From Quantity to Quality: Delivering a Home-based Parenting Intervention through China's Family Planning Cadres

Sean Sylvia, Nele Warrinnier, Renfu Luo, Ai Yue,
Orazio Attanasio, Alexis Medina, and Scott Rozelle*

Draft July 2017

Abstract

A key challenge in developing countries interested in providing early childhood development programs at scale is whether these programs can be effectively delivered through existing public service infrastructures. We present the results of a randomized experiment evaluating the effects of a home-based parenting program delivered by cadres in China's Family Planning Commission (FPC) - the former enforcers of the one-child policy. We find that the program significantly increased cognitive development after six months and that increased investments by caregivers was a major mechanism through which this occurred. Compliance with the program was associated with the degree to which participants had a favorable view of the FPC, which also increased due to the program.

JEL Classification: J13, I21, I28, H11

Keywords: Early Childhood Development, Parenting, China, Poverty, Family Planning

---

*Sylvia: University of North Carolina at Chapel Hill (email:sean_sylvia@unc.edu); Warrinnier: LICOS, KULeuven and CEP, London School of Economics (email:n.warrinnier@lse.ac.uk); Luo: Peking University (email:luorf.ccap@pku.edu.cn); Yue (Corresponding Author): Shaanxi Normal Univeristy (email:yueai@163.com); Attanasio: University College London (email:o.attanasio@ucl.ac.uk); Medina: Stanford University (email:amedina5@stanford.edu); Rozelle: Stanford University (email:rozelle@stanford.edu).
We thank Cai Jianhua and the China National Health and Family Planning Commission for their support on this project. We are also grateful for funding from the Stanford Policy Implementation Lab and the International Initiative for Impact Evaluation (3ie). We would also like to thank Jim Heckman for his support and conversations and also acknowledge the support of Shasha Jumbe and the Gates Foundation's Healthy Birth, Growth and Development Knowledge Integration (HBGDki), China Program.

1 Introduction

A growing body of cross-disciplinary research highlights the importance of a child's environment in the first years of life for skill development and outcomes over the life course (Knudsen et al., 2006). This period is thought to be important for human capital accumulation both because very young children are sensitive to their environment and because advantage or deprivation during this period can have long-term consequences. Cognitive science research suggests that malleability of cognitive ability is highest in infancy and decreases over time (Nelson and Sheridan, 2011). Due to the hierarchical nature of brain development – whereby higher level functions depend and build on lower level ones – cognitive deficiencies in early life can permanently hinder skill development. The nature of cognitive development may further lead to important dynamic complementarities in the production of human capital where early skills increase the productivity of later human capital investments and encourage more investment as a result (Cunha et al., 2010; Attanasio et al., 2015).

These mechanisms may explain findings of large long-run effects of early childhood interventions (Cunha and Heckman, 2007). Long-term follow-up studies of early childhood interventions to improve nutrition and create stimulating environments have found large and wide-ranging effects into adulthood. These studies found programs to have increased college attendance, employment, and earnings as well as cause reductions in teen pregnancy and criminal activity (Heckman et al., 2010; Walker et al., 2011; Gertler et al., 2014).

Findings from this body of research provide strong support for investments in early childhood programs (Carneiro and Heckman, 2003). Particularly in low and middle-income countries, the social returns to early intervention could be substantial due to the large number of children that are at risk of becoming developmentally delayed. Estimates indicate that 250 million children (43%) younger than 5 years in low-income and middle-income countries are at risk of not reaching their full development potential (Lu et al., 2016). While there are several reasons that so many children are at risk in developing countries, a significant factor is that children often lack a sufficiently stimulating environment (Black et al., 2017). Partly as a result of this evidence, Early Childhood Development (ECD) has moved up high on the policy agenda, as evidenced by its inclusion in the United Nation's Sustainable Development Goals (Nations, 2015).

A key challenge facing policy makers is how to deliver ECD programs cost effectively at scale (Berlinski et al., 2016; Richter et al., 2017). Providing ECD interventions at scale is challenging largely due to the infrastructure required to deliver services effectively to families in need, many of whom live in hard-to-reach communities such as urban slums and sparsely populated rural areas. Because building a new infrastructure to support ECD services alone would be costly, some have suggested integrating ECD programs into existing public service infrastructures (Richter et al., 2017). For example, international agencies including the World Bank, the Inter-American Development Bank, the United Nations and the World Health Organization have called for ECD to be integrated into health and nutrition programs (Chan, 2013; Black and Dewey, 2014). Whether such a strategy can be successful is an open question. It is unclear, for example, if existing personnel who have been working in other areas and have little or no background in early childhood education can be trained to effectively deliver an ECD program. Moreover, it is often the case that public sector agencies resist new tasks, particularly if they are perceived as misaligned with the organization's existing mission (Wilson, 1989; Dixit, 2002).

We study the promotion of ECD in rural China through a home-based parent training intervention implemented by one of the world's largest bureaucracies, the China Family Planning Commission (FPC). In recent years, the Chinese government has relaxed its family planning laws and, since January 2016, has allowed all parents to conceive two children without penalty. Relaxation of the One Child Policy (OCP) and changing fertility preferences have greatly diminished the need for enforcement, and the FPC has begun to shift focus to other areas including ECD (Wu et al., 2012). Delivering ECD policies through the infrastructure of the FPC has promise but also potentially significant challenges. It is therefore unclear – even if an intervention itself is efficacious – whether it can be effectively delivered through the apparatus of the FPC.[1] This study investigates whether it is possible to re-train cadres formerly responsible for enforcing the OCP into effective parenting teachers. In other words, can the local knowledge and infrastructure of the FPC – which has been responsible for managing the quantity of human capital – be used to effectively raise the quality of human capital in China?

To study the effects of an FPC-delivered home-based parenting intervention, we conducted a cluster-randomized controlled trial across 131 villages in Shaanxi Province, located in northwestern China. We worked with the FPC to re-train 70 cadres to deliver a structured curriculum aimed at improving parenting practices in early childhood through weekly home visits. Loosely modeled on the Jamaican Early Childhood Development Intervention (Grantham-McGregor et al., 1991), the curriculum was designed with ECD experts in China and aimed to train and encourage caregivers to engage in stimulating activities with their children.

[1] See China Central Television (CCTV) News report: How will a Million Family Planning Workers Transition? https://youtu.be/84WIe1C3XTM

We find that the intervention increased cognitive development outcomes of children assigned to receive weekly home visits by approximately 0.25 standard deviations, but find no effect on psychomotor development or social-emotional development. Effects on cognitive outcomes were accompanied by increases in parental investment, changed beliefs regarding parenting practices and a change in disciplining practices. However, of these, parental investment was the main mediator of effects on cognitive development. The average effect of the program was diminished by imperfect compliance. We find evidence that one of the primary factors hindering compliance – unfavorable public perception of the FPC – was also significantly reduced as a result of the program. This suggests that compliance may improve as the program is implemented over time.

Our findings add to an emerging literature studying how ECD can be integrated into existing infrastructure in developing countries to facilitate delivery at scale. Attanasio et al. (2014) found that a parenting intervention integrated into an existing conditional cash transfer program in Colombia and delivered by local volunteers successfully improved cognitive development outcomes, and, like the program we study in China, did so primarily through increased parental investments (Attanasio et al., 2015). In Pakistan, Yousafzai et al. (2014) find significant improvements in early childhood outcomes of children enrolled in a parenting intervention integrated in a community-based health service. Local community workers employed by the Lady Health Worker Program - a government initiative set up in the early 90's to increase access to primary health services in rural and urban slum areas - were retrained to deliver a psychosocial stimulation intervention along side their primary health care services during monthly household visits. A follow-up study of the intervention shows that program effects persist 2 years after termination of the parenting intervention (Yousafzai et al., 2016). Our study adds to the literature by providing evidence on the effectiveness of an ECD intervention integrated into local government services in China: specifically whether the infrastructure and personnel of the FPC can effectively implement a home-based parenting program and reduce the high prevalence of cognitive delay among infants and toddlers in China's poor rural regions.

The remainder of the paper is structured as follows. In the next section we discuss the FPC and how their role is changing with the abolishment of the One Child Policy. In section 3 we describe the experimental design and data collection. In section 4 we present the main effects on cognitive development and analyze drivers of compliance and impact heterogeneity. Section 5 concludes.

2 Background: The Changing Role of the FPC

The Family Planning Commission (FPC)[2] is the entity responsible for the implementation of population and family planning policies in China. From 1980, a large part of the agency's mandate included enforcement of the One Child Policy – a policy comprised of a set of regulations governing family size.[3] Although there were several, now well-documented, unintended consequences of the policy, the government at the time considered population containment necessary to improve living standards as the country faced an impending baby boom (Hesketh et al., 2005).

The implementation of China's One China Policy required close interaction between families and local FPC cadres to ensure universal access to contraceptive methods, to monitor for violations, and to enforce penalties. Though details of how the policy was implemented varied across regions and time, at its most intense phase of implementation families were required to obtain birth permits before pregnancy and births were to be registered with the local FPC cadre. Once families met their number of allowed children, FPC officers often encouraged or forced sterilization (Greenhalgh, 1986). The FPC also enforced penalties for out-of-plan births which included substantial fines and loss of employment. There is also wide documentation of forced abortion of out-of-plan pregnancies.

Given the numerous and complicated set of policy instruments, and the close interaction with families that this entailed, implementation of the One Child Policy required a large bureaucracy. As of 2005, the FPC had more than 500,000 administrative staff and more than 1.2 million village-level FPC operatives.[4] In 2016, the budget supporting the FPC's activities exceeded 8.85 billion dollars.[5] However, after debates in recent years about the necessity of the One Child Policy's continuation, the government announced in October 2015 that the policy would be formally terminated as of January 1, 2016.[6] Termination of the policy has also called into question the future role of the FPC.[7]

[2]In March 2013, the National Population and Family Planning Commission was merged with the Ministry of Health to form the current National Health and Family Planning Commission.

[3]Despite its name, most families were not restricted to having only one child. In many rural areas, families were allowed two children and there were a number of other exemptions including for minority groups and for parents who worked in high-risk occupations. See Hesketh et al. (2005) and Hesketh et al. (2015) for good overviews of the policy and implementation.

[4]See NPFPC, 2006, Statistical Bulletin of Forth National Population and Family Planning System Statistical, http://www.nhfpc.gov.cn/guihuaxxs/s10741/201502/f68e73331a9147e78209ab81bd156a39.shtml

[5]Includes funding for health and family planning activities. See NHFPC, 2016, The Departmental budget report of National Health and Family Planning commission of the PRC, http://www.nhfpc.gov.cn/caiwusi/s3574/201604/3582098e060144148a1e3b4f3f1a4fe0.shtml

[6]The Central Committee of the Communist Party of China, 2015. Bulletin of Fifth Plenary Session of 18th CPC Central Committee.

[7]See Sonmez, F., Wall Street Journal, 2015. After the One-Child Policy: What Hap-

Some have argued that an appropriate future focus of the FPC would include early childhood care and education, which falls within the technical purview of the agency (Wu et al., 2012). Currently, responsibility for providing these services is spread across multiple entities, which in practice has led to a gap in service provision (Wu et al., 2012). Whether the FPC would be able to effectively fill this role is an open question, however. On one hand, the FPC has the ideal infrastructure to provide early childhood services: a large, well-functioning organization with representation in every village and community in the country; a relatively well-educated work force; and the ability to maintain information on every family and child. On the other hand, it may be difficult for FPC cadres to retrain and effectively deliver ECD services. More significantly, the agency's history and reputation could limit its effectiveness. Although the enforcement of the policy relaxed over time, the agency's at times draconian measures may have created lasting social animosity toward the family planning commission that could hinder its effective delivery of ECD services.[8]

3 Experimental Design and Data Collection

3.1 Sampling and Randomization

The study sample was selected from Shangluo prefecture located in Shaanxi province. Shaanxi is a relatively poor province located in Northwest China, ranking 19th out of 31 provinces nationally in terms of GDP per capita. Shangluo prefecture is located in a mountainous and relatively poor region of Shaanxi province.

To identify the sample, we first selected townships from four nationally-designated poverty counties in Shangluo. All townships in each county were included except the one township in each county that housed the county seat. Within each township, government data was used to compile a list of all villages reporting a population of at least 800 people. We then randomly selected two villages from the list in each township. These exclusion criteria were chosen to ensure a rural sample and increase the likelihood that sampled villages had a sufficient number of children in the target age range. Our final sample consisted of 131 villages total. All children in sample villages between 18 and 30 months of age were enrolled in the study. At baseline, a total of 592 children were sampled.

Following baseline data collection (described below), 65 villages were randomly assigned to the parenting intervention group and the remaining 66 to a control group. The

pens to China's Family-Planning Bureaucracy? http://blogs.wsj.com/chinarealtime/2015/11/12/after-the-one-child-policy-what-happens-to-chinas-family-planning-bureaucracy/

[8]See China Central Television (CCTV) News report: How will a Million Family Planning Workers Transition? https://youtu.be/84WIe1C3XTM

randomization procedure was stratified by county, child cohort, and experimental group of an earlier trial.[9] Each trainer was assigned a maximum of four families chosen randomly from treatment villages to be enrolled in the program. In treatment villages, a total of 212 children were enrolled and the remaining 79 were not. In the analysis, we test for spillover effects on these children in treatment villages who were not selected to participate.

3.2 Parenting Program

Parenting trainers, selected by the FPC from among their cadres in each township, delivered a structured curriculum through weekly home visits to households in treatment villages for a period of six months (from November 2014 to April 2015). Based loosely on the Jamaican home visiting model (Grantham-McGregor et al., 1991) and adapted by child development psychologists in China to the local setting, the goal of the intervention was to train caregivers to interact with their children through stimulating and developmentally-appropriate activities.

The curriculum delivered by the parenting trainers was developed by the research team in collaboration with the FPC and outside ECD experts in China. The curriculum was stage-based and fully scripted. Weekly age-appropriate sessions were developed targeting children from 18 months of age to 36 months of age. Each weekly session contained modules focused on two of four total developmental areas: cognition, language, socio-emotional, and (fine and gross) motor skills. Every two weeks, caregivers would encounter one activity from each category. In addition to developmental activities, the curriculum also included one weekly module on child health/nutrition.

During sessions, parent trainers would introduce caregivers to the activity and assist caregivers to engage in the activity with their child. Typically the only caregiver that participated was the primary caregiver (usually mother or grandmother), though other caregivers sometimes observed. At the end of each weekly session, the materials used for that week's activities (toys and books) were left in the household to be returned at the next visit.

Parenting trainers were selected and deployed by the FPC office in each township. Summary statistics on trainer characteristics are shown in Table 8. Around 60 percent of

[9] The children included in this study were previously part of the sample for a randomized trial testing the effect of micronutrient powders aimed at reducing anemia. For that trial, children were recruited when they were between 6 and 12 months of age. Recruitment was done in two six-month cohorts. The findings of that trial, reported in Luo et al. 2016, show no effect of micronutrient powder on hemoglobin or anemia at 18 months. The treatment assignment for the parenting intervention evaluated in this study was stratified on the arms of the earlier trial. There is no evidence that effects of the parenting intervention vary across arms of the earlier trial.

the parenting trainers deployed by the FPC office were men. The majority of parenting trainers were married and had children themselves. The parenting trainers were well educated with most of them having enjoyed a community college higher education and around 30 percent have obtained a bachelor degree. On average, parenting trainers were 34 years old and had worked 12 years for the Family Planning Commission.

Fully scripting the curriculum eliminated the need for extensive training of parent trainers. All parenting trainers underwent an initial, centralized one-week intensive training at the beginning of the program which covered theories and principles of early childhood development, parenting skills, and the curriculum. This initial training consisted of both classroom-based instruction as well as field practice. Throughout the program, trainers received periodic training by phone on curriculum activities which would vary according to the ages of children to whom they were assigned.

FPC offices assigned parent trainers to enrolled families in their township. Most trainers were assigned families in only one village.

3.3 Data Collection

We conducted our baseline survey in October 2014 and our follow-up survey in May 2015. Teams of enumerators collected detailed information on children, caregivers and households. Each child's primary caregiver was identified and administered a survey on child, parent and household characteristics including each child's gender, birth order, maternal age and education. Each child's age was obtained from his or her birth certificate. The primary caregiver was identified by each family as the individual most responsible for the infant's care (typically the child's mother or grandmother).

Children's cognitive, psychomotor and social-emotional development were assessed in each round. At baseline, all children were assessed using the Bayley Scales of Infant Development (BSID) Version I, a standardized test of infant cognitive and motor development (Bayley, 1969). The test was formally adapted to the Chinese language and environment in 1992 and scaled according to an urban Chinese sample (Yi et al., 1993; Huang et al., 1993). Following other published studies that use the BSID to assess infant development in China (Li et al., 2009; Chang et al., 2013; Wu et al., 2011), it was this officially adapted version of the test that was used in this study (Yi, 1995). All BSID enumerators attended a week-long training course on how to administer the BSID, including a 2.5 day experiential learning program in the field. The test was administered in the household using a standardized set of toys and detailed scoring sheet. The BSID takes into consideration each child's age in days, as well as whether he or she was premature at birth. These two factors, combined with the child's performance on a series of tasks using the standardized toy kit, contribute to the establishment of the two sub-indices: the Mental Development Index (MDI), which evaluates memory, habitation, problem solving, early number concepts, generalization, classification, vocalization and language to produce a measure of cognitive development; and the Psychomotor Development Index (PDI), which evaluates gross motor skills (rolling, crawling and creeping, sitting and standing, walking, running and jumping) and fine motor skills to produce a measure of psychomotor development (Bayley, 1969).

Because the BSID-I is not designed to assess outcomes for children older than 30 months, only children aged 30 months or under at follow-up (approximately half of the sample) were administered the BSID in the follow-up survey. Older children were assessed using the Griffith Mental Development Scales (GMDS-ER 2-8) (Luiz et al., 2006), which has been shown to be comparable in its assessment of early childhood development to the BSID-I (Cirelli et al., 2015).[10]

Enumerators were trained for two days on how to administer the Griffith Mental Development Scales. As is the case with the BSID, a standard kit with equipment such as building blocks and a drawing book is used to test different skill sets of the children, and enumerators are supplied with a standard scoring form. Testing also took place in the household. The GMDS-ER 2-8 comprises six sub scales: locomotor, personal-social, language (receptive and expressive), hand and eye coordination, performance, practical reasoning.[11]

In our main results, we present estimates pooling both cohorts by combining sub-scales of both tests to construct a cognitive development score and a motor development score.[12] The cognitive development measure uses scores from the MDI sub-index of the BSID for the children up to 30 months and uses scores from the Performance sub-index of the GMDS-ER for children older then 30 months. The motor development measure uses scores from the PDI sub-index of the BSID for children up to 30 months and score from the Locomotor sub-index of the GMDS-ER for older children.

For the analysis, raw scores are standardized separately by sub-index. Since raw scores are increasing in age, we internally standardize to remove the effect of age (measured in months). More specifically, we compute age-adjusted internal z-scores by subtracting from the raw scores the empirical age-conditional means and dividing by the empirical age-conditional standard deviations estimated using non-parametric regression methods.

[10]The Pearson correlation coefficient between the BSID and GMDS is found to be higher than 0.8.

[11]The last sub-scale of the GMDS-ER, practical reasoning, is only used to assess development of older children, hence was not registered to this particular age group.

[12]Estimates for each cohort separately using all sub-indexes of both BSID and GMDS-ER are shown in Appendix Table A2.

This method is less sensitive to outliers and small sample size within age-category and gives us normally distributed internally standardized scores with mean zero across the age range (in months)(Attanasio et al., 2015).[13]

Additionally, in each wave children's social-emotional behavior was assessed using the Ages and Stages Questionnaire: Social Emotional (ASQ:SE) (Squires et al., 2003). The items in this questionnaire (which vary by age) measure a child's tendency towards a set of behaviors such as ability to calm down, accept directions, demonstrate feelings for others (empathy), communicate feelings, initiate social responses to parents and others, and respond without guidance (move to independence). Main caregivers were asked to indicate whether the child exhibits these behaviors *most of the time*, *sometimes*, or *never*. Depending on the desirability of the behavior, answered are scored either 0, 5, or 10 points. Children who score 60 or more are considered to require further assessment for social-emotional problems.

Our hypothesis is that effects of the parenting intervention on child development are driven (at least partially) through increased quality of the home stimulation environment and increased parental knowledge of child development. Hence, we collect detailed information in each survey wave on the above. We measure the quality of the home stimulation environment by asking whether the main caregiver engaged in a set of child-rearing activities, such as story-telling and playing with toys, the previous day and how many children's books they have in the house. Parental knowledge of child development was measured by asking the main care giver how they feel about the importance of reading and play for child development using a 7-point likert scale. Additionally, we asked the main caregiver to indicate whether they believe they know how to read to or play with their child, again using a 7-point likert scale. As harsh parental disciplinary practices are found to be strongly associated with social-emotional problems in children (Baumrind, 1966; Kiernan and Huerta, 2008), we add questions about how parents exert control over their offspring in both survey waves. Parental disciplinary practices are measured on a 4-point likert scale by asking parents how often they used certain disciplinary techniques (such as *yelling* or *spanking*).

Information on compliance – including whether the weekly parenting sessions took place and, if not, the reason they did not take place – as well as details of the interaction were collected on a monthly basis from caregivers and on a weekly basis from parenting trainers through telephone interviews. In our analysis, we use parenting trainer reports as these data are more complete. The difference in average compliance for these two measures is insignificant and the two measures are highly correlated (correlation of 0.69).

[13]The non-parametric method is described further in the Web Appendix B.4. of Attanasio et al. (2015).

3.4 Summary Statistics, Balance, and Attrition

Summary statistics and tests for balance across control and treatment groups are shown in Table 1. Differences between study arms in individual child and caregiver characteristics are insignificant. A joint significance test across all baseline characteristics also confirms the study arms are balanced.[14] Appendix Table A1 shows that characteristics of untreated children in treatment villages (the ”spillover group”) are also balanced with those of children in the treatment and control groups.

Children in our sample are on average just over 24 months old at the start of the program as reported in Panel A in Table 1. Less than 5% of children are born with low birth weight. A large part of the children in our sample are first born in the family (60%). More than 80% of children were breastfed and more than 30% were breastfed for more than one year. More than 20% percent of sample children were anemic according to the WHO-defined threshold of 110 g/L. On average children were reported to be ill 4 days over the previous month.[15] At baseline we find that around 40 percent of the sample is cognitively delayed with Bayley MDI scores below 80 points. Fewer children (10%) are delayed in their motor development. We further find that around 30 percent of the children are at risk of social-emotional problems at baseline.

Around 26 percent of the sample receives social security support through the *dibao*, China’s minimum living standard guarantee program, as reported in Panel B of Table 1. The biological mother is the primary caregiver in only 60 percent of households, with grandmothers often taking over child rearing when mothers out-migrate to join the labor force in larger cities. We find that slightly more than 70 percent of primary caregivers in the sample (mothers or grandmothers as appropriate) have at least 9 years of formal schooling. On average households report being somewhat indifferent in their feelings toward the FPC at baseline.[16]

Baseline statistics on parental inputs shown in Panel C of Table 1 show that caregivers engage in few stimulating activities with their children. Only 11% of caregivers told a story to their child the previous day. Less than 5% read a book to their child (on average households have only 1.6 books). Only around 1 in 3 caregivers report playing or singing to their child the previous day.

Overall attrition between November 2014 and May 2015 was less than 1 percent and

[14] We test this by regressing treatment status on all baseline characteristics reported in Table 1 and test that the coefficients on all characteristics were jointly zero. The p-value of this test is 0.529.

[15] Caregivers were asked whether the child had suffered from fever, cough, diarrhea, indigestion or respiratory cold over the previous month.

[16] We asked caregivers to rate their perception of local Family Planning Commission on a 5-point scale (1 *very much like*; 2 *like*; 3 *Neither like nor dislike*; 4 *dislike*; 5 *very much dislike*).

insignificantly correlated with treatment status. We define attrition as missing a cognitive development measurement at endline for children with a BSID baseline measure.

3.5 Empirical Strategy

Given random assignment of households into treatment and control groups, comparison of outcome variable means across treatment arms provides unbiased estimates of the effect of the parenting intervention on outcomes. However, to increase power (and to account for our stratified randomization procedure) we condition our estimates on randomization strata (Bruhn and McKenzie, 2009). We also show estimates from specifications that additionally control for baseline values of the outcome variables in each regression.

We use ordinary least-squares (OLS) to estimate the effects of the parenting intervention with the following specification:

$$Y_{ij} = \alpha_1 + T_j'\beta_1 + X_{ij}'\gamma_1 + \tau_s + \epsilon_{ij1} \tag{1}$$

where Y_{ij} is the outcome for child i in village j; T_j is a vector of dummy variables indicating the treatment assignment of village j; X_{ij} is a vector of control variables and τ_s is a set of strata fixed effects. We adjust standard errors for clustering at the village level using the cluster-corrected Huber-White estimator. We use Equation (1) to estimate effects on primary development outcomes (cognitive development, motor development, and social-emotional problems). To estimate spillover effects we use the same specification but replace treated children with untreated children in treatment villages in the estimation sample.

To examine the mechanism of the parenting intervention, we estimate effects on secondary outcomes using a similar specification:

$$I_{ij} = \alpha_2 + T_j'\beta_2 + X_{ij}'\gamma_2 + \tau_s + \epsilon_{ij2} \tag{2}$$

where I_{ij} is a parenting measure for child i in village j. Our set of parenting measures includes parental investment, beliefs about parenting and disciplinary practices.

Because we estimate treatment effects on multiple outcomes, we present p-values adjusted for multiple hypothesis testing using the step-down procedure of Romano and Wolf (2005) which controls for the familywise error rate (FWER). We adjust across the three main outcomes (cognitive, motor, and social-emotional) and across secondary outcome variables within each domain (investments, beliefs, and disciplinary practices).

To further explore mechanisms, we conduct a mediation analysis following Imai et al.

(2010a,b). The objective of this analysis is to estimate how much of the total treatment effect is mediated through effects on each domain of parenting practice (investments, beliefs, and disciplinary practices). Following the parametric two-step algorithm described in Imai et al. (2010a),[17] we simultaneously estimate Equation (2) and an augmented outcome equation (control directly for endline parenting practices in Equation 1):

$$Y_{ij} = \alpha_3 + T_j'\beta_3 + X_{ij}'\gamma_3 + I_{ij}'\zeta + \tau_s + \epsilon_{ij3} \quad (3)$$

We construct three aggregate measures of parenting practices, I_{ijc}, by extracting the first factor from our secondary outcome measures[18] and standardize these by the distribution of the control group.

Assuming "sequential ignorability" (SI), the average treatment effect can be decomposed into the the average causal mediation effect (ACME) and the average direct effect (ADE). The ACME is obtained by the product of estimated coefficients, $\hat{\beta_2}\hat{\zeta}$, and tells us what proportion of the average treatment effect of the parenting intervention on child development is transmitted by changes in various dimensions of parenting practices (investments, beliefs about child-rearing and disciplinary practices) induced by the intervention. The ADE equals $\hat{\beta_3}$ and captures the remaining proportion of the average treatment effect resulting from all other possible causal mechanisms such as the weekly trainer-child interactions or independent learning activities of the infants.

Our randomization design ensures that the first ignorability assumption is satisfied.[19] The second part of the SI assumption, however, may not be satisfied if there exist confounding factors that affect child development and parenting practices simultaneously that we fail to control for in the mediation analysis. When estimating Equations (2) and (3) for the causal mediation analysis, we therefore extend X_{ij} to include a larger set of pre-treatment measures of child and household characteristics. In addition to strata fixed effects and baseline levels of cognitive skills, we also control for baseline hemoglobin concentration, child gender, whether child is firstborn, whether child had low birth weight,

[17]We use the mediation package in Stata (Hicks et al., 2011). For theoretical proof we refer to Appendix D and E in Imai et al. (2010a)

[18]Exploratory factor analysis indicates that the dimensionality of the measures of each of the three families of secondary outcomes equals one. We determined the number of relevant factors using Horn's parallel analysis(Horn, 1965). For each family of secondary outcomes we extract the first principal component from the several observed measures. The parental investment factor is derived from variable (1)-(5) in Table 3a; the parental beliefs factor is derived from variable (1)-(4) in Table 3b; the parental disciplining factor is derived from variable (1)-(5) in Table 3c. Factors are rotated using the direct quartimin oblique rotation method.

[19]The first ignorability assumption states that treatment assignment is assumed to be statistically independent of potential outcomes and potential mediators.

the primary carer's educational attainment, number of months the mother stayed at home after child birth, whether household is a welfare recipient under the *dibao* program, and a household asset index. [20]

Following the mediation analysis, we estimate average treatment effects within sub-groups defined by pre-treatment characteristics. To do so, we create dummy variables indicating treatment for each subgroup and include these, along with dummy variables indicating subgroups, in Equation (1). A t-test is used to test whether estimated treatment effects differ significantly across sub-groups.

In addition to the main intention-to-treat results, we also present estimates of the dose-response relationship between the number of completed household visits and cognitive development outcomes using control function methods. We do this first assuming a linear relationship and then allowing for a concave relationship by adding a squared term for household visits completed. For both of these, we instrument the number of visits with the treatment assignment, the distance between the village and the FPC township office, and the interaction between these two variables.[21]

4 Results

4.1 Short-term Average Treatment Effects on Development Outcomes

Table 2 shows the average treatment effect of the parenting intervention on measures of child development at the end of the 6-month intervention. We find that the parenting intervention program significantly increased cognitive development after 6 months as measured by the cognitive development score. The household visits by the parenting trainers led to an average increase of 0.27 standard deviations in cognitive development. We find no significant program effects on child psychomotor development or on social-emotional development. Estimates for each cohort separately using all sub-indexes of both the BSID and GMDS-ER are shown in Appendix Table A2.

These results are similar to the findings of Attanasio et al. (2014), who report that their home-based parenting intervention in Colombia led to an increase of 0.26 standard

[20]We further test the robustness of our results to violations of the SI assumption by running a sensitivity analysis as described in Imai et al 2010b. Denoting the correlation between the error terms of the mediation Equation (2) and the outcome Equation (3) as $\rho = corr(\epsilon_{ij2}; \epsilon_{ij3})$, in the presence of unobserved confounding factors that both affect child cognitive development and parenting practices $\rho \neq 0$, violating the key identifying assumption. We can express the ACME as a function of ρ which allows us to analyze for which value of ρ the ACME becomes insignificant.

[21]Linear estimates of the dose-response relationship between the number of completed household visits and cognitive development outcomes are similar when instrumenting compliance with only treatment assignment.

deviations in cognitive development but no significant improvement in psychomotor development. Despite a very similar average treatment effect, the Colombia study lasted one year longer than our study (18 months in total) and enrolled younger children (12-24 months). The integrated parenting intervention in Pakistan (Yousafzai et al., 2014) delivered ECD services with lower intensity (monthly instead of weekly household visits) but over a more extended period of 2 year. They report larger effect sizes on both cognitive development (0.6 standard deviations) and motor developments (0.5 standard deviations) at the end of the intervention.

Estimated effects on untreated children in treatment villages are shown in Appendix Table A3. None of these estimates reach statistical significance (possibly due to limited power), though coefficients are in the same direction as estimated impacts in the treatment group.

4.2 Mechanisms

In Table 3a we report average treatment effects of the early childhood intervention on measures of parental investment. We find that caregivers in the treatment group significantly increased the time they spend with their children actively engaging in developmentally-friendly activities. As a result of the weekly home visits, caregivers spend more time reading, story telling and singing to their children. Households in the treatment group also had more books in their home at the end of the program compared to households in the control group. Our aggregate parental investment measure indicates that the parenting program improved parental investment with 0.73 of a standard deviation.

We also find that the intervention significantly affected caregiver beliefs about child-rearing (Table 3b). Households receiving the parenting intervention reported a stronger belief in the importance of reading for child development after completion of the parenting intervention. Caregivers in the treatment group also reported that they felt they knew better how to read to their child, possibly indicating they were more confident in their child-rearing abilities after the intervention. The intervention did not, however, enhance beliefs about the importance of play, nor did parents feel they knew better how to play with their child. Effects on disciplinary practices are reported in Table 3c. Caregivers in the treatment group were less likely to report yelling or spanking their children – practices that are deemed harmful to children's social-emotional development.

The mediation analysis (Table 4) suggests that a substantial proportion of the average treatment effect of the program is mediated through a treatment-induced increase in parental time investment (column 1). More specifically, we find that the ACME of parental investment is estimated at 0.067 of a standard deviation and is significantly different from zero. This implies that about 25% of the average treatment effect of the program on child cognitive development resulted from increased parental engagement rather than the weekly household visits of the parenting trainers. Despite our attempt to control for a large set of pre-treatment covariates it is still possible that an unobserved factor affects both child development and the parental investment decision. The sensitivity analysis indicates that the ACME would be zero if there were to be a correlation of the error terms of the mediation Equation (2) and the outcome Equation (3) of about 0.11. This implies that a modest violation of the SI assumption would render the estimated mediation effect insignificant. However, concern that the SI assumption is not satisfied should be reduced given that we control directly for a large set of pre-treatment factors that are most likely to confound the relationship between parental investment and child cognitive outcomes - such as maternal education, socio-economic status and baseline levels of parental investment, child cognitive skills and child health.

Despite the positive effect of the intervention on parental beliefs about childrearing and parental disciplining behavior we find no evidence that effects on these dimensions of parenting were responsible for raising cognitive outcomes for children. The ACME of parental beliefs and parental disciplining are found to be close to zero and statistically insignificant, suggesting that these factors did not mediate the effects of the intervention on child outcomes independently of mediation through parental investment. Our finding that increased investment on the part of parents is an important factor in the success of the early childhood intervention is in line with Attanasio et al. (2015) who find a similar result in an early childhood intervention among disadvantaged children in Colombia.

4.3 Heterogeneous Treatment Effects

The subgroup analysis is presented in Table 5. Point estimates of treatment effects are larger for male children, for children with lower MDI scores at baseline, for children with mothers who had not migrated before they were two years of age, and for children with more highly educated mothers. Average treatment effects do not, however, differ significantly at the 10 percent level or lower across any of the subgroups examined, which could be due to insufficient power for the sub-sample analysis. Table A2 in the appendix shows average treatment effects for both age cohorts separately. We find treatment effects to be slightly higher for the older age cohort (children above 30 months old at study completion) but again this difference in not significantly different.

4.4 Compliance

On average, 16.4 visits (out of 24 total planned visits) were completed for each household during the course of the study based on reports from parent trainers. To assess the drivers of incomplete compliance, we regress the number of reported household visits on child, family, and trainer characteristics as well as the distance from the village to the closest FPC office (Table 6).

Compliance is most strongly correlated with three factors: whether the child is male, distance from the village to the FPC office in the township, and caregiver perception of the FPC. Male children receive on average slightly more household visits, possibly a reflection of male preference in rural China. Compliance is negatively correlated with the distance to the FPC office, which may reflect supply-side compliance failure as parenting trainers chose to visit remote households less frequently, though it may also be capturing correlated demand-side characteristics of households.

Once all variables are included in the compliance regression, the most important demand-side factor associated with compliance appears to be whether households had an unfavorable view of the FPC at baseline. Households with a more unfavorable view of the agency completed significantly fewer visits, decreasing by almost 2 visits for each point in their likert scale rating. If the program were to be implemented in the future, however, this may become less of an obstacle to implementation as we find that the program itself has a significant positive effect on public perception of the FPC (Table 3b).

4.5 Dose-Response

Table 7 shows control function estimates of the dose-response function. Column (1) shows estimates assuming a linear relationship between the number of household visits and cognitive child development. We estimate that each session completed increases cognitive scores by 0.015 standard deviations. Results in Column (2) which allow for non-linearity do not suggest that this relationship is concave. Assuming a linear relationship up to 24 visits, estimates suggest that full compliance would increase cognitive scores by 0.36 standard deviations on average (approximately 0.1 SD higher than the intention-to-treat estimate).

5 Conclusion

This paper reports the results of a randomized trial of a home-based parenting program delivered by cadres employed by China's Family Planning Commission. We find that the program significantly increased the cognitive outcomes of children after only six months. There were no significant effects on motor development or social-emotional outcomes. The program also had corresponding positive effects on measures of parental investment, led to significant changes in caregiver beliefs about parenting, and improved parental disciplining practices. These effects occurred despite lackluster compliance with the program which appears to have been driven primarily by a combination of supply-side implementation failures and an unfavorable perception of the FPC by beneficiary households. The program itself, however, had a positive effect on views of the FPC suggesting that public perception may be a less significant obstacle as the program is implemented over time. Efforts to improve supply-side compliance will likely have the greatest impact on improving program effectiveness.

Our study faces a number of limitations. First, the study took place in one poor rural area in Northwest China; results may differ in other regions. Second, children were already over 18 months of age at the start of the trial. It is possible that effects would be larger if children were enrolled at an earlier age and/or the intervention took place over a longer period of time. Finally, we estimate effects only at one point in time at the conclusion of the intervention. Longer-run follow-up of the children in the study will be necessary to determine if the gains we find are lasting or fade out over time.

Despite these limitations, our results imply that an ECD program can be effectively delivered through the existing infrastructure of the National Health and Family Planning Commission. Future research should explore alternative interventions to improve ECD outcomes and compare relative cost-effectiveness across alternative delivery models.

References

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2015). Estimating the production function for human capital: Results from a randomized control trial in colombia. Discussion paper, National Bureau of Economic Research.

Attanasio, O. P., Fernández, C., Fitzsimons, E. O., Grantham-McGregor, S. M., Meghir, C., and Rubio-Codina, M. (2014). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in colombia: cluster randomized controlled trial. *BMJ*, 349:g5785.

Baumrind, D. (1966). Effects of authoritative parental control on child behavior. *Child development*, pages 887–907.

Bayley, N. (1969). *Manual for the Bayley scales of infant development*. Psychological Corporation.

Berlinski, S., Schady, N., et al. (2016). *The early years: Child well-being and the role of public policy*. Springer.

Black, M. M. and Dewey, K. G. (2014). Promoting equity through integrated early child development and nutrition interventions. *Annals of the New York Academy of Sciences*, 1308(1):1–10.

Black, M. M., Walker, S. P., Fernald, L. C., Andersen, C. T., DiGirolamo, A. M., Lu, C., McCoy, D. C., Fink, G., Shawar, Y. R., Shiffman, J., et al. (2017). Early childhood development coming of age: science through the life course. *The Lancet*, 389(10064):77–90.

Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, 1(4):200–232.

Carneiro, P. M. and Heckman, J. J. (2003). Human capital policy.

Chan, M. (2013). Linking child survival and child development for health, equity, and sustainable development. *Lancet*, 381(9877):1514–1515.

Chang, S., Zeng, L., Brouwer, I. D., Kok, F. J., and Yan, H. (2013). Effect of iron deficiency anemia in pregnancy on child mental development in rural china. *Pediatrics*, 131(3):e755–e763.

Cirelli, I., Graz, M. B., and Tolsa, J.-F. (2015). Comparison of griffiths-ii and bayley-ii tests for the developmental assessment of high-risk infants. *Infant Behavior and Development*, 41:17–25.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2):31–47.

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

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of human resources*, pages 696–727.

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S. M., and Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in jamaica. *Science*, 344(6187):998–1001.

Grantham-McGregor, S. M., Powell, C. A., Walker, S. P., and Himes, J. H. (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the jamaican study. *The Lancet*, 338(8758):1–5.

Greenhalgh, S. (1986). Shifts in china's population policy, 1984-86: Views from the central, provincial, and local levels. *Population and Development Review*, pages 491–515.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavitz, A. (2010). The rate of return to the highscope perry preschool program. *Journal of public Economics*, 94(1):114–128.

Hesketh, T., Lu, L., and Xing, Z. W. (2005). The effect of china's one-child family policy after 25 years.

Hesketh, T., Zhou, X., and Wang, Y. (2015). The end of the one-child policy: lasting implications for china. *Jama*, 314(24):2619–2620.

Hicks, R., Tingley, D., et al. (2011). Causal mediation analysis. *Stata Journal*, 11(4):605.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2):179–185.

Huang, H., Tao, S., Zhang, Y., et al. (1993). Standardization of bayley scales of infant development in shanghai. *Chin J Child Health*, 1(3):158–160.

Imai, K., Keele, L., and Tingley, D. (2010a). A general approach to causal mediation analysis. *Psychological methods*, 15(4):309.

Imai, K., Keele, L., Yamamoto, T., et al. (2010b). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical science*, 25(1):51–71.

Kiernan, K. E. and Huerta, M. C. (2008). Economic deprivation, maternal depression, parenting and children's cognitive and emotional development in early childhood1. *The British journal of sociology*, 59(4):783–806.

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

Li, Q., Yan, H., Zeng, L., Cheng, Y., Liang, W., Dang, S., Wang, Q., and Tsuji, I. (2009). Effects of maternal multimicronutrient supplementation on the mental development of infants in rural western china: follow-up evaluation of a double-blind, randomized, controlled trial. *Pediatrics*, 123(4):e685–e692.

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. *The Lancet Global Health*, 4(12):e916–e922.

Luiz, D., Barnard, A., Knoesen, N., Kotras, N., Horrocks, S., McAlinden, P., Challis, D., and O'Connell, R. (2006). Griffiths mental development scales: Extended revised. two to eight years. administration manual. *Hogrefe, Oxford, UK*.

Nations, U. (2015). Transforming our world: The 2030 agenda for sustainable development. *New York: United Nations, Department of Economic and Social Affairs*.

Nelson, C. A. and Sheridan, M. A. (2011). Lessons from neuroscience research for understanding causal links between family and neighborhood characteristics and educational outcomes. *Whither opportunity*, pages 27–46.

Richter, L. M., Daelmans, B., Lombardi, J., Heymann, J., Boo, F. L., Behrman, J. R., Lu, C., Lucas, J. E., Perez-Escamilla, R., Dua, T., et al. (2017). Investing in the foundation of sustainable development: pathways to scale up for early childhood development. *The Lancet*, 389(10064):103–118.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Squires, J., Bricker, D., and Twombly, E. (2003). The asq: Se user's guide for the ages & stages questionnaires, social-emotional: A parent completed, child-monitoring system for social-emotional behaviors. *Baltimore: Paul H. Brookes Publishing Co*.

Walker, S. P., Chang, S. M., Vera-Hernández, M., and Grantham-McGregor, S. (2011). Early childhood stimulation benefits adult competence and reduces violent behavior. *Pediatrics*, 127(5):849–857.

Wilson, J. Q. (1989). *Bureaucracy: What government agencies do and why they do it*. Basic Books.

Wu, K. B., Young, M. E., and Cai, J. (2012). *Early child development in China: Breaking the cycle of poverty and improving future competitiveness*. World Bank Publications.

Wu, W., Sheng, D., Shao, J., and Zhao, Z. (2011). Mental and motor development and psychosocial adjustment of chinese children with phenylketonuria. *Journal of paediatrics and child health*, 47(7):441–447.

Yi, S. (1995). Manual of bayley scales of infant development, chinese revision. xiangya school of medicine.

Yi, S., Luo, X., Yang, Z., and Wan, G. (1993). The revising of bayley scales of infant development (bsid) in china. *Chin J Clin Psychol*, 1:71–5.

Yousafzai, A. K., Obradović, J., Rasheed, M. A., Rizvi, A., Portilla, X. A., Tirado-Strayer, N., Siyal, S., and Memon, U. (2016). Effects of responsive stimulation and nutrition interventions on children's development and growth at age 4 years in a disadvantaged population in pakistan: a longitudinal follow-up of a cluster-randomised factorial effectiveness trial. *The Lancet Global Health*, 4(8):e548–e558.

Yousafzai, A. K., Rasheed, M. A., Rizvi, A., Armstrong, R., and Bhutta, Z. A. (2014). Effect of integrated responsive stimulation and nutrition interventions in the lady health worker programme in pakistan on child development, growth, and health outcomes: a cluster-randomised factorial effectiveness trial. *The Lancet*, 384(9950):1282–1293.

Table 1: Descriptive Statistics and Balance

	(1) Control (N=301)	(2) Treatment (N=212)	(3) p-value
Panel A. Child Characteristics			
(1) Age in months	24.46	24.45	0.747
	(0.20)	(0.22)	
(2) Male	0.45	0.51	0.185
	(0.03)	(0.04)	
(3) Low birth weight	0.04	0.04	0.774
	(0.01)	(0.01)	
(4) First born	0.59	0.61	0.366
	(0.03)	(0.04)	
(5) Ever breastfed	0.85	0.87	0.974
	(0.03)	(0.04)	
(6) Breastfed $\geq$ 12 months	0.35	0.39	0.867
	(0.05)	(0.05)	
(7) Anemia (Hb <110 g/L)	0.23	0.27	0.849
	(0.03)	(0.04)	
(8) Days ill past month	4.32	4.55	0.620
	(0.33)	(0.37)	
(9) Cognitive Delay (BSID MDI<80)	0.46	0.39	0.206
	(0.04)	(0.03)	
(10) Motor Delay (BSID PDI<80)	0.12	0.10	0.476
	(0.02)	(0.02)	
(11) Social-Emotional Problems (ASQ:SE>60)	0.25	0.28	0.401
	(0.03)	(0.03)	
Panel B. Household Characteristics			
(1) Social security support recipient	0.28	0.25	0.832
	(0.03)	(0.03)	
(2) Mother at home	0.68	0.62	0.116
	(0.04)	(0.05)	
(3) Caregiver education $\geq$ 9 years	0.72	0.74	0.487
	(0.03)	(0.04)	
(4) Unfavourable perception of FPC	2.87	2.85	0.824
	(0.06)	(0.05)	
Panel C. Parental Inputs			
(1) Told story to child yesterday	0.11	0.11	0.960
	(0.02)	(0.02)	
(2) Read book to child yesterday	0.05	0.04	0.872
	(0.01)	(0.01)	
(3) Sang song to child yesterday	0.37	0.35	0.651
	(0.03)	(0.04)	
(4) Played with child yesterday	0.34	0.34	0.996
	(0.03)	(0.03)	
(5) Number of books in household	1.60	1.90	0.615
	(0.24)	(0.29)	

P-values account for clustering within villages. Unfavourable perception of FPC is measured on a 5-point likert scale.

Table 2: Average Treatment Effects on Early Childhood Development Outcomes

	Cognitive Development		Motor Development		Social-Emotional Problems	
	(1) Cognitive Score	(2) Cognitive Score	(3) Motor Score	(4) Motor Score	(5) ASQ:SE > 60	(6) ASQ:SE > 60
Treatment Group	0.245***	0.271***	-0.009	0.006	-0.023	-0.017
	(0.090)	(0.091)	(0.076)	(0.076)	(0.043)	(0.043)
	{0.005}	{0.002}	{0.889}	{0.941}	{0.357}	{0.465}
Baseline Control		✓		✓		✓
Observations	508	503	508	503	508	508
R^2	0.04	0.14	0.03	0.15	0.05	0.09

In all regressions we control for strata fixed effects. Cognitive development for children under 30 months was assessed using the Mental Development Index (MDI) from the Bayley Scales of Infant Development (BSID) whereas older children were assessed using the Performance subscale from the Griffith Mental Development Scales (GMDS-ER 2-8). Motor development was measured using the BSID-I Psychomotor Development Index (PDI) for children younger than 30 months and the Locomotor subscale from the Griffith Mental Development Scales for the older children. Both development indexes are non-parametrically standardized for each age group (measured in months). Social-emotional problems are represented by a dummy variable indicating a ASQ:SE score higher then 60. Additional controls indicate we condition our estimate on baseline MDI score (column 2), PDI score (column 4) or ASQ:SE score (column 6). Estimates of treatment effects on social-emotional development are additionally conditioned on child gender and age (columns 5 and 6). All standard errors are clustered at the village level. Significance stars indicate significance after adjusting for multiple hypotheses using the step-down procedure of Romano Wolf (2005) to control for the familywise error rate (FWER). P-values accounting for multiple hypotheses are reported in curly brackets.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table 3a: Average Treatment Effect on Parental Investments

	(1) Play with toys	(2) Story-telling	(3) Reading books	(4) Singing songs	(5) Number of books	(6) **Parental investment**
Treatment Group	0.166***	0.201***	0.201***	0.128***	1.297***	0.727***
	(0.043)	(0.038)	(0.031)	(0.044)	(0.374)	(0.103)
	{0.001}	{0.001}	{0.001}	{0.020}	{0.001}	{0.001}
Observations	475	475	472	475	458	453
R^2	0.10	0.13	0.12	0.11	0.52	0.31
Control mean	0.341	0.144	0.064	0.401	2.381	0.000

In all regressions we control for baseline parental inputs and strata fixed effects. We measure parental time inputs by asking whether the main caregiver engaged in a set of child-rearing activities yesterday. The aggregate parental investment measure is constructed as the first factor from outcome variables of column (1)-(5) using the polychoric correlation matrix. The factor is standardized by the distribution of the control group. All standard errors account for clustering at the village level. Significance stars indicate significance after adjusting for multiple hypotheses using the step-down procedure of Romano and Wolf (2005) to control for the familywise error rate (FWER). P-values accounting for multiple hypotheses are reported in curly brackets.

$*p < 0.10, **p < 0.05, ***p < 0.01$.

Table 3b: Average Treatment Effects on Parental Beliefs

	(1) Belief in importance of reading	(2) Belief in importance of play	(3) Belief in own ability to read to child	(4) Belief in own ability to play with child	(5) **Parental beliefs**	(6) Unfavorable perception FPC
Treatment Group	0.432***	0.085	0.595***	0.142	0.337***	-0.334**
	(0.117)	(0.097)	(0.168)	(0.148)	(0.071)	(0.093)
	{0.001}	{0.257}	{0.001}	{0.280}	{0.001}	{0.007}
Observations	472	474	473	473	471	474
R^2	0.05	0.02	0.08	0.06	0.08	0.08
Control Mean	4.408	5.083	3.983	4.780	-0.000	3.797

In all regressions we control for baseline parental beliefs and strata fixed effects. Parental beliefs about the importance of reading and play are measured by asking the main caregiver how they feel about each behavior using a 7-point likert scale. The same 7-point likert scale is used to assess whether parents believe they know how to read and play with their child. The aggregate parental beliefs measure is constructed as the first factor from outcome variables of column (1)-(5).The factor is standardized by the distribution of the control group. Unfavourable perception about the Family Planning Commision was measured using a 5-point likert scale. All standard errors account for clustering at the village level. Significance stars indicate significance after adjusting for multiple hypotheses using the step-down procedure of Romano and Wolf (2005) to control for the familywise error rate (FWER). P-values accounting for multiple hypotheses are reported in curly brackets.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table 3c: Average Treatment Effects on Parental Disciplinary Practices

	(1) Use yelling to discipline	(2) Use spanking to discipline	(3) Take away toy to discipline	(4) Explain unreasonable behaviour to discipline	(5) **Parental disciplining**
Treatment	-0.171***	-0.159**	-0.086	0.081	-0.237**
	(0.066)	(0.076)	(0.089)	(0.072)	(0.099)
	{0.005}	{0.030}	{0.312}	{0.701}	{0.011}
Observations	474	475	473	471	465
R^2	0.09	0.08	0.05	0.04	0.09
Control Mean	3.077	2.823	2.181	3.262	0.000

In all regressions we control for baseline parental disciplinary techniques and strata fixed effects. Parental disciplinary behavior is measured by asking parents how often they use a certain disciplinary technique using a 4-point likert scale. The aggregate parental disciplining measure is constructed as the first factor from outcome variables of column (1)-(5). The factor is standardized by the distribution of the control group. All standard errors account for clustering at the village level. Significance stars indicate significance after adjusting for multiple hypotheses using the step-down procedure of Romano and Wolf (2005) to control for the familywise error rate (FWER). P-values accounting for multiple hypotheses are reported in curly brackets.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table 4: Mediation of Program Treatment Effects

	(1) I_ijc = Parental investment factor	(2) I_ijc = Parental belief factor	(3) I_ijc = Parental disciplining factor
Average causal mediation effect	.067	0.021	-0.001
	[0.018, 0.142]	[-0.011, 0.058]	[-0.026, 0.021]
Average direct effect	0.194	0.226	0.240
	[0.002, 0.378]	[0.042, 0.400]	[0.057, 0.415]
Average total effect	0.261	0.247	0.239
	[0.072, 0.433]	[0.065, 0.416]	[0.062, 0.408]
% Average total effect mediated	0.251	0.085	-0.006
	[0.153, 0.832]	[0.051, 0.323]	[-0.020, -0.003]
Observations	431	446	440
ρ at which ACME=0	0.108	0.059	0.002

Outcome is cognitive score as in Table 1 column (1) and (2). In square brackets are 95% percentile confidence intervals. The model for both the outcome and mediator is a linear regression model. In both models we control for baseline levels of child cognitive skills and hemoglobin, child gender, whether child is firstborn, whether child had low birth weight, carer's educational attainment, number of months carer stayed at home after child birth, whether household is a welfare recipient and the first factor of a set of household assets which proxies household wealth. We additionally condition our estimates further on strata fixed effects. ρ represents the correlation between the error terms of the mediation and outcome models.

Table 5: Average Treatment Effects on Cognitive Development by Sub-Groups

	(1) Cognitive Score	(2) Cognitive Score	(3) Cognitive Score	(4) Cognitive Score	(5) P-value on t-test
Child Gender					
Male	0.326**				0.462
	(0.131)				
Female	0.218*				
	(0.113)				
Child Cognition					
Low MDI		0.335**			0.381
		(0.123)			
High MDI		0.162			
		(0.122)			
Maternal Migration					
Mother at home < 2 years			0.199*		0.307
			(0.101)		
Mother at home ≥ 2 years			0.377**		
			(0.159)		
Caregiver Education					
Caregiver education < 9 years				0.042	0.131
				(0.174)	
Caregiver education ≥ 9 years				0.335***	
				(0.103)	
Observations	503	503	503	501	
R^2	0.14	0.070	0.14	0.10	

In all regressions we control for baseline MDI score and strata fixed effects. Low MDI is a dummy indicating the child is in the bottom half of the baseline Bayley MDI distribution, high MDI indicates the child is in the top half of the baseline Bayley MDI distribution at baseline. Maternal migration is measured by a dummy variable indicating that the biological mother did not outmigrate to a different city in the first two years after child birth. Column (5) reports p-values of a t-tests of equality of treatment effects for different sub-groups. All standard errors are clustered at the village level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table 6: Determinants of Compliance

	(1) HH Visits	(2) HH Visits	(3) HH Visits	(4) HH Visits	(5) HH Visits
Male	1.433*	1.708**	1.701**	1.619*	1.271
	(0.813)	(0.836)	(0.829)	(0.845)	(0.800)
Age in months	-0.093	-0.058	-0.054	-0.026	-0.032
	(0.121)	(0.113)	(0.112)	(0.116)	(0.094)
Low ability	1.483	1.570*	1.447	1.278	1.206
	(0.957)	(0.931)	(0.874)	(0.934)	(0.971)
Number of days ill	0.120	0.072	0.074	0.058	-0.039
	(0.133)	(0.134)	(0.133)	(0.132)	(0.126)
Mother home > 2 years		0.191	0.244	-0.172	-0.660
		(1.018)	(0.973)	(0.970)	(0.873)
Maternal education > 9 year		0.976	0.762	0.931	0.391
		(0.979)	(0.951)	(0.909)	(0.968)
Social security support recipient		-1.434	-1.717	-1.697	-1.343
		(1.041)	(1.034)	(1.084)	(1.091)
Distance to FPC office		-0.311***	-0.314***	-0.318***	-0.288**
		(0.112)	(0.112)	(0.116)	(0.110)
Unfavourable perception of FPC			-1.392***	-1.508***	-1.786***
			(0.473)	(0.479)	(0.465)
Trainer is male				-1.066	-1.551
				(1.384)	(1.355)
Trainer work experience FPC				0.136	0.161
				(0.113)	(0.109)
Trainer has bachelor degree				0.135	-0.343
				(1.453)	(1.049)
County FE					✓
Observations	211	211	211	211	211
R^2	0.03	0.11	0.14	0.16	0.27

Low MDI is a dummy indicating the child is in the bottom half of the Bayley MDI distribution at baseline. Mother at home is a dummy variable indicating that the biological mother did not outmigrate to a different city in the first two years after child birth. Unfavourable perception of Family Planning Commission is measured on a 5-point likert scale. Trainer work experience is measured by the number of years worked as cadre for the Family Planning Commission. All standard errors are clustered at the village level.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table 7: Dose-Response Relationship

	(1) Cognitive Score	(2) Cognitive Score
Number of HH Visits	0.015***	0.014
	(0.006)	(0.038)
Number of HH Visits^2		0.000
		(0.002)
Observations	472	472
R^2	0.09	0.09

Column (1) gives a control function estimate of the treatment effect of one household visit, assuming a linear relationship between the number of household visits and the child's cognitive development. Column (2) gives control function estimates of the treatment effect of on household visit, assuming a concave relationship. Residuals used in the control function estimation are derived from regressing the number of household visits on treatment status, distance to the Family Planning Commission office, perception of Family Planning Commission and the interaction of both the distance and perception measure with treatment status. In all regressions we control for baseline cognitive development score and strata fixed effects. All standard errors are clustered at the village level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table 8: Trainer Summary Statistics (N=69)

Variable	**Mean**	**Std. Dev.**
Male	0.609	0.492
Age	33.8	5.8
Married	0.884	0.323
Has child	0.841	0.369
Age of youngest child	6.731	6.074
Has bachelor degree	0.304	0.464
Monthly Salary (RMB)	3229.2	477.7
Work experience FPC (years)	11.681	6.895

Table A1: Descriptive Statistics and Balance

	(1) Control (N=300)	(2) Treatment (N=212)	(3) Spillover (N=79)	(4) P-value Control vs. Treatment	(5) P-value Control vs. Spillover	(6) P-value Treatment vs. Spillover
Panel A. Child Characteristics						
(1) Age in months	24.47	24.45	24.38	0.800	0.689	0.856
	(0.20)	(0.22)	(0.33)			
(2) Male	0.45	0.51	0.58	0.196	0.017	0.110
	(0.03)	(0.034)	(0.05)			
(3) Low birth weight	0.04	0.04	0.05	0.777	0.683	0.808
	(0.01)	(0.01)	(0.03)			
(4) First born	0.58	0.61	0.66	0.358	0.427	0.985
	(0.03)	(0.04)	(0.06)			
(5) Ever breastfed	0.85	0.87	0.87	0.997	0.502	0.562
	(0.03)	(0.04)	(0.06)			
(6) Still breastfed ≥ 12 months	0.35	0.39	0.33	0.900	0.929	0.690
	(0.05)	(0.05)	(0.08)			
(7) Anemia (Hb <110 g/L)	0.23	0.27	0.16	0.846	0.871	0.831
	(0.03)	(0.04)	(0.05)			
(8) Days ill past month	4.32	4.55	4.77	0.702	0.995	0.725
	(0.34)	(0.37)	(0.84)			
(9) Cognitive Delay (BSID MDI<80)	0.46	0.39	0.36	0.243	0.846	0.611
	(0.04)	(0.03)	(0.08)			
(10) Motor Delay (BSID PDI<80)	0.12	0.10	0.13	0.568	0.832	0.476
	(0.023)	(0.024)	(0.055)			
(11) Social-Emotional Problems(ASQ:SE>60)	0.25	0.28	0.32	0.414	0.286	0.957
	(0.03)	(0.03)	(0.05)			
Panel B. Household Characteristics						
(1) Social security support recipient	0.28	0.25	0.29	0.852	0.895	0.910
	(0.03)	(0.03)	(0.06)			
(20 Mother at home	0.68	0.62	0.66	0.120	0.481	0.102
	(0.04)	(0.04)	(0.06)			
(3) Caregiver education ≥ 9 years	0.72	0.74	0.78	0.460	0.201	0.886
	(0.03)	(0.04)	(0.04)			
(4) Unfavourable perception of FPC	2.87	2.85	2.86	0.882	0.946	0.707
	(0.06)	(0.05)	(0.09)			
Panel C. Parental Inputs						
(1) Told story to child yesterday	0.11	0.11	0.09	0.985	0.694	0.330
	(0.02)	(0.02)	(0.04)			
(2) Read book to child yesterday	0.05	0.04	0.02	0.859	0.404	0.350
	(0.01)	(0.01)	(0.02)			
(3) Sang song to child yesterday	0.37	0.35	0.46	0.653	0.400	0.082
	(0.03)	(0.04)	(0.08)			
(4) Played with child yesterday	0.33	0.34	0.38	0.952	0.296	0.756
	(0.03)	(0.03)	(0.06)			
(5) Number of books in household	1.60	1.89	2.30	0.599	0.148	0.179
	(0.24)	(0.29)	(0.64)			

P-values account for clustering within villages.

Table A2: Average Treatment Effects on Early Childhood Development Outcomes

	Bayley Assessment		Griffith Assessment				
	(1) Cognitive	(2) Psychomotor	(3) Performance	(4) Personal-Social	(5) Language	(6) Locomotor	(7) Eye-hand coordination
Treatment Group	0.237**	-0.000	0.279***	0.278**	0.009	-0.013	0.110
	(0.112)	(0.117)	(0.105)	(0.107)	(0.107)	(0.113)	(0.114)
	{0.048}	{0.837}	{0.010}	{0.007}	{0.897}	{0.950}	{0.291}
Observations	226	226	277	277	277	276	277
R^2	0.31	0.21	0.25	0.25	0.27	0.30	0.17

In all regressions we control for baseline developmental score and strata fixed effects. Bayley assessment scores are taken from the Bayley Scales of Infant Development test and registered to all children younger than 30 months at endline. Griffith Assessment scores are from the GMDS-ER 2-8 test and registered to all children older than 30 months at endline. All assessment scores are non-parametrically standardized for each age group (measured in months). Standard errors account for clustering at the village level. Significance stars indicate significance after adjusting for multiple hypotheses using the step-down procedure of Romano and Wolf (2005) to control for the familywise error rate (FWER). P-values accounting for multiple hypotheses are reported between brackets.

$*p < 0.10, **p < 0.05, ***p < 0.01$.

Table A3: Average Treatment Effects on Early Childhood Development Outcomes of Non-treated Children in Treatment Villages

	Cognitive Development		Psychomotor Development		Socio-Emotional Problems	
	(1) Cognitive Score	(2) Cognitive Score	(3) Motor Score	(4) Motor Score	(5) ASQ:SE > 60	(6) ASQ:SE > 60
Spillover group	0.097	0.055	0.075	0.020	-0.071	-0.079
	(0.143)	(0.145)	(0.104)	(0.102)	(0.062)	(0.061)
	{0.420}	{0.433}	{0.491}	{0.523}	{0.992}	{0.389}
Baseline Control		✓		✓		✓
Observations	373	367	373	367	372	323
R^2	0.01	0.12	0.03	0.16	0.04	0.08

In all regressions we control for strata fixed effects. Cognitive development for children under 30 months was assessed using the Mental Development Index (MDI) from the Bayley Scales of Infant Development (BSID) whereas older children were assessed using the Performance subscale from the Griffith Mental Development Scales (GMDS-ER 2-8). Motor development was measured using the BSID-I Psychomotor Development Index (PDI) for children younger than 30 months and the Locomotor subscale from the Griffith Mental Development Scales for the older children. Both development indexes are non-parametrically standardized for each age group (measured in months). Social-emotional problems are represented by a dummy variable indicating a ASQ:SE score higher then 60. Additional controls indicate we condition our estimate on baseline MDI score (column 2), PDI score (column 4) or ASQ:SE score (column 6). Estimates of treatment effects on social-emotional development are additionally conditioned on child gender and age (columns 5 and 6). All standard errors are clustered at the village level. Significance stars indicate significance after adjusting for multiple hypotheses using the step-down procedure of Romano Wolf (2005) to control for the familywise error rate (FWER). P-values accounting for multiple hypotheses are reported in curly brackets.

$*p < 0.10, **p < 0.05, ***p < 0.01$.