increased cognition considerably, by about 8.4%, although it is not precisely estimated. Therefore, according to this specification, parental investments play some role in mediating the intervention but there is a large direct effect, at least in terms of the point estimate.

A key problem with these results, though, is that they assume investments are invariant to unobserved shocks to child development that occur between baseline and the end of the intervention 18 months later. The IV results presented in columns 2 to 5 of Table 6 address this issue and present a different story: the increase in material investments now explains a large fraction of the observed impact. Thus, a good description of cognitive development among children aged 3-4 is that it is driven by initial cognition (measured at aged 1-2), stimulation provided by older siblings, and material investments provided by parents, and the main channel through which the intervention affects cognitive development is by shifting these investments. The model in column 5 with material investments alone achieves a good fit of the data and captures over 91% of the overall impact of the intervention. Finally, we find no evidence that any other coefficient of the production function changed. In Table E.2 we allow all coefficients to vary by treatment status. The overall p-value is 0.32, while for the non-intercept coefficients it is 0.29.

These results highlight the importance of material investments but leave no role for quality time investments. This is perhaps surprising because one would expect quality time to be an important input as well. However, this result could at least partly reflect the measurement issue we discussed earlier (see Section 4.2) and that needs to be addressed in future research: material investments are measured through interviewer observations, while time investments are self-reported and hence carry the risk to be over-reported, particularly in the intervention communities where the importance of quality time has been continuously emphasized throughout the duration of the intervention. Material investments provided by the parents require real effort through

making or buying toys, and they are likely to correlate strongly with actual activities carried out with the child. For example, providing a toy, whether it is home made or bought, will likely include spending time with the child in this activity. The actual observation of play materials may therefore better reveal the developmental activities carried out by parents. On the other hand the materials left behind by the intervention and captured by the intervention/treatment dummy may not reflect actual engagement by the parent. This again emphasizes that the program works to the extent that it shifts actual parental behavior.[34]

6.3 The production function for socio-emotional skills

In Table 7, we present estimates of several specifications of the production function for socio-emotional skills. As with cognitive skills, we observe that the accumulation process of socio-emotional skills exhibits substantial amount of persistence (regardless of the specification considered). The point estimates, however, are lower than in the case of cognition: about 0.50, compared to about 0.70 for cognitive skills. The intervention also has no direct effect in any specification. The lagged value of cognitive development is marginally significant in the various specifications, indicating a role for cognitive development in generating socio-emotional skills.

Parental background variables, such as mothers' cognitive and socio-emotional skills, do not seem to matter for the development of children's socio-emotional skills. Instead, the number of siblings has a positive and significant impact on these skills, consistent with what we found for cognitive development. This result is robust across specifications.

As with cognitive skills, we experiment with a number of specifications. In none of

---

[34]It may be interesting to consider whether similar impacts could be achieved by subsidizing prices of toys. However, the treatment effect on investment is clearly very large compared to the price effects. In a sense this reflects the fact that the intervention induce parents to change their behavior towards their children in quite a radical way that cannot be reasonably achieved by shifting prices.

Table 7: Estimates of the production function for socio-emotional skills

	(1) OLS	(2)	(3)	(4)	(5)
			IV		
Instruments:		Prices, conflict	Prices, conflict	Prices, conflict, treatment	Prices, conflict, treatment
Intercept	-0.008	-0.007	0.000	-0.026	-0.022
	[-0.08,0.063]	[-0.09,0.065]	[-0.083,0.066]	[-0.095,0.058]	[-0.092,0.051]
Treatment	-0.012	-0.083	-0.102		
	[-0.124,0.093]	[-0.318,0.158]	[-0.308,0.123]		
Log child's cognitive skill (t)	0.105	0.086	0.087	0.104	0.098
	[0.018,0.192]	[-0.018,0.211]	[-0.014,0.188]	[0.005,0.22]	[0.004,0.191]
Log child's socio-emotional skill (t)	0.51	0.494	0.5	0.512	0.513
	[0.403,0.659]	[0.38,0.669]	[0.388,0.662]	[0.388,0.665]	[0.404,0.656]
Log mother's cognitive skill (t)	-0.073	-0.091	-0.143	-0.035	-0.083
	[-0.217,0.049]	[-0.438,0.305]	[-0.361,0.085]	[-0.293,0.329]	[-0.23,0.075]
Log mother's socio-emotional skill (t)	0.036	0.041	0.06	0.046	0.044
	[-0.058,0.119]	[-0.055,0.144]	[-0.054,0.135]	[-0.047,0.134]	[-0.041,0.126]
Log of number of children (t)	0.098	0.128	0.131	0.103	0.101
	[0.047,0.153]	[0.015,0.235]	[0.03,0.223]	[0.026,0.17]	[0.036,0.164]
Log of material investment (t+1)	0.142	-0.079		-0.117	
	[0.06,0.256]	[-0.645,0.424]		[-0.698,0.341]	
Log of time investment (t+1)	0.119	0.558	0.542	0.436	0.324
	[-0.006,0.213]	[-0.175,1.285]	[-0.054,1.15]	[-0.096,1.022]	[0.023,0.687]
Gap in output between treated and control:					
(a) Measured in the data			0.087		
(b) Predicted by the model estimates	0.057	0.087	0.078	0.121	0.108

Note: Dependent variable is the log socio-emotional skills of the child at follow up. At the top of each of the IV columns we state the exclusion restrictions used. t and $t+1$ denote that a variable was measured at baseline and endline, respectively. 95% Confidence intervals in brackets based on 1,000 bootstrap replications of the entire estimation process - clustering at the municipality level.

the specifications does the direct treatment effect play any role: the coefficient is always negative and very imprecise. When we use OLS we find that both material and quality time investments enhance socio-emotional development. When we turn to IV (columns 2 to 5), none of the investment coefficients remain significant, with the exception of the results in column 5. There, we exclude material investments, which always enter with a negative coefficient when instrumented, and the treatment dummy, which never plays a role. In this case the results suggest that our measure of quality time can explain the improvement in socio-emotional skills resulting from the intervention.

Based on the OLS and using a specification that excludes the direct impact of the intervention (which is negative and insignificant) all the reported coefficients in column

1 remain unchanged. That specification explains 79% of the impact (it predicts a 0.069 log point improvement). The alternative specification in column 5 over predicts the impact. We also tested whether the production function differed beyond an intercept shift between treatment and control: the differences were completely insignificant, even for the more precise OLS specification with a p-value of 0.38. Whether we take the OLS results or the IV results presented in column 5, their implications are clear and similar: the intervention acted by improving parental engagement with their children but there is no evidence that it also had a direct impact on their development.[35]

6.4   Are the instruments weak? Monte Carlo evidence

Before discussing further the implications of the estimates we first report results demonstrating that the instruments we use are strong enough to avoid biases due to a weak instrument problem. Weak instruments can lead to estimates that are severely biased towards OLS.[36] Although the F-statistics and the rank tests reported in Table 5 suggest that the instruments are sufficiently strong, these criteria have not been derived and validated for our nonstandard framework, which includes a first stage with factor analysis, two endogenous variables on the right hand side and a cluster structure with intracluster spatial correlation. Instead of relying on unvalidated diagnostics, we investigate directly whether we have a weak instrument problem by performing a Monte Carlo simulation.

Using our parameter estimates as true values[37] we simulate investments and cog-

---

[35]The experiment was stratified across three regions: Central, Oriental and Cafetera. The probability of allocation to treatment was equal in each stratum and hence we do not need to control for stratum fixed effects in the experimental analysis. However, they may be relevant in the estimation of the investment equations and the production functions. Including the stratum fixed effects, as expected, weakens a bit the instruments, whose variation is geographic, although prices and conflict are still significant. However, the overall conclusions do not change. We present these additional results in Appendix Tables E.12 and E.13.

[36]see Bound et al. (1997), Staiger and Stock (1997), Andrews and Stock (2005) amongst others.

[37]We interpret the coefficient of time investments in the cognitive production function as 0. In other non-reported simulations, we also allowed for a positive coefficient on time investment and none of the conclusions were affected.

nition. The simulations use the actual data on the exogenous variables[38], which are kept fixed in repeated samples. We draw errors based on a covariance matrix that replicates the stochastic structure in the data, including the cluster structure. We use two alternative covariance matrices (Cov A and Cov B), the latter being particularly conservative (Cov B).[39] By setting the coefficients of the first stage in the data generating process to be the same as the ones we estimate in the investment equations and using the appropriate covariance structure for the errors we replicate data which leads to a first stage effectively equal in strength as the actual one in the data. All details are provided in Appendix section E.3.

Table 8: Monte Carlo Simulations

| | True Values | | Covariance A | | Covariance B | |
| | Material | Time | Material and time investment functions on simulated data | | | |
			Material	Time	Material	Time
Toy price	-0.095	-0.02	-0.089	-0.019	-0.082	-0.017
			[-0.143,-0.036]	[-0.077,0.037]	[-0.154,-0.013]	[-0.101,0.073]
Food price	0.096	0.042	0.089	0.04	0.081	0.037
			[0.024,0.154]	[-0.027,0.103]	[0,0.163]	[-0.061,0.142]
Conflict	-0.011	-0.096	-0.01	-0.093	-0.007	-0.085
			[-0.064,0.048]	[-0.152,-0.037]	[-0.076,0.07]	[-0.172,0.001]
Rank test			0.047		0.029	
Cragg-Donald test			0.059		0.023	
	True Values		Cognitive production function on simulated data			
			Investment treated as:		Investment treated as:	
			Exogenous	Endogenous	Exogenous	Endogenous
Material investment	0.463		0.111	0.418	0.131	0.432
			[0.045,0.176]	[-0.084,1.059]	[0.066,0.197]	[-0.132,1.103]
Time investment	0		0.162	-0.003	0.185	0.017
			[0.097,0.23]	[-0.655,0.516]	[0.118,0.253]	[-0.794,0.642]

Note: The Monte Carlo simulations employ 1000 replications. Average estimates across replications with bottom and top 2.5th percentile in square brackets. We produce a simulated observation for each household in the data using the actual values for the observed exogenous variables, including the instruments. So the Simulated data has exactly the same structure as the actual data. The data generating process includes the transformation from latent factors to measurements using the estimated measurement system. Each replication implements the entire estimation procedure as described in the full model, including the factor scoring. Investments are endogenous in the simulated data: the error structure assumed in the simulation allows for a covariance between the investments and cognition residuals and for spatial dependence. The errors are normal.

We compare the assumed true values to the average estimates obtained in 1000

---

[38]Prices, Conflict, Treatment status, Parental background and Child baseline characteristics.
[39]Both covariance matrices are reported in Appendix Table E.9.

simulated data sets. The difference is the bias when our estimator is applied to data of our structure and size. The results for some key parameters are shown in Table 8. The top panel compares the average estimates of the coefficients on the instruments for the first stage to the assumed true values used in generating the data. As expected they are estimated with no substantial bias, the average being very close to the true values.

The lower panel of Table 8 shows the coefficients on the two investments in the cognitive production function. Under the column "exogenous investments" we erroneously ignore the endogeneity of investments in which case we obtain a strong downward bias for the coefficient on material investments (0.111 and 0.131 under covariance A and B respectively instead of 0.463) and a strong upward bias for the coefficient on time investments as we would expect for the assumed structure of the covariance matrix of the errors. However, when we allow for endogeneity the bias almost completely disappears. For example, under the more conservative Covariance B we get 0.432 when the true value is 0.463 for material investments; for time investments we obtain 0.017 where the true value is zero. These results are strong evidence that our estimates do not suffer from weak instrument bias.[40]

6.5   Complementarities between inputs of the production function

We now go back to considering the implications of our estimates. The Cobb-Douglas specification implied by the data means that the inputs are complementary (Cunha et al., 2006; Heckman and Mosso, 2014). The complementarity is imposed by the functional form. However, in earlier versions of the paper, when we estimated the substitution elasticity, we always found it very close to 1. Given the metric we use for the latent factors, the return to investment is higher for children with better initial conditions. In Figure 2(a) we show the effect of a one standard deviation increase in

---

[40]The full set of Monte Carlo results are presented in Appendix Table E.11.

material investment on cognitive skills and in Figure 2(b) the effect of a one standard deviation increase in both material and time investments on socio-emotional skills. The y-axes are in standard deviation units of the outcome.

Figure 2: Complementarity between investments and baseline skills

(a) Production function for cognitive skills   (b) Production function for socio-emotional skills
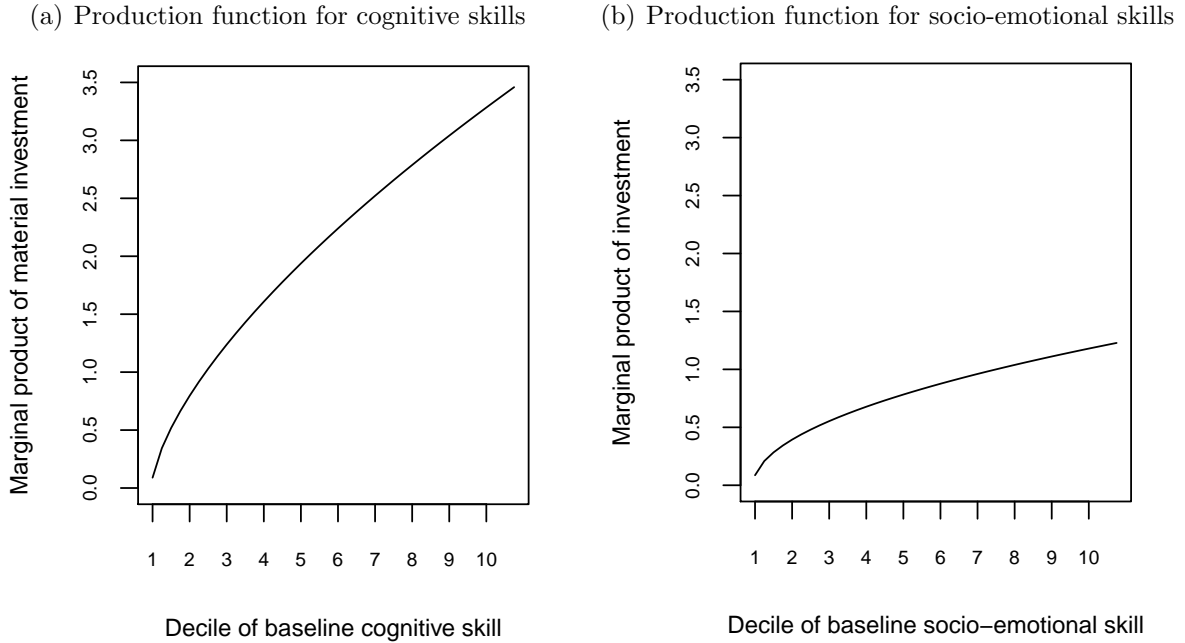
Notes: Figure 2.a (2.b) is based on the estimates of the production function for cognitive skills (socio-emotional skills) reported in Column 5 (1) of Tables 6 and 7. The figures above are constructed by evaluating the increase in cognitive (socio-emotional skills) in standard deviation units resulting from an increase in one standard deviation of investments at different deciles of $\theta_{i,t}^C$ for (a) and $\theta_{i,t}^S$ for (b) and holding all remaining inputs of the production function at their mean values across the sample.

The complementarity of investments with initial conditions of the child may appear contradictory to the set of studies indicating that early interventions benefit low-achieving children the most (Bitler et al., 2014; Elango et al., 2016). However, one needs to allow for the differences in the populations concerned. Our intervention targets the 20% poorest children in Colombia. While these children do not live in extreme poverty, they may still be poorer and of lower ability at baseline than disadvantaged children targeted by programs such as Head Start in the US. Our results imply that, in this subset of the population, those with a better start benefited more. However, one

can imagine that with a population that extends more broadly in the socio-economic distribution, diminishing returns could set in unless perhaps we design an intervention better attuned to higher ability children.

6.6 Implications for longer term outcomes

The results have implications for what to expect in the longer term. Under the assumption that the patterns of self-productivity and complementarities documented here remain the same at least in the medium-term, the fact that the return to investments are complementary to the prior level of achievement implies that, if parents keep investing at the higher levels induced by the intervention, the skills of the treated children should continue to improve, subject of course to the impact of investments at later ages. And since ability and investments are complementary, future investments would further increase the skills of the intervention group. However, the estimates also reveal fadeout between the baseline and the follow up: the coefficient on past cognitive (socio-emotional) skills is about 0.7 (0.5) in the production function for cognitive (socio-emotional) skill for a time lapse of about 18months. If such mean reversion continues beyond the ages that we consider and if parents revert to the level of activity in the control group, we can expect the impact of the intervention to become much smaller in the long-run. In fact, as shown in Andrew et al. (2018), parents in our intervention reduced their level of engagement to that of the control group when interviewed again two years after the end of the intervention. And consistent with the results here the impact of the intervention also faded.

Partial fadeout was also observed following the Jamaica intervention (Walker et al., 2005), although perhaps because the original effect was as large as 80% of a standard deviation, half the original impact remained. This underscores two key lessons. First, we should not underestimate the challenges involved in scaling up successful small scale efficacy trials, such as the Jamaica study, and in achieving comparable initial impacts.

Second, we need to better understand what motivates parents to continue investing in children in the longer run and whether sustained intervention is needed to preserve and reinforce initial gains in such environments.

7 Conclusion

Children from poor backgrounds accumulate developmental deficits from a very early age. Causes include not only the risky environments in which they live but also the lack of stimulation, which prevents the brain from developing its full potential. Such adverse early experiences are at the heart of the intergenerational transmission of poverty.

In this paper, we present results from an 18-month long early childhood intervention carried out in Colombia that promoted suitable parenting and stimulation to children between one and two years old at baseline. The intervention involved weekly home visits delivered by local women who had no prior specific knowledge of child development, but were trained as part of the intervention to deliver a structured stimulation curriculum that progressed in difficulty. The evaluation by randomized controlled trial showed improvements in a number of developmental dimensions, including cognition, language and socio-emotional development, though impacts on the latter are smaller.

We use data from the experiment to estimate a model of parental investments in children and production functions for children's cognition and socio-emotional skills. The aim is to improve our understanding of the development of child skills from a very early age and to provide an interpretation of how the intervention affected child development. The model estimates trace some of the origins of social inequalities to the beginning of life: children with higher initial skills obtain more investments from their parents, and, given the child's skills, mothers with higher levels of cognition invest more in their children.

The estimates of the production functions also provide evidence of several important

features of skill development among children below the age of 4. First, we find strong evidence of self-productivity of skills. That is, the current stock of cognitive (socio-emotional) skills strongly affects the development of future cognitive (socio-emotional) skills, but also imply mean reversion. Second, we find evidence of cross-productivity: the current stock of cognitive skills fosters the development of future socio-emotional skills, but the reverse does not seem to be the case. This result contrasts with that reported by Cunha, Heckman, and Schennach (2010), who find socio-emotional skills to be important for the accumulation of future cognitive skills though at a different age. While the presence of siblings in the household reduces parental investments, it improves the cognitive and socio-emotional development of our subject child (who is the youngest child in the family in most cases), most likely through interactions and imitation, over and beyond the effect of parental investments.

Most importantly for the question addressed in this paper, as well as more broadly, our results show that investments help develop both cognitive and socio-emotional skills. The program increased investments substantially and our key conclusion is that it is this increase that led to the estimated impacts of the intervention on children's skills. Specifically, the intervention increased the cognitive development of the children by 0.115 log points and socio-emotional development by 0.087 log points. Our best estimates of the production functions imply that the increase in parental investments induced by the program account for just over 91% of the intervention impact on cognition and at least 66% of its impact on socio-emotional skills. There was no direct impact of the intervention. It is thus the parenting component – where the home visitor directly involves the mother in developing the stimulation activities and encourages to continuously engage with the child based on the stimulating activities – that underlies its success.

Our study answers some important questions but raises many more, calling for further experimentation and analysis. We need to understand how to better target

and treat the most disadvantaged of society. Moreover, the analysis raises the question of how sustainable the effects of the intervention are and how salient improvements at this age are for longer term outcomes. This requires long-term follow ups of the children participating in the intervention and calls for further research with systematic measurements and interventions at various stages of life. Finally, an important lesson from our results is that involving parents in interventions is key to promoting child development in the short-term. Going forward, it is crucial that we better understand how to ensure continued parental investments after the intervention has ended.

References

Agostinelli, F. and Wiswall, M. (2016). Identification of dynamic latent factor models: The implications of re-normalization in a model of child development. NBER Working Paper No. 22441.

Almlund, M., Duckworth, A., Heckman, J. J., and Kautz, T. (2011). Personality, psychology and economics. In E. Hanushek, S. Machin, and L. Woessmann, eds., *Handbook of the Economics of Education*. Elsevier Science.

Almond, D. and Currie, J. (2011). Human capital development before age five. In O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics, Volume 4b*. Elsevier Science, Amsterdam.

Anderson, T. and Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 5, 111–150. University of California Press, Berkeley.

Andrew, A., Attanasio, O., Fitzsimons, E., Grantham-McGregor, S., Meghir, C., and Rubio-Codina, M. (2018). Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in colombia. *PLoS Med* 15(4): e1002556. Https://doi. org/10.1371/journal.pmed.1002556.

Andrews, D. W. K. and Stock, J. H. (2005). Inference with weak instruments.

Attanasio, O., Fernandez, C., Fitzsimons, E., Grantham-McGregor, S., Meghir, C., and Rubio-Codina, M. (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., Fitzsimons, E., Gomez, A., Gutiez, M.-I., Meghir, C., and Mesnard, A. (2010). Children's schooling and work in the presence of a conditional cash transfer program in rural colombia. *Economic Development and Cultural Change* 181–210.

Attanasio, O., Meghir, C., and Nix, E. (2017). Investments in children and the development of cognition and health in india. NBER Working Paper No. 21740.

Attanasio, O., Meghir, C., and Santiago, A. (2012). Education choices in Mexico: Using a structural model and a randomized experiment to evaluate Progresa. *Review of Economic Studies* 79(1):37–66.

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

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

Betancourt, T. S. (2015). Intergenerational effects of war. *JAMA Psychiatry* 72(3):199–200.

Bitler, M., Hoynes, H., and Domina, T. (2014). Experimental evidence on distributional impacts of head start. NBER Working Paper 20434.

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

Blundell, R., Duncan, A., and Meghir, C. (1998). Estimating labor supply responses using tax reforms. *Econometrica* 66(4):827–761.

Bound, J., Jaeger, D. A., and Baker, R. (1997). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90:443–450.

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

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

Campbell, F. A. and Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development* 65:684 – 698. Doi:10.2307/1131410 Medline.

Carneiro, P., Hansen, K., and Heckman, J. (2003). Estimating distribution of counterfactuals with an application to the returns to schooling and measurement of the effect of uncertainty on schooling choice. *International Economic Review* 44(2):361–422.

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

Creech, S. K. and Misca, G. (2017). Parenting with ptsd: A review of research on the influence of ptsd on parent-child functioning in military and veteran families. *Frontiers in Psychology* 8. Doi:10.3389/fpsyg.2017.01101.

Cunha, F. and Heckman, J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources* 43(4):738–782.

Cunha, F., Heckman, J., Lochner, L., and Masterov, D. (2006). Interpreting the evidence on life cycle skill formation. In E. Hanushek, S. Machin, and L. Woessmann, eds., *Handbook of the Economics of Education, Volume 1.* Elsevier Science, Amsterdam.

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

DelBoca, D., Flinn, C., and Wiswall, M. (2014). Household choices and child development. *Review of Economic Studies* 81:137–185.

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

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

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

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

Frongillo, E., Sywulka, S., and Kariger, P. (2003). UNICEF psychosocial care indicators project. Final report to UNICEF. Mimeo, Cornell University.

Gelber, A. and Isen, A. (2010). Children's schooling and parents' behavior: Evidence from the head start impact study. *Journal of Public Economics* 101:25 ? 38.

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

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

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

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

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

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

Heckman, J., Moon, S., Pinto, R., Savelyev, P., and Yavitz, A. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Journal of Quantitative Economics* 1:1–46.

Heckman, J. and Mosso, S. (2014). The economics of human development and social mobility. *Annual Reviews of Economics* 6:689–733.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103:2052–2086.

Helmers, C. and Patnam, M. (2011). The formation and evolution of childhood skill acquisition: Evidence from India. *Journal of Development Economics* 95(2):252–266.

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

Hu, Y. and Schennach, S. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica* .

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

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

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

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

Knudsen, E. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience* 16(1):1412–1425.

Knudsen, E., Heckman, J., Cameron, J., and Shonkoff, J. (2006). Economic, neurobiological and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Science* 103(27):10155–10162.

Lancet (2016). Early child development in developing countries 2016 .

Lu, C., Black, M. M., and Richter, L. M. (2016). Risk of poor development in young children in low-income and middle-income countries: an estimation and analysis at the global, regional, and country level. *Lancet Global Health 4* e916?e922.

Nielsen, E. R. (2015). Achievement gap estimates and deviations from cardinal com- parability. Finance and Economics Discussion Series 2015-040. Washington: Board of Governors of the Federal Reserve System, http://dx.doi.org/10.17016/FEDS.2015.040.

Putnam, S. P., Gartstein, M., and Rothbart, M. K. (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., Jacobs, J., Garstein, M., and Rothbart, M. (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.* Harcourt Assessment, San Antonio, TX.

Robin, J. and Smith, R. (2000). Tests of rank. *Econometric Theory* 16:151–175.

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

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica* 39(3):577–591.

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

Schennach, S. (2004). Estimation of nonlinear models with measurement error. *Econometrica* 72:33–75.

Schennach, S. (2007). Instrumental variable estimation of nonlinear errors-in-variables models. *Econometrica* 75:201–239.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica* 65:557–586.

Todd, P. and Wolpin, K. (2006). Using experimental data to validate a dynamic behavioral model of child schooling and fertility: Assessing the impact of a school subsidy program in Mexico. *American Economic Review* 96(5):1384–1417.

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

Walker, S., Chang, S., C.Powell, E.Simonoff, and Grantham-McGregor, S. (2006). Effects of psychosocial stimulation and dietary supplementation in early childhood on psychosocial functioning in late adolescence: follow-up of randomized controlled trial. *British Medical Journal* 333(7566):472.

Walker, S. P., Chang, S. M., Powell, C. A., and Grantham-McGregor, S. M. (2005). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study. *The Lancet* 366:1804 – 1807.

Walker, S. P., Wachs, T. D., Grantham-McGregor, S. M., Black, M., Nelson, C., and Huffman, S. (2011). Inequality in early childhood: risk and protective factors for early child development. *Lancet* 378(1325-38).

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*

May 12, 2019

WEB APPENDIX

*Attanasio: University College London and Institute for Fiscal Studies (o.attanasio@ucl.ac.uk). Cattan: Institute for Fiscal Studies (sarah_c@ifs.org.uk). Fitzsimons: UCL Institute of Education and Institute for Fiscal Studies (e.fitzsimons@ucl.ac.uk). Meghir: Yale University, NBER, IZA and Institute for Fiscal Studies (c.meghir@yale.edu). Rubio-Codina: Inter-American Development Bank and Institute for Fiscal Studies (martarubio@iadb.org).

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 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 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 chil-

dren 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 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 nutritional supplementation) and in the group who received none (including those who received the nutritional 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.

5

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	*pval*	*RW pval*
Children:							
Boy	0.510	(0.500)	0.481	(0.500)	0.029	*0.268*	*0.992*
Age (months)	18.02		18.04		-0.019	*0.933*	*0.992*
Premature	0.098		0.074		0.024	*0.287*	*0.992*
Mean birth weight (g)	3236.1	(526.4)	3246.0	(499.8)	-9.924	*0.762*	*0.992*
Stunted	0.141		0.134		0.007	*0.729*	*0.992*
First born	0.420		0.367		0.053	*0.078*	*0.888*
Bayley: cognitive	-0.005	(1.003)	-0.034	(1.049)	0.029	*0.788*	*0.908*
Bayley: receptive language	-0.009	(0.999)	-0.006	(0.934)	-0.003	*0.976*	*0.992*
Bayley: expressive language	-0.020	(0.992)	0.004	(1.018)	-0.024	*0.812*	*0.992*
Bayley: fine motor	-0.006	(1.003)	-0.066	(1.094)	0.060	*0.639*	*0.992*
Mac Arthur: words the child can say	-0.018	(0.933)	0.003	(0.974)	-0.021	*0.730*	*0.992*
ICQ: difficult (-)	-0.002	(1.001)	0.062	(1.000)	-0.064	*0.359*	*0.992*
ICQ: unsociable (-)	0.001	(0.997)	-0.035	(1.002)	0.037	*0.534*	*0.992*
ICQ: unadaptable (-)	0.001	(0.998)	-0.035	(1.036)	0.037	*0.685*	*0.992*
ICQ: unstoppable (-)	0.006	(0.996)	0.092	(1.026)	-0.087	*0.335*	*0.992*
Mother:							
Education (years)	7.575	(3.534)	7.351	(3.423)	0.224	*0.343*	*0.992*
Married	0.690		0.673		0.017	*0.512*	*0.992*
CESD: Bothered by things? (-)	0.004	(0.988)	0.092	(0.943)	-0.089	*0.020*	*0.490*
CESD: Trouble keeping mind on doing? (-)	-0.008	(0.999)	-0.070	(1.076)	0.062	*0.018*	*0.438*
CESD: Feel depressed? (-)	0.005	(0.993)	0.108	(0.976)	-0.103	*0.001*	*0.056*
CESD: Everything was an effort? (-)	0.001	(1.001)	0.120	(0.908)	-0.119	*0.002*	*0.092*
CESD: Feel fearful? (-)	0.005	(0.997)	-0.077	(1.058)	0.082	*0.097*	*0.936*
CESD: Was sleep restless? (-)	-0.003	(1.004)	0.070	(0.974)	-0.073	*0.012*	*0.343*
CESD: Feel happy?	-0.005	(1.005)	0.030	(0.996)	-0.035	*0.015*	*0.382*
CESD: Feel lonely? (-)	0.004	(0.999)	0.037	(0.974)	-0.033	*0.006*	*0.223*
CESD: Feel you couldn't get going? (-)	0.005	(0.999)	0.134	(1.037)	-0.130	*0.038*	*0.673*
Household:							
No of rooms/household size	0.594	(0.281)	0.591	(0.310)	0.003	*0.875*	*0.992*
Wealth index	-0.010	(0.993)	0.016	(1.006)	-0.027	*0.708*	*0.992*
Home ownership	0.370		0.364		0.006	*0.829*	*0.992*
FCI: Types of play materials	3.708	(1.688)	3.644	(1.697)	0.064	*0.600*	*0.992*
FCI: Types of play activities over past 3 days	3.256	(1.609)	3.223	(1.513)	0.033	*0.777*	*0.992*
FCI: No of adult books	1.071	(1.044)	1.047	(1.059)	0.023	*0.775*	*0.992*
FCI: No of magazines and newspapers	0.984	(1.140)	0.960	(1.160)	0.025	*0.772*	*0.992*
Log village-level toy prices	8.072	(0.189)	8.000	(0.221)	0.074	*0.081*	*0.884*
Log village-level food prices	8.075	(0.147)	8.101	(0.135)	-0.026	*0.372*	*0.992*
Maternal childhood exposure to conflict	.056	(0.081)	0.070	(0.094)	-0.013	*0.469*	*0.992*

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 et al. (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 et al. (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 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 et al. (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 et al. (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 Frongillo et al. (2003). 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 stan-

dardize, we compute the age-conditional mean using the fitted values of the following regression, estimated by kernel-weighted local polynomial smoothing methods:

$$Y_i = f(X_i) + \epsilon_i \tag{7}$$

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 (7) on age of the child as shown in the following kernel-weighted local polynomial regression:

$$(Y_i - \hat{f}_i)^2 = g(X_i) + v_i \tag{8}$$

Our estimate of the age-conditional standard deviation is the square root of the fitted values in equation (8). 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 (7) and dividing by the within sample age-conditional standard deviation obtained from (8). More specifically:

$$ZY_i = \frac{(Y_i - \hat{f}_i)}{\sqrt{\hat{g}_i}} \tag{9}$$

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

12

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 an 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 aspect 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 in 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 resulted 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. Most methods indicate that two factors should be extracted from the measures of child's development at follow-up. Between 1 and 3 factors should be extracted from the measures of child's development at baseline. Most methods also suggest that two factors should be extracted from the measures of parental investments and from the measures of maternal skills. Finally, only one factor seem to underlie the measures of household wealth, according to all methods.

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

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
Dimensions to measure:				
Child's skills at follow-up	2	2	2	3
Child's skills at baseline	1	2	1	3
Parental investments at follow-up	2	2	2	3
Maternal skills at baseline	2	2	2	4

we describe now.

C.3 Specifying the dedicated measurement system

Once we have a strong indication 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.[1]

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.

Table A2 reports rotated factor loadings for each measure. Note that we have

---

[1]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.

assumed two factors for the measures of child development at baseline and at follow-up, two factors for the measures of mother's skills, two factors for the measures of parental investments and one factor for the measures of household wealth. Below we comment on these results and how they informed the final configuration of our 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 support 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 s 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 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 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 to 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 one 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: No. words the child can say	0.370*	0.143*
MacArthur: No. words the child can understand	0.047	0.097*
ICQ: Unsociable	0.118*	0.382*
ICQ: Difficult	-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 sentences 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: Number of different types of play material	1.007*	-0.006*
FCI: Number of picture books	0.334*	0.048
FCI: Number of colouring books	0.433*	-0.016
FCI: Number of home-made toys	0.102*	0.093*
FCI: Number of bought toys	0.492*	0.042
FCI: Number of toys to learn movement	0.490*	0.081*
FCI: number of toys to learn shapes	0.613*	0.025
FCI: Number of different types of play activities	0	0.985*
FCI: Times told stories in last 3 days	-0.008	0.519*
FCI: Times read in last 3 days	0.002	0.520*
FCI: Times played with toys in last 3 days	0.049	0.528*
FCI: Times labelled things in last 3 days	-0.006	0.545*

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

Mothers' years of education	0.685*	-0.018
Mother's vocabulary	0.775*	-0.001
Mother's Raven's score (IQ)	0.735*	0.034
FCI: No. of adult books at home	0.410*	-0.003
FCI: No. magazines and newspapers at home	0.249*	0.016
CESD: Are you bothered by what usual don't? (-)	-0.049	0.411*
CESD: Did you have trouble keeping mind on doing? (-)	-0.027	0.504*
CESD: Did you feel hopeful about your future? _S	0.151*	0.047
CESD: Did you feel depressed? (-)	0.000	0.674*
CESD: Did you feel everything you did was an effort? (-)	0.037	0.500*
CESD: Did you feel fearful? (-)	-0.018	0.560*
CESD: Was your sleep restless? (-)	-0.206*	0.458*
CESD: Did you feel happy?	0.05	0.454*
CESD: How often did you feel lonely in last 7 days? (-)	0.015	0.510*
CESD: Did you feel you couldn't get going? (-)	0.02	0.530*

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 and correlation between the latent factors for the treated and control households separately

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

Table C.6: Estimates of factor loadings

Measures	Latent factor							
	θ_{t+1}^C	θ_t^C	θ_{t+1}^S	θ_t^S	I_{t+1}^M	I_{t+1}^T	P_t^C	P_t^S
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	0.541	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	0.725
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	0.351	(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)
Parental material investments	0.227	(0.069)
Parental time investments	0.302	(0.068)
Maternal cognitive skills	-0.016	(0.042)
Maternal socio-emotional skills	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

$theta_{t+1}^C$	1	0.667	0.363	0.296	0.396	0.280	0.496	0.148
$theta_{t+1}^S$	0.667	1	0.257	0.261	0.267	0.225	0.345	0.044
$theta_t^C$	0.363	0.257	1	0.622	0.372	0.275	0.290	0.183
$theta_t^S$	0.296	0.261	0.622	1	0.290	0.235	0.406	0.242
I_{t+1}^M	0.396	0.267	0.372	0.290	1	0.624	0.603	0.123
I_{t+1}^T	0.280	0.225	0.275	0.235	0.624	1	0.394	0.053
P_t^C	0.496	0.345	0.290	0.406	0.603	0.394	1	0.070
P_t^S	0.148	0.044	0.183	0.242	0.123	0.053	0.070	1

Treated group

$theta_{t+1}^C$	1	0.681	0.320	0.213	0.292	0.172	0.257	0.087
$theta_{t+1}^S$	0.681	1	0.245	0.203	0.213	0.134	0.180	0.034
$theta_t^C$	0.320	0.245	1	0.445	0.282	0.284	0.234	0.183
$theta_t^S$	0.213	0.203	0.445	1	0.086	0.045	0.274	0.342
I_{t+1}^M	0.292	0.213	0.282	0.086	1	0.451	0.457	0.021
I_{t+1}^T	0.172	0.134	0.284	0.045	0.451	1	0.276	0.054
P_t^C	0.257	0.180	0.234	0.274	0.457	0.276	1	0.098
P_t^S	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 nutritional components of the intervention

	Any stimulation		Any nutrition		Nutrition and stimulation	
	Point est	St err	Point est	St err	Point est	St err
A - Child's cognitive skills at follow-up						
Bayley: cognitive	0.310***	(0.0843)	-0.0331	(0.0726)	-0.121	(0.123)
Bayley: expressive language	0.236**	(0.0909)	-0.0475	(0.0952)	-0.122	(0.125)
Bayley: receptive language	0.0805	(0.0783)	-0.0878	(0.0924)	-0.0990	(0.126)
Bayley: fine motor	0.135	(0.0889)	-0.139	(0.0840)	-0.127	(0.119)
MacArthur: words the child can say	0.143	(0.0865)	-0.118	(0.0775)	-0.106	(0.131)
MacArthur: complex phrases the child can say	0.00473	(0.0785)	0.0675	(0.0728)	0.109	(0.106)
B - Child's socio-emotional skills at follow-up						
ICQ: unsociable	0.00143	(0.0827)	0.0103	(0.0844)	0.0797	(0.110)
ICQ: difficult	0.119*	(0.0672)	-0.00156	(0.0604)	-0.0896	(0.0898)
ICQ: unstoppable	0.0448	(0.0771)	-0.0107	(0.0737)	-0.0256	(0.116)
ECBQ: inhibitory control	0.0289	(0.0782)	0.00716	(0.0789)	-0.0647	(0.116)
ECBQ: attentional focusing	0.0907	(0.0664)	-0.126*	(0.0725)	-0.0466	(0.0982)
C - Material investment at follow-up						
FCI: no. of different types of play materials	0.282***	(0.0801)	-0.116	(0.0958)	-0.138	(0.128)
FCI: no. of coloring and drawing books	-0.124	(0.0759)	0.0136	(0.0901)	-0.0172	(0.113)
FCI: no. of toys to move	0.0589	(0.0834)	-0.160	(0.108)	-0.219	(0.136)
FCI: no. of toys to learn shapes	0.443***	(0.0985)	-0.0694	(0.100)	-0.0572	(0.174)
FCI: no. of shop-bought toys	0.0455	(0.0817)	-0.121	(0.0991)	-0.0460	(0.130)
D - Time investment at follow-up						
FCI: no. of different types of play activities in last 3 days	0.287***	(0.0745)	-0.0664	(0.0678)	-0.0210	(0.105)
FCI: no. of times read to child in last 3 days	0.353***	(0.0921)	0.0410	(0.0790)	0.0201	(0.120)
FCI: no. of times told a story to child in last 3 days	0.208**	(0.0845)	-0.0839	(0.0711)	-0.143	(0.119)
FCI: no. of times played with toys in last 3 days	0.115	(0.0778)	-0.0337	(0.0700)	0.119	(0.119)
FCI: no. of times named things to child in last 3 days	0.0783	(0.0607)	-0.0219	(0.0694)	0.118	(0.0992)

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.

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***
	(0.0637)	(0.0656)
FCI: no. of books to pain and draw	-0.133**	-0.136**
	-(0.056)	(0.0656)
FCI: no. of toys bought	0.0240	0.0477
	(0.0611)	(0.0621)
FCI: no. of toys to learn movement	-0.0476	-0.0505
	(0.0646)	(0.0674)
FCI: no. of toys to learn shape	0.416***	0.421***
	(0.0877)	(0.0924)
B - Time investment at follow-up		
FCI: no. of different types of play activities in last 3 days	0.277***	0.273***
	(0.0503)	(0.0516)
FCI: no. of times told a story to child in last 3 days	0.138**	0.113*
	(0.0601)	(0.0602)
FCI: no. of times read to child in last 3 days	0.362***	0.350***
	(0.0625)	(0.0651)
FCI: no. of times played with toys in the last 3 days	0.175***	0.193***
	(0.0597)	(0.0632)
FCI: no. of times named things to child in last 3 days	0.137***	0.134***
	(0.0483)	(0.0498)
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	4.26	(6.38)
Toys made with the home visitor	0.84	(2.64)
Intervention books	5.20	(3.39)

Note: Sample size is 667 observations.

E Specification and robustness checks

In this section of the appendix, we report results associated with specification and robustness checks we discuss in the main text.

Table E.1 reports the estimates of the investment functions where the investment function is fully interacted with the treatment indicator.

Table E.2 report the estimates of the Cobb-Douglas production functions for cognitive skills and socio-emotional skills where all the share parameters (as well as the total productivity factor A) 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.3 reports the estimates of the production functions in the case where the control group is made of children who neither received the stimulation nor the nutritional supplementation component of the intervention and the treated group is only made of children who received the stimulation, but not the nutritional supplementation component.

Table E.4 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.

E.1 Invariance of the measurement system between the treated and control groups

Tables E.5 and E.6 report estimates of the factor loadings in the measurement system where we allow the loadings associated with measures of child's skills and parental

31

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.7 report the estimates of our favorite specifications for the production function for cognitive and socio-emotional skills where the latent factors are estimated using this measurement system.

Table E.1: Material and time investment equations where all parameters are interacted with treatment

	Material investment	Time investment
Constant	-0.017	0.002
	[-0.112,0.086]	[-0.092,0.094]
x 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 (t)	1.088	0.581
	[0.642,1.522]	[0.151,0.961]
x Treatment	-0.185	-0.126
	[-0.436,0.094]	[-0.343,0.098]
Log mother's socio-emotional skill (t)	0.171	-0.072
	[-0.099,0.43]	[-0.333,0.183]
x Treatment	-0.061	0.068
	[-0.223,0.105]	[-0.111,0.242]
Log number of children (t+1)	-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]
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 variables are the log of material and time investment factors measured at follow-up. t and $t+1$ denote that a variable was measured at baseline and endline, respectively. 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.2: Production functions where all parameters are allowed to vary with treatment

	Cognitive skill	Socio-emotional skill
	IV (prices, treatment)	OLS
TFP	-0.013	-0.009
	[-0.103,0.073]	[-0.085,0.063]
x Treatment	0.148	-0.002
	[-0.069,0.383]	[-0.11,0.107]
Log child's cognitive skill (t)	0.476	0.077
	[0.193,0.816]	[-0.172,0.356]
x Treatment	0.133	0.02
	[-0.073,0.332]	[-0.148,0.169]
Log child's socio-emotional skill (t)	0.043	0.828
	[-0.178,0.328]	[0.529,1.159]
x Treatment	-0.038	-0.201
	[-0.229,0.131]	[-0.385,-0.027]
Log mother's cognitive skill (t)	0.366	-0.288
	[-0.165,1.241]	[-0.703,0.122]
x Treatment	-0.23	0.127
	[-0.931,0.301]	[-0.118,0.376]
Log mother's socio-emotional skill (t)	0.152	0.017
	[-0.106,0.41]	[-0.254,0.275]
x Treatment	-0.038	0.015
	[-0.181,0.128]	[-0.138,0.179]
Log of number of children (t)	0.165	0.13
	[-0.049,0.327]	[-0.023,0.274]
x Treatment	-0.048	-0.022
	[-0.185,0.111]	[-0.122,0.08]
Log of material investment (t+1)	0.33	0.243
	[-0.79,0.95]	[0.006,0.534]
x Treatment	0.052	-0.062
	[-0.634,0.907]	[-0.21,0.07]
Log of time investment (t+1)		-0.009
		[-0.404,0.295]
x Treatment		0.067
		[-0.102,0.298]

P-value of F-test of joint significance of interactions with treatment dummy		
Excluding the interaction with the intercept	0.288	0.378
Including the interaction with the intercept	0.318	0.47

Note: Dependent variable in column 1 (2) is the log of cognitive (socio-emotional) skill measured at follow-up. 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. t and $t+1$ denote that a variable was measured at baseline and endline, respectively. 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: Estimates of the production functions on the sample excluding children who received nutritional supplementation

	(1) Cognitive skill	(2) Socio-emotional skill
	IV (prices, treatment)	OLS
TFP	-0.006	-0.007
	[-0.122,0.063]	0.103,0.087]
TFP × 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 (t)	-0.123	-0.142
	[-0.64,0.197]	[-0.383,0.088]
Log mother's socio-emotional skill (t)	0.115	-0.007
	[-0.021,0.258]	[-0.14,0.131]
Log number of children (t)	0.073	0.106
	[-0.016,0.163]	[0.028,0.182]
Log material investment (t+1)	0.453	0.14
	[0.154,0.97]	[0.022,0.299]
Log time investment (t+1)		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. 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. t and $t+1$ denote that a variable was measured at baseline and endline, respectively. The sample excludes children who received the nutritional 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.4: 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
TFP	-0.019	-0.011
	[-0.117,0.051]	[-0.069,0.045]
TFP × 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 (t)	-0.143	0.014
	[-0.659,0.241]	[-0.128,0.137]
Log mother's socio-emotional skill (t)	-0.039	0.057
	[-0.187,0.084]	[-0.024,0.129]
Log number of children (t)	0.093	0.072
	[0.018,0.183]	[0.023,0.121]
Log material investment (t+1)	0.617	0.11
	[0.153,1.236]	[0.037,0.206]
Log time investment (t+1)		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. 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. t and $t+1$ denote that a variable was measured at baseline and endline, respectively. 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.5: Estimates of factor loadings in the control group

Measures	θ_{t+1}^C	θ_t^C	θ_{t+1}^S	θ_t^S	I_{t+1}^M	I_{t+1}^T	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.6: Estimates of factor loadings in the treated group

Measures	θ_{t+1}^C	θ_t^C	θ_{t+1}^S	θ_t^S	I_{t+1}^M	I_{t+1}^T	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.7: 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 (t)	-0.104	-0.083
	[-0.495,0.201]	[-0.222,0.042]
Log child's socio-emotional skill (t)	0.073	0.041
	[-0.024,0.151]	[-0.054,0.125]
Log number of chidlren (t)	0.086	0.168
	[0.017,0.181]	[0.074,0.277]
Log material investment (t+1)	0.517	0.11
	[0.16,1.03]	[0.003,0.198]
Log time investment (t+1)		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. 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. t and $t+1$ denote that a variable was measured at baseline and endline, respectively. 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.

E.2   Reduced form production functions

In this section we present the reduced form production functions, where the material and time investments have been substituted out.

Table E.8: Reduced form Production Functions

	Cognitive	Socioemotional
Intercept	-0.019	-0.01
	[-0.097,0.062]	[-0.086,0.066]
Treatment dummy	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 skills (t)	-0.002	0.527
	[-0.087,0.084]	[0.409,0.668]
Log mother's cognitive skill (t)	0.279	0.073
	[0.169,0.411]	[-0.049,0.201]
Log mother's socio-emotional skill (t)	0.11	0.047
	[0.039,0.181]	[-0.049,0.13]
Log number of children (t)	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: 95% Bootstrap confidence intervals in brackets

E.3 Monte Carlo for evaluating the presence of Weak Instruments bias

Here 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.[2] 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

---

[2]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.

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:

$$u_{3n} \sim N(0, \Sigma_i) \tag{10}$$

Table E.9: Covariances used for the Monte Carlo Simulation

	Covariance A					
	Within Person Σ_d			Spatial Covariance Σ_{off}		
	Material	Time	Cognition	Material	Time	Cognition
Material	0.623	0.202	-0.179	0.04	0	0
Time	0.202	0.561	0.038	0	0.04	0
Cognition	-0.179	0.038	0.506	0	0	0.04
	Covariance B					
	Within Person Σ_d			Spatial Covariance Σ_{off}		
	Material	Time	Cognition	Material	Time	Cognition
Material	0.823	0.351	-0.200	0.147	0.132	-0.006
Time	0.351	0.815	0.052	0.132	0.207	0.011
Cognition	-0.200	0.052	0.595	-0.006	0.011	0.046

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 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 E.9.

The simulated data is generated as follows: we draw errors for all individuals jointly in each community based on equation (10). 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 (0.463) 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 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.[3]

Table E.10: Standard errors from estimated model and standard deviations from the Monte Carlo Simulations

	Estimated Model		Covariance A		Covariance B	
	Investment Equations					
	Material	Time	Material	Time	Material	Time
Toy Price	0.036	0.031	0.028	0.031	0.037	0.044
Food Price	0.043	0.038	0.033	0.033	0.043	0.051
Conflict	0.037	0.028	0.028	0.030	0.038	0.043
	Production Function					
Material Investment	0.299		0.270		0.338	
Time Investment	0.374		0.316		0.377	

In Table E.10 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 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 (as also implied by comparing the rank tests). Under covariance A the price coefficients have a standard deviation a bit lower than the standard errors in the estimated model but this time the coefficient on the conflict

---

[3]The estimated production function is a bit more precise because we use Maximum likelihood to estimate the distribution of the factors as in Attanasio et al. (2017). 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 E.11: Monte Carlo Simulations

| | True Values | | Covariance A | | Covariance B | |
| | | | Material and Time investment functions on Simulated Data | | | |
	Material	Time	Material	Time	Material	Time
Intercept	-0.016	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	0.204	0.333	0.188	0.33	0.173	0.291
			(0.063)	(0.064)	(0.087)	(0.096)
			[0.059,0.313]	[0.202,0.454]	[0.003,0.336]	[0.106,0.479]
Baseline Cognition	0.13	0.067	0.132	0.071	0.122	0.064
			(0.041)	(0.042)	(0.045)	(0.048)
			[0.056,0.214]	[-0.008,0.155]	[0.03,0.21]	[-0.031,0.153]
Mother's Cognition	0.754	0.367	0.749	0.39	0.692	0.349
			(0.049)	(0.045)	(0.050)	(0.047)
			[0.659,0.849]	[0.299,0.481]	[0.596,0.793]	[0.26,0.445]
Price of Toys	-0.095	-0.02	-0.089	-0.019	-0.082	-0.017
			(0.028)	(0.031)	(0.037)	(0.044)
			[-0.143,-0.036]	[-0.077,0.037]	[-0.154,-0.013]	[-0.101,0.073]
Price of Food	0.096	0.042	0.089	0.04	0.081	0.037
			(0.033)	(0.033)	(0.043)	(0.051)
			[0.024,0.154]	[-0.027,0.103]	[0,0.163]	[-0.061,0.142]
Conflict	-0.011	-0.096	-0.01	-0.093	-0.007	-0.085
			(0.028)	(0.030)	(0.038)	(0.043)
			[-0.064,0.048]	[-0.152,-0.037]	[-0.076,0.07]	[-0.172,0.001]

Rank test			0.047		0.029	
Cragg-Donald test			0.059		0.023	

| | True Values | Cognitive Production function on Simulated Data | | | |
| | | Investment treated as | | Investment treated as | |
		Exogenous	Endogenous	Exogenous	Endogenous
Intercept	-0.02	0	0	0	0
		(0.000)	(0.001)	(0.000)	(0.001)
		[0,0]	[-0.001,0.001]	[0,0]	[-0.001,0.001]
Treatment	0.076	0.085	0.067	0.075	0.054
		(0.055)	(0.122)	(0.052)	(0.133)
		[-0.018,0.194]	[-0.144,0.276]	[-0.031,0.175]	[-0.203,0.293]
Baseline Cognition	0.663	0.635	0.611	0.607	0.586
		(0.037)	(0.052)	(0.035)	(0.055)
		[0.563,0.706]	[0.506,0.714]	[0.544,0.675]	[0.478,0.687]
Mother's Cognition	0.018	0.182	0.02	0.162	0.014
		(0.043)	(0.212)	(0.042)	(0.231)
		[0.104,0.264]	[-0.407,0.385]	[0.076,0.245]	[-0.445,0.467]
Material investment	0.463	0.111	0.418	0.131	0.432
		(0.033)	(0.270)	(0.033)	(0.338)
		[0.045,0.176]	[-0.084,1.059]	[0.066,0.197]	[-0.132,1.103]
Time investment	0	0.162	-0.003	0.185	0.017
		(0.033)	(0.316)	(0.034)	(0.377)
		[0.097,0.23]	[-0.655,0.516]	[0.118,0.253]	[-0.794,0.642]

Note: The Monte Carlo simulations employs 1000 replications. Average estimates across replications with standard deviation in parentheses and bottom and top 2.5th percentile in square brackets. We produce a simulated observation for each household in the data using all their observed exogenous variables, including the instruments. So the simulated data has exactly the same structure as the actual data. The data generating process includes the transformation from latent factors to measurements to produce data similar to the one we observe. Each replication implements the entire estimation procedure as described in the full model, including the factor scoring. Investments are endogenous in the simulated data: the error structure assumed in the simulation allows for a covariance between the investments and cognition residuals and for spatial dependence as in the data. The errors are normal.

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 Table E.11 we show the corresponding Monte Carlo simulations for the three equation system of material and time investments and cognitive skills. The coefficients on 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.

E.4 Allowing for 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. The results are substantially the same as those reported in the main paper and none of the conclusions change. Where there are some differences these are always insignificant.

Table E.12: Investment equations with stratum fixed effects

	Material investment	Time investment
Intercept	0.187	0.095
	[-0.006,0.382]	[-0.033,0.232]
Treatment dummy	0.187	0.283
	[0.006,0.356]	[0.122,0.44]
Log child's cognitive skill (t)	0.082	0.065
	[-0.021,0.193]	[-0.045,0.18]
Log child's socio-emotional skill (t)	-0.018	0.031
	[-0.124,0.116]	[-0.085,0.162]
Log mother's cognitive skill (t)	0.848	0.443
	[0.675,1.032]	[0.221,0.607]
Log mother's socio-emotional skill (t)	0.081	0.043
	[0.006,0.153]	[-0.041,0.136]
Log number of children (t)	-0.018	-0.008
	[-0.064,0.031]	[-0.057,0.037]
Log toy price	-0.058	0.001
	[-0.134,0.017]	[-0.065,0.067]
Log food price	0.141	0.094
	[0.036,0.25]	[0.009,0.192]
Maternal childhood exposure to conflict	0.003	-0.063
	[-0.082,0.085]	[-0.121,-0.005]
Central region	-0.217	0.005
	[-0.467,0.036]	[-0.19,0.193]
Oriental region	-0.366	-0.284
	[-0.621,-0.1]	[-0.53,-0.048]
Rank test	0.063	
Cragg-Donald test	0.061	
Test for joint Significance - p-values		
Toy price, food price, conflict	0.044	0.026
Toy price, food price, conflict, treat	0.005	0.000
Toy price, food price	0.020	0.167
Toy price, food price, treat	0.001	0.000
Conflict, treat	0.143	0.000

Note: Investment equations including stratum fixed effects. 95% Confidence Intervals in brackets. Zona Cafetera is the reference region.

Table E.13: Cognitive and socio-emotional production functions with stratum fixed effects

	Cognition		Socio-Emotional	
	OLS	IV	OLS	IV
		Prices, conflict		Prices, conflict
Intercept	-0.022	-0.053	0.116	0.161
	[-0.128,0.082]	[-0.203,0.056]	[0,0.229]	[-0.019,0.303]
Treatment	0.086	0.129	0.002	-0.124
	[-0.014,0.191]	[-0.061,0.286]	[-0.106,0.106]	[-0.288,0.148]
Log child's cognitive skill (t)	0.68	0.719	0.041	-0.011
	[0.591,0.777]	[0.579,0.804]	[-0.069,0.149]	[-0.114,0.16]
Log child's socio-emotional skill (t)	-0.034	-0.021	0.541	0.51
	[-0.125,0.049]	[-0.123,0.099]	[0.418,0.689]	[0.391,0.693]
Log mother's cognitive skill (t)	0.191	0.073	-0.061	0.065
	[0.059,0.338]	[-0.39,0.336]	[-0.211,0.076]	[-0.391,0.426]
Log mother's socio-emotional skill (t)	0.095	0.092	0.034	0.042
	[0.026,0.167]	[-0.017,0.168]	[-0.057,0.121]	[-0.063,0.151]
Log of number of children (t)	0	-0.002	0.008	-0.006
	[-0.044,0.046]	[-0.046,0.054]	[-0.035,0.054]	[-0.049,0.061]
Log of material investment (t+1)	0.068	0.335	0.133	-0.507
	[-0.011,0.143]	[-0.007,0.984]	[0.042,0.231]	[-0.845,0.29]
Log of time investment (t+1)	0.064	-0.21	0.094	0.984
	[-0.024,0.159]	[-0.927,0.324]	[-0.023,0.199]	[-0.057,1.409]
Central region	-0.132	-0.054	-0.176	-0.335
	[-0.273,0.006]	[-0.221,0.199]	[-0.326,-0.027]	[-0.531,-0.055]
Oriental region	0.162	0.183	-0.182	-0.131
	[0.026,0.309]	[0.009,0.392]	[-0.341,-0.038]	[-0.401,0.082]

Note: Production functions including stratum fixed effects. 95% confidence Intervals in brackets. Zona Cafetera is the reference region.