Courant Research Centre

'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 276

Two Scalable Interventions to Promote Health and Mental Development in Early Childhood: A Randomized Controlled Trial in Rural India

Cara Ebert, Esther Heesemann, Sebastian Vollmer

First version: July 2020
This version: June 2023

Platz der Göttinger Sieben 5 · 37073 Goettingen · Germany
Phone: +49-(0)551-3921660 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

Two Scalable Interventions to Promote Health and Mental Development in Early Childhood: A Randomized Controlled Trial in Rural India*

Cara Ebert, RWI - Leibniz Institute for Economic Research

Esther Heesemann, University of Mannheim & Center for Evaluation and Development

Sebastian Vollmer, University of Goettingen

May 31, 2023

Abstract: Early childhood interventions are important for poverty reduction, but difficult to target. We study two interventions in rural India for children below two that are simple and easy to scale: a durable device for iron fortification at home and caregiver training in dialogic reading. Outcomes were measured blinded. When delivered together, the interventions improved receptive language skills of children who were not anemic at baseline, suggesting that health endowment complements parental investments in the production of human capital. The dialogic reading intervention further shows positive spillovers on untreated children. Overall, shorter-lived interventions are potentially effective in healthier populations.

Keywords: Early Childhood Development, Parental Investment, Nutrition, Health Behavior, Human Capital, India

JEL Codes: I12, I15, I24, J13, O12, O15

*Corresponding author: Cara Ebert, email: cara.ebert@rwi-eesen.de. Previous versions of the paper were titled "Encouraging parents to invest: A randomized trial with two simple interventions in early childhood". The trial is registered at the American Economic Association Registry (AEARCTR-0002696). Ethical clearance was received by the ethic committees of the Indian Institute of Technology Gandhinagar (IEC/2016-17/2/MS/025) and the University of Göttingen (no number). See the end of the manuscript for funding sources, acknowledgments and a conflict of interest statement.

I INTRODUCTION

Early disadvantages in children's health and mental functions have a lifetime impact as human capital accumulation is dynamic and path-dependent (Cunha and Heckman, 2007; Attanasio et al., 2020*b*; Attanasio, Meghir and Nix, 2020). For children living in poverty there are many factors, such as malnutrition and poor mental stimulation, which can impair their early development and lifetime human capital outcomes (Walker et al., 2007, 2011). An important policy intervention to alleviate the intergenerational transmission of poverty is to reduce the risk factors in early childhood development (Attanasio, 2015; García and Heckman, 2022).

A major challenge for effectively promoting early childhood development at scale in low- and middle-income countries is in targeting young children (Richter et al., 2017; Attanasio et al., 2022). Whereas slightly older children can be reached through preschools and schools, most young children are not enrolled in institutional daycare where interventions can reach them. Nutrition interventions delivered via daily school meals in India, for example, improved nutritional and educational outcomes effectively and at low cost (Afridi, 2010; Chakraborty and Jayaraman, 2019; Krämer, Kumar and Vollmer, 2021*b*).

At a small scale, several programs to improve early childhood development have been tested and implemented effectively in developing countries. A seminal and oft-replicated program with one-year-old children in Jamaica increased children's mental development in the short term, even earnings 20 years later as well as cognitive and psychosocial skils and behavior 30 years later (Grantham-McGregor et al., 1991; Walker et al., 1991; Gertler et al., 2014; Walker et al., 2022). The Jamaican program was very intensive and involved weekly home visits for parenting training over a two-year period, cross-randomized with the provision of nutrition supplementation via enriched formula. More recent study designs integrated similar programs into existing networks of community health workers, or organized group sessions to test the feasibility of such modes of delivery on a larger scale (Attanasio et al., 2014; Yousafzai et al., 2014, 2016; Andrew et al., 2020; Sylvia et al., 2020; Attanasio et al., 2022). Overall, little is known about how to scale effective early childhood interventions that are less intensive. The evidence that does exist is largely based on programs with weekly or bi-weekly meetings over a longer period of time.

Our study proposes an alternative early childhood development program that empowers parents to improve nutrition and stimulation of their children based on interventions that are remarkably simple but long lasting, and have shown to be effective in efficacy trials. The first intervention targets iron deficiency as a risk factor for early development. In one home visit, we distribute an iron cooking utensil, called Lucky Iron Leaf^{TM} (hereafter Lucky Iron Leaf), for home iron fortification. The Lucky Iron Leaf leaches iron when placed in boiling water with some fruit acid. The enriched water fortifies foods like rice and lentils during regular cooking. The intervention is simple to deliver to households because the Lucky Iron Leaf lasts as an iron source for up to five years and does not require replenishment as conventional fortification technologies do (e.g., multi-micronutrient powder). This is the first study to rigorously test how effective the fortification technology is in reducing the prevalence of anemia among children and in a large-scale intention-to-treat design. The

second intervention targets the lack of psychosocial stimulation as a risk factor of early development. During four home visits caregivers receive three durable picture books and learn methods of dialogic reading, which is an interactive mode of sharing picture books to stimulate speaking and learning. In addition, we nudge the regular exercise of dialogic reading using book sharing calendars as reminders in a random subset of dialogic reading intervention households. To our knowledge, this is the first trial which evaluates the impact of dialogic reading in a large-scale intention-to-treat design, and in India.

The program is much simpler than those previously tested and better suited to contexts where service delivery via public infrastructure is not an option, but it would also be less costly than the intensive programs if it were integrated into functional services. In many places where early childhood interventions are needed, the public infrastructure that might be used for service delivery is dysfunctional or resource constrained. For example, in Bihar, the Northeast Indian state where this study takes place, evidence shows that a quarter of Anganwadi centers – which offer take-home food rations, vaccinations and health check-ups for children below three, plus other services for older children – were closed during inspection visits and only 1.6 percent of children aged 0-71 months had a health check-up in a three month period (Fraker, Shah and Abraham, 2013; Government of India and UNICEF, 2015).

We combine the psychosocial stimulation intervention with the nutrition intervention, because of potential complementarities of parental investments with other inputs in the human capital production technology (Cunha and Heckman, 2007; Cunha, Heckman and Schennach, 2010). In a previous estimation of a production function of child development in India, it has been shown that good health complements parental investment in children (Attanasio, Meghir and Nix, 2020). In this paper, we provide a simple framework for the production technology of human capital that structures our ideas and understanding of how our interventions affect children's health and skills.

We use a factorial design, in which we assigned 1,480 households with 10- to 20-month old children in Madhepura district (Bihar, India) to one of four experimental arms: one stand-alone Lucky Iron Leaf arm, one stand-alone dialogic reading arm, one combined interventions arm, which received the Lucky Iron Leaf and the dialogic reading treatment, and one no-intervention control arm. Among dialogic reading households we further cross-randomized the book sharing calendar. Factorial studies designed to test stimulation and nutrition interventions individually and in combination are common in the early childhood development literature of developing countries, including studies in Bangladesh, Colombia, India, Jamaica and Pakistan (Grantham-McGregor et al., 1991; Walker et al., 1991; Tofail et al., 2013; Yousafzai et al., 2014; Andrew et al., 2020; Grantham-McGregor et al., 2020; Attanasio et al., 2022).[1] Randomization was conducted centrally by the authors using a random number generator in Stata. Assignment was concealed and only known to participants and intervention facilitators at the moment of treatment.

[1]Tofail et al. (2013) do not cross-randomize the nutrition and stimulation interventions but use two samples, one with iron-deficient anemic children and one with non-anemic, non-iron deficient children, and randomize the stimulation intervention across these two samples.

The primary outcomes are children's hemoglobin levels and cognitive, expressive and receptive language, motor, and socioemotional functions. While the outcomes immediately follow from the program and study design, the trial had not been pre-registered in 2016. In the supplementary materials, we include the approved ethics proposal, which describes the study design, to alleviate concerns regarding pre-registration. The trial was registered with the AEA Registry in early 2018.

One year after implementation, we find no impact of the Lucky Iron Leaf intervention, the dialogic reading intervention (with or without calendar) or the combination of both on hemoglobin levels and mental functions on average. However, we find that the combined intervention improved receptive language skills by 0.54 standard deviations for non-anemic children at baseline, who make up 30 percent of the sample. This result suggests that children's initial health endowment complements parental investments, corroborating the earlier findings from India (Attanasio, Meghir and Nix, 2020). Given the salient effect in the combined intervention arm, it also suggests that parental investments complement each other in a dynamic way: iron fortification increases hemoglobin levels and avoids anemia which makes dialogic reading more productive. However, only 25 percent of households used the fortification technology regularly. The complementarity of the two interventions, therefore, must stem from impacts among Lucky Iron Leaf users. The anemia rate among children in families that used the Lucky Iron Leaf regularly reduced considerably by 40 percent (not significant after multiple hypothesis testing). Biologically, non-anemic children have an advantage in the efficacy of book sharing over their anemic peers because they can engage more actively, whereas anemic children tend to be more tired and fearful, and less attentive (Lozoff et al., 1998; Lozoff, 2007, 2011). The home visit for the delivery of the Lucky Iron Leaf intervention, which was focused primarily on the child's well-being, may have also reinforced dialogic reading investments.

Given the simplicity of our interventions, the effect size for non-anemic children is considerable. Comprehensive home-based parenting interventions patterned after the Jamaica-program in China, Colombia and India, with weekly home visits for six (China), 18 (Colombia and India I) or 24 months (India II), improved receptive language by about 0.2 standard deviations and cognition by about 0.3 standard deviations (Attanasio et al., 2014; Andrew et al., 2020; Sylvia et al., 2020; Grantham-McGregor et al., 2020).[2] A parenting program in Pakistan with monthly visits over 24 months resulted in larger impacts of 0.7 and 0.6 standard deviations for language and cognition, respectively (Yousafzai et al., 2014). Compared to our study population, anemia rates in the studies in Colombia and China are much lower at 19 to 27 percent only (anemia rates for studies in India and Pakistan are not known). If we take the Colombian iron deficiency anemia prevalence of 19 percent and our effect size for receptive language of 0.54 standard deviations, the hypothetical average effect is 0.44 standard deviations ($0.81 * 0.54 + 0.19 * 0$). However, the interventions tested in this study only significantly affected receptive language skills, whereas the more comprehensive home-visiting programs also improved cognition and, in some cases, expressive language

[2]For a comprehensive overview of the short and long-term effects of home visiting programs such as the Jamaica program and those programs patterned after Jamaica see García and Heckman (2022)

and motor skills. Overall, our results imply that more short-lived early childhood development programs have potential in populations where the prevalence of anemia among children is low.

We further contribute to the literature by examining how the family environment changes via intervention spillovers through neighborhood networks, using variation in treatment saturation rates. We find that if the number of dialogic reading households increases from none to all study households in 100 meters distance to an untreated household, then the development score of children in untreated households increases by half a standard deviation. This result provides novel evidence on positive neighborhood externalities of early childhood programs. We are aware of one other, to date unpublished, study that finds large spillovers of an early childhood program via neighborhood networks in the United States (List, Momeni and Zenou, 2019). Such positive spillovers suggest even greater returns from early interventions than previously estimated (Heckman, 2008; García et al., 2020, 2021).

The paper is organized as follows. In section II we introduce a simple framework for the production technology of human capital. In section III we describe the interventions, in section IV the research design and data, and in section V the estimation strategy. In section VI we present estimates of the program impact on children's development and anemia as well as effect heterogeneities. In section VII we present intervention take-up rates, treatment effects under consideration of spillovers and complier average causal effects. In section VIII we examine the interventions' impact on other domains of caregiving. In section IX we discuss contextual challenges for intervention take-up. We conclude in section X.

II Framework

Following work by Cunha and Heckman (2007), Cunha, Heckman and Schennach (2010) and others, we present a simple framework for the production function for human capital of children. Children's human capital accumulates over time depending on their initial endowment, investments into the child, parental characteristics and the family environment. We model the accumulation of two types of skills S, i and j, and health H at different ages t of the child. We think of S_i as cognitive skills and S_j as motor skills.

In the skill production function, each skill in $t+1$ is a function of own endowment at the previous age t and the other skill at t as well as health at t. Skill accumulation is further shaped by factors external to the child, namely by emotional or financial investments by parents I_t (e.g., attentiveness, time, nutrition or toys), parental characteristics P_t (e.g., educational background, employment status or caregiving knowledge) and the family's environment V_t (e.g., neighborhood influences or public health services). The stock of skill S_i at $t+1$ can be described by the following function:

$$S_{i,t+1} = f^{(t)}(S_{i,t}, S_{j,t}, H_t, I_t, P_t, V_t) \tag{1}$$

$f^{(t)}$ is increasing in its arguments. Skills are self-productive, so that $S_{i,t+1}$ is higher the higher is $S_{i,t}$ ($\partial f^{(t)}/\partial S_{i,t} > 0$). Further, skills and health are cross-productive, which means

that endowments at time t in other skills, $S_{j,t}$, or health, H_t, result in higher skills in the next period, $S_{i,t+1}$ ($\partial f^{(t)}/\partial S_{j,t} > 0$ and $\partial f^{(t)}/\partial H_t > 0$). Moreover, skills and health complement investments into the child, I_t, so that, for example, cognitive stimulation in t is more productive in children with higher skills or better health in t (e.g., for skill j: $\partial^2 f^{(t)}/\partial S_{j,t}\partial I_t > 0$). Investments, I_t, are mediated by parental characteristics, P_t, and the environment, V_t. For example, higher educational attainment may facilitate cognitive stimulation.

We further assume that children's health can be described by a production function that depends on health in the previous period, H_t, investments in the child, I_t, parental characteristics, P_t, and the family environment, V_t:

$$H_{i,t+1} = g^{(t)}(H_t, I_t, P_t, V_t) \tag{2}$$

$g^{(t)}$ is increasing in its arguments. Health is self-productive ($\partial g^{(t)}/\partial H_t > 0$) and complements parental investments ($\partial^2 g^{(t)}/\partial H_t \partial I_t > 0$), as previously estimated in the Indian context (Attanasio, Meghir and Nix, 2020). Further, investments, I_t, are mediated by parental characteristics, P_t, and the environment, V_t. The skill and health production technologies imply that investments into the child at age t affect skills directly at $t+1$, and indirectly at $t+2$, through their impact on health in $t+1$.

Our interventions are designed to enter the human capital accumulation process initially via parental investments, i.e., practicing dialogic reading and feeding iron fortified food, at age t. Complementary with initial endowments, dialogic reading at t is hypothesized to directly augment skills at $t+1$, whereas feeding iron fortified food at t is hypothesized to directly augment health, by reducing or avoiding iron deficiency anemia, and skills at $t+1$. In $t+2$, health is reinforced through improved health at $t+1$ (self-productivity). Similarly, skills in $t+2$ are reinforced through improved skills at $t+1$ (self-productivity and cross-productivity with the other skill), and through improved health at $t+1$ (cross-productivity with health). Continued dialogic reading and feeding of fortified food further augment skills and health over time in a dynamic complementary manner. The neighborhood environment, V, may also improve over time as other households practice dialogic reading and use the Lucky Iron Leaf.

III THE INTERVENTIONS

III.A Background

The study is located in the district of Madhepura in Bihar, India. Bihar is the third largest state of India and has the lowest GDP per capita. Of Bihar's 104 million inhabitants, 88 percent reside in rural areas (Census of India, 2011). Madhepura district comprises about two million inhabitants and is almost entirely rural.

Chronic and acute malnutrition are common in Madhepura. In 2015/16, about half of the children under the age of five were stunted and/or underweight, and 61 percent were anemic (International Institute for Population Sciences, 2017). As part of the Government's

2013 National Iron+ Initiative, community health workers are tasked to monitor the administration of iron syrup in every household with young children, adolescent girls and women of reproductive age in weekly to biweekly home visits. In addition, children are entitled to receive iron supplements free of charge from public health providers, such as health centers and hospitals. However, the outreach of the public distribution system is limited (Kapil and Bhadoria, 2014). In 2015/16, only 22 percent of children under the age of five in Bihar consumed any iron supplements in the seven days prior to the interview (International Institute for Population Sciences and ICF, 2017).

There are no public programs for the psychosocial stimulation of children under three years of age in rural Bihar. The Integrated Child Development Services (ICDS) in Anganwadi centers offer day care only to children from three years of age and up.[3] For children under the age of three the ICDS only offers health services - i.e. macro-nutrient food rations, vaccinations, and health check-ups.

III.B Dialogic reading

Dialogic reading is a method for sharing picture books with children. It is solely based on having a conversation about images in the book and does not require caregivers to be literate. Dialogic reading techniques emphasize strategic questioning and feedback to encourage children to think, speak and learn. Rather than listen passively to readings, children become story tellers based on images in the book and assisted by the caregiver's questions and encouragement.

The effectiveness of dialogic reading is well-studied in high-income countries.[4] Studies from low- and middle-income countries include small-scale, closely controlled or short-term trials in Mexico, Bangladesh, Egypt, Kenya, South Africa, and Turkey. In most countries, the efficacy of dialogic reading in class rooms was tested (Valdez-Menchaca and Whitehurst, 1992; Opel, Ameer and Aboud, 2009; Elmonayer, 2013; Simsek and Erdogan, 2015), whereas the programs in Kenya and South Africa provided training to caregivers (Cooper et al., 2014; Vally et al., 2015; Murray et al., 2016; Knauer et al., 2020). The evidence from caregiver programs suggests that a four to eight week program with weekly meetings improved children's receptive and expressive language, attention and socioemotional behavior. To our knowledge, this is the first trial that evaluates the impact of dialogic reading for implementation at large scale and in the Indian context.

The training program for the dialogic reading intervention of this trial was developed by the Mikhulu Child Development Trust and is a shorter version of the eight-week caregiver training program tested in South Africa (Cooper et al., 2014; Vally et al., 2015; Murray et al.,

[3]Even at the older age group of three to five year-olds take-up is far from universal, with only a third of the age-eligible children in Bihar having ever attended the preschool programs over the course of 12 months (International Institute for Population Sciences and ICF, 2017).

[4]For book sharing of caregivers with toddlers (up to three years) see Whitehurst et al. (1988); Arnold et al. (1994); Huebner (2000); in care institutions with pre-school children (aged three to six years) see Whitehurst et al. (1994); Bus, van IJzendoorn and Pellegrini (1995); Dale et al. (1996); Lonigan and Whitehurst (1998); Hargrave and Sénéchal (2000); Lever and Sénéchal (2011).

2016). During four home visits of 45 to 60 minutes, women facilitators taught and practiced methods of dialogic reading with the primary caregiver, who is usually the mother, and the child.[5] The first home visit focused on training basic behavior rules during book sharing, such as that the child must always have fun, that the caregiver should stay positive at all times and that book sharing should be practiced regularly. The second to fourth dialogic reading home visits included a recap of the previous meetings and then covered specific dialogic reading strategies, such as point and say or asking questions (session 2), linking pictures in the book to real contexts or experiences (session 3), and talking about feelings (session 4). These concepts were taught with the help of presentation slides, pictures and videos on a tablet, and handouts made of durable material which summarized the lessons learned of each training session in colorful pictures (see appendix Figure A.I.1). During each home visit the contents of books to be shared were discussed and the caregiver practiced book sharing with her child, guided by the facilitator. In each of the first three visits a picture book was left at the house, so that families in this treatment received three books in total to use for book sharing. Appendix I details further aspects of the intervention.

In addition to the books, all treatment households received handouts which summarize the lessons learned by the training sessions in simple colorful pictures and simple instructions in Hindi language (see appendix Figure A.I.1 for the English version).

The first two dialogic reading training sessions were delivered immediately after baseline data collection and one week apart from each other. Sessions three and four were delivered three months later, again one week apart from each other. The implementation of dialogic reading trainings was closely monitored through daily spot-checks and attendance sheets, which logged the time and duration of the training, and had to be signed by participating households.

To prepare for each of the four dialogic reading home visits, facilitators participated in a five to six-day training program, led by an experienced trainer from the Mikhulu Child Development Trust. In the training program, facilitators learned the principles and benefits of dialogic reading with the aid of power point presentations and videos, were familiarized with the books to be distributed to households, learned to apply the principles of dialogic reading to these books, and practiced dialogic reading with toddlers themselves.

In the study population, picture book sharing was essentially not practiced at baseline and the increase in book sharing at the extensive margin is one of the targets for improving children's skill development (Fletcher and Reese, 2005). In particular, we expect there to be benefits from increasing responsive parent-child interactions (Landry, Smith and Swank, 2006) and from exposing children to more spoken words (Weisleder and Fernald, 2013). For children to engage with caregivers and books attentively, so that they achieve optimal learning outcomes, children's interest in sharing picture books is key (Fletcher and Reese,

[5]In preparing this study we asked about 400 women, who were from the study region but not the study sample, who in the household spends the most time with their youngest child below the age of five. In 78% of interviews women reported that they themselves, as mothers, spend the most time with the child. The second most common response was mother-in-law, in 11% of cases. We also asked in our study sample at endline who in the household usually shares books with the child, allowing for multiple responses. 86% of book-sharing households indicated the child's mother, 37% indicated a sibling, 35% the father and 21% the mother-in-law.

2005; Ortiz, Stowe and Arnold, 2001).

To encourage regular and continued book sharing, a random subset of households also received a dialogic reading calendar during the third home visit. On the calendar, caregivers could mark with colorful stickers each day that books were shared. The calendar was intended to be a reminder and was inspired by the medical literature, in which calendar boxes or blisters are commonly used to improve compliance with medication (Zedler et al., 2011). The calendar was structured by the lunar months and included colorful graphics to indicate national holidays and festivities. The design was targeted to an illiterate population and was made of robust material to last throughout the intervention period (and beyond).

III.C The Lucky Iron Leaf™

The second intervention was the distribution of Lucky Iron Leafs. The Lucky Iron Leaf is a 7-cm long leaf-shaped iron ingot made from electrolytic iron. When placed in a cooking vessel with boiling water and some fruit acid – such as some drops of lemon juice or half a tomato – it leaches iron into the water. After 10 minutes of boiling, the Lucky Iron Leaf can be removed from the pot and uncooked food added, so that it can be fortified by cooking in the iron-enriched water.[6] The Lucky Iron Leaf can be used for daily cooking for five years at a NGO-sales price below USD 10. In contrast to supplements or multi-micronutrient powder, it does not require any replenishment. This makes it an easy-to-deliver and remarkably sustainable nutrition intervention. This is the first study to rigorously test the effectiveness of the fortification technology in reducing the prevalence of anemia among children and in a large-scale intention-to-treat design.

The Lucky Iron Leaf is the Indian alternative to the Lucky Iron Fish™, a fish-shaped iron ingot, which effectively reduced anemia among Cambodian women (Charles et al., 2010, 2015).[7] The Lucky Iron Leaf resembles a leaf of the *tulsi* tree which is a holy plant in Hinduism and is commonly used for herbal tea against cough and colds. We changed the shape of the iron ingot to make it more attractive to Indian households, following the results of a qualitative study in the target population prior to the baseline survey. During semi-structured interviews, interviewees associated positive characteristics with tulsi leaves and were enthusiastic about using a tulsi-shaped iron ingot for cooking.

In a single home visit, male facilitators distributed the Lucky Iron Leaf to the main meal-maker and, if available, at least one decision-maker of the household, at no cost. Women are typically meal-makers, whereas men are decision-makers, including decisions about dishes to be cooked. Because we were running our experiment in a patriarchal society, we expected the Lucky Iron Leaf to have higher acceptance and higher take-up if it was introduced by a male facilitator to a decision-maker, rather than by a female facilitator. The home visit started with the male facilitator providing information on iron deficiency anemia, its most common

[6]The safety of using such an iron ingot for cooking is documented by Armstrong, Dewey and Summerlee (2017).

[7]The Lucky Iron Leaf and the Lucky Iron Fish™ are otherwise of similar size and weight. They also do not differ meaningfully in their fortification ability. The Lucky Iron Leaf was produced by the Lucky Iron Fish™ Enterprise for the purpose of this study. For more information, see https://luckyironfish.com.

causes, symptoms and consequences. He also highlighted that young children and women are particularly prone to anemia. Next, facilitators explained how to use the Lucky Iron Leaf with the help of a short video, in which a woman from the study region demonstrates how she uses the Lucky Iron Leaf for cooking, and an instruction manual. Seeing a local woman using the Lucky Iron Leaf was intended to increase its acceptance and uptake. The instruction manual showed six images, one for each step of using the Lucky Iron Leaf: 1) wash it before use, 2) boil it in water together with a tomato (or other fruit acid) for 10 minutes, 3) take the iron ingot out of the water, 4) add lentils (as daal is one of the most common foodstuffs; however, the facilitator also explained how it can be used in other dishes), 5) dry the iron ingot, and 6) store it (see Appendix Figure A.I.2 for the manual). The picture manual was designed to be easy to understand for the mostly non-literate study population and had been previously assessed in focus groups in Pune, India.[8] The facilitators were then instructed to discuss with participants how to integrate the Lucky Iron Leaf into their everyday cooking and how to clean and store it safely. Finally, facilitators offered to stay on to practice cooking together with the Lucky Iron Leaf, but this offer was typically accepted only when participants were about to cook anyway. The roll-out of the Lucky Iron Leaf closely followed the Lucky Iron Fish^{TM} implementation schedule of the trials in Cambodia. The distribution of Lucky Iron Leafs was closely monitored through regular spot-checks and implementation sheets logging intervention times. Appendix I details further aspects of the intervention.

In preparation for these home visits, facilitators participated in a five-day training program. In the training program, facilitators learned about iron deficiency anemia, its causes, symptoms, consequences, and about population groups most vulnerable to anemia. Next, they were introduced to the Lucky Iron Leaf and its benefits, and were trained in its use. Part of this was using the Lucky Iron Leaf firsthand and practicing with it. Each facilitator received a Lucky Iron Leaf to keep and use. In the last step, facilitators were trained in the procedures of the home visit, such as how to explain the instruction manual to households.

The iron ingot is a low dose fortification tool and overdosing is unlikely. Laboratory tests conducted at the University of Guelp showed that the Lucky Iron Leaf enriches one serving of the typical lentils-based dish *daal* by 4.25-5 mg of bioavailable iron. At baseline, caregivers reported that children eat on average 2.6 times per day and consume about half of an adult-equivalent portion at each sitting. We therefore expect children to consume in the ideal case 5.5-6.5 mg of supplementary iron per day. The WHO recommends 10-12.5 mg elemental iron for under two-year olds and 30 mg for two- to five-year olds for three consecutive months in high risk regions such as Bihar (World Health Organization, 2016).

We expect the Lucky Iron Leaf intervention to improve skill development by reducing or avoiding iron deficiency and iron deficiency anemia. Iron-deficient children are less attentive and more fearful than their peers and, hence, interact less with their environment, which can inhibit learning and mental development (Lozoff et al., 1998; Lozoff, 2007, 2011). This mechanism assumes a complementarity of children's health and parental investments as described in the framework for the production technology of human capital in section

[8]The Lucky Iron Leaf is manufactured in Pune.

II. Iron deficiency is also theorized to directly influence the neurodevelopment of children (described as cross-productivity in section II) because iron is needed for myelination and the dopamine metabolism (Lozoff, 2007; Beard, 2008; Lozoff, 2011). Myelin insulates the axons (which transmit information to different neurons, muscles and glands) and, thereby, increases the speed of information transmission and processing in the brain. Dopamine is a neurotransmitter which regulates the award seeking behavior and is, therefore, closely linked to motivation and learning (Wise, 2004; Flagel et al., 2011). While these bio-medical processes as well as evidence from rodent trials suggest strong adverse effects of iron deficiency on children's mental development, the evidence of the effectiveness of iron supplementation in iron-deficient anemic children to improve mental development is mixed and raises doubts about the reversibility of mental delays through supplementation. This trial expands the existing evidence base on the effectiveness of iron supplementation on mental functions and, in particular, studies the complementarity of improved health, through the use of the Lucky Iron Leaf, with dialogic reading. A positive side effect of the Lucky Iron Leaf intervention might further be general awareness-raising among caregivers about child nutrition and anemia, which could positively impact complementary feeding practices (Dewey and Adu-Afarwuah, 2008).

IV Experimental design and data

IV.A Sample selection

Sampling and randomization

The study population is recruited from a listing of 1,995 households with pregnant women from 2015. In 2016, these women had children in our target age group of 10 to 20 months. The listing was gathered from rural Anganwadi center registers in 140 villages.[9] Sixty-eight gram panchayats, which comprise 180 villages, were randomly sampled from a total of 95 gram panchayats in six of Madhepura's thirteen blocks (sub-districts). Because in some villages the lists of pregnant women were not made available, the number of gram panchayats and villages reduced to 56 and 140, respectively. The number of households sampled per village ranges from 5 to 49.

The list of households with pregnant women was generated in 2015 for an impact evaluation of a participatory learning and action approach program, called Gram Varta, designed to improve health, nutrition, water, sanitation and hygiene (HNWASH) outcomes of adults and children. The intervention consisted of women's group meetings, using existing savings groups, and was completed before our intervention started. The HNWASH program was randomized at the gram panchayat level and was overall ineffective, including final outcomes such as health and upstream outcomes such as HNWASH beliefs and behavior (Bogler et al., 2023). The endline of this evaluation was used as the baseline of our study.

[9] In 2015/16, 75 percent of all pregnant women in Madhepura had registered their pregnancies (International Institute for Population Sciences, 2017).

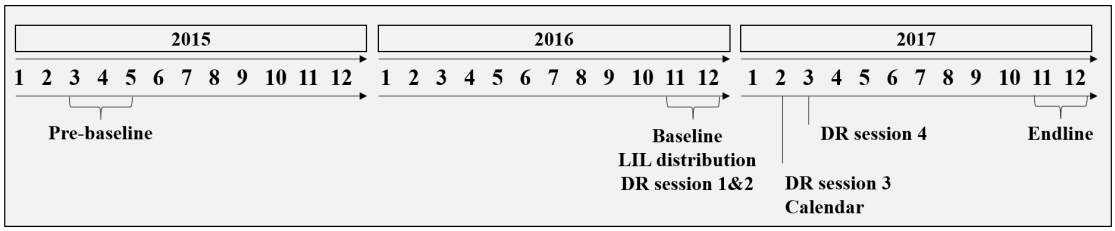


Figure 1: Timeline of data collections and intervention implementation

We randomly assigned the 2015 sample of 1,995 households with pregnant women to one of four treatment arms: (i) a dialogic reading group ($N = 498$), (ii) a Lucky Iron Leaf group ($N = 499$), (iii) a dialogic reading and Lucky Iron Leaf group ($N = 499$), and (iv) a no-intervention control group ($N = 499$). Randomization was conducted centrally by the authors using a random number generator in Stata and prior to the baseline survey, because the intervention implementation was conducted subsequent to the survey. Assignment was concealed and only known to participants and intervention facilitators at the moment of treatment.

The factorial design allows us to capture potential interactions of the nutrition and psychosocial stimulation interventions, either resulting from complementarities of parental investments, i.e., dialogic reading and iron fortification, or through a complementarity of improved health, through iron fortification, and dialogic reading.

During the intervention implementation which households received the Lucky Iron Leaf intervention and participated in each of the four dialogic reading training sessions was precisely recorded. We collected proof of delivery through signatures or fingerprints of participants, and field supervisors and field research assistants conducted regular unannounced spot-checks throughout the implementation period. Ninety-six percent of households assigned to the Lucky Iron Leaf intervention received the treatment and 94 percent of dialogic reading households received at least one training session: the average number of training sessions is 3.5. Based on these records and qualitative reports from spot-checks, we have no reason to be concerned about the fidelity of the implementation of interventions. Section VII.A reports on utilization rates of the interventions.

After the baseline survey, the implementation of the Lucky Iron Leaf intervention, and the first two dialogic reading household visits, treatment groups (i) and (iii) were further randomized into a calendar ($N = 499$) and no-calendar group ($N = 498$). The randomization was conditional on the caregiver having participated in at least one dialogic reading training session, which was a pre-requisite for being revisited in March 2017 for training sessions 3 and 4, when the calendar was distributed.

Figure 1 shows the timeline of events. The baseline survey, the Lucky Iron Leaf distribution and the first two dialogic reading training sessions began in November 2016 and lasted for eight weeks. Dialogic reading training sessions three and four took place in February and March 2017. The calendars were distributed during the third dialogic reading training session. We collected endline data exactly one year after the baseline survey in December 2017.

The experimental design and outcome indicators had been registered at the American Economic Association's registry for randomized controlled trials shortly after the completion of the trial. We did not pre-specify an analysis plan for the estimation strategy and heterogeneous treatment effect analyses, and therefore the analysis may be considered explorative. However, the main outcomes – hemogloin levels and mental functions – follow immediately from the tested interventions. In the supplementary materials, we include the approved ethics proposal, which describes the study design and primary outcomes, to alleviate concerns regarding pre-registration.

Attrition

In 2015, prior to our study, 1,995 households with pregnant women had been sampled. One and a half years later, in 2016, 80 percent of formerly pregnant women could be reached for our baseline data collection. The main reason for dropout before baseline was the incorrect documentation of a woman's permanent household location, because of the tradition of women moving to their parental home for pregnancy and child birth. Other households dropped out of the sample because not all conceived children survived until baseline data collection in 2016.[10] The baseline sample of households with alive children thus totaled 1,483 (74% of pre-baseline sample). Because the reduction in sample size occurred prior to the intervention roll-out, there are no biases from selective attrition at this stage.

The final estimation sample consists of 1,164 households, corresponding to 78.5 percent of baseline households.[11] The main reasons for attrition between baseline and endline are the absence of households or children at the time of data collection (N=152) and incomplete child development tests (N=83). Less common causes of attrition are migration of households across village borders, incorrect household identification information and child death. Ten households refused to participate.

To check for selective attrition from baseline to endline as well as from pre-baseline to endline, we regress an indicator for attrition on the three treatment group indicators. The coefficients are small and none of them approach statistical significance (appendix Table A.II.1). Further, the baseline balance in the estimation sample is similar to the baseline balance in a sample which is not restricted to whether we observe endline outcomes (see appendix A.II.2).

IV.B Data

Outcome indicators

We use the early childhood development test FREDI 0-3 to measure children's skills and behavior (Mähler, Cartschau and Rohleder, 2016; Macha and Petermann, 2017). FREDI 0-3

[10]The magnitude of child deaths is comparable to the state-wide statistics. The infant mortality rate in rural Bihar is 50 per 1,000 life births and only 92 percent of all pregnancies lead to live births in Bihar (Kochar et al., 2014; International Institute for Population Sciences and ICF, 2017).

[11]We define the estimation sample as observations with at least one non-missing child development score.

tests cognitive, language, motor, and socioemotional development. It includes playful tasks administered to the child and interview questions posed to the caregiver. FREDI 0-3 was normed to German children and its language items were validated to the tests ELAN-R and SETK-2 (Kiese-Himmel, 2013, 2014). The test was adapted for Bihar by the same psychologists who developed the original test. The adapted test comprises around 40 items and we administered different test versions for children younger or older than 15 months at baseline or 27 months at endline. The FREDI 0-3 was administered at home by local women testers, who were trained by the developers of the FREDI 0-3 test for two weeks. The testers were blind to treatment assignment at baseline and endline. Daily spot-checks were conducted to ensure a high quality administration of the test. For the data analysis, we standardize scores of cognitive, receptive language, expressive language, motor and psychosocial skills relative to the experimental control group and with respect to test year, age group, a linear age in months trend and a heteroskedastic residual variance.

Iron deficiency is the most common cause of anemia, which we measure through low hemoglobin (Hb) concentration in the blood.[12] We used HemoCue 301® machines to determine the Hb concentration from capillary blood samples at the point of care and informed all patients about the results directly after testing. Anemia cut-offs follow WHO recommendations, according to which 6-59 month old children are classified as anemic if Hb$<$11 g/dl (World Health Organization, 2011). For ethical reasons, we implemented the following action plan. In case of mild (9-10.9 g/dl) or moderate anemia (7-8.9 g/dl), we recommended that caregivers seek treatment at the local primary health center or with the community health worker, and change the child's diet towards more iron-rich foods. Severely anemic children (Hb$<$7 g/dl) were offered transport and funding for the cost of treatment in the local hospital or health center. At baseline, about 70 percent of children in the sample were anemic, but only 1 percent were severely anemic. There is potential for interaction effects of baseline testing and the Lucky Iron Leaf intervention, however, in a population and set-up very similar to ours, the behavior and health of participants did not change in response to recommendations or referrals (Krämer, Kumar and Vollmer, 2021*a*,*b*). Additionally, we test for heterogeneous take-up by baseline anemia status and do not find differential effects. The field staff conducting the hemoglobin measurements were blind to treatment assignment at baseline and endline.

Sample characteristics and balance

Table I presents means of baseline background characteristics and outcomes for each of the four main experimental groups in the estimation sample. Most participants are Hindus and members of the caste category "other backward class". The highest education level in most households and among mothers is uncompleted primary school. About 90 percent of mothers were breastfeeding their children at baseline and 59 percent breastfed exclusively. About 35 percent of children had received iron supplements in the last three months. Of the eight stimulating activities we enquired about, 5.6 activities were conducted with the

[12]While anemia can also be caused by infectious diseases, deficiencies of vitamin A or B12, or genetic disorders, iron deficiency is globally the most common cause of anemia (Kassebaum et al., 2014).

Table I: Baseline balance across treatment arms in the estimation sample

	Control		Dialogic reading			Lucky Iron Leaf			Lucky Iron Leaf & Books		
	Mean SD	N	Mean SD	Std. Diff. p-value	N	Mean SD	Std. Diff. p-value	N	Mean SD	Std. Diff. p-value	N
Household characteristics:											
Household size	5.69	283	5.77	-0.04	299	5.75	-0.03	294	5.75	-0.03	285
	2.01		2.30	0.63		2.26	0.71		2.34	0.72	
Hindu	0.87	283	0.82	0.13	300	0.85	0.05	295	0.86	0.02	286
	0.34		0.38	0.13		0.36	0.52		0.34	0.84	
Caste category:											
Scheduled caste	0.30	282	0.30	-0.01	300	0.31	-0.02	295	0.28	0.04	285
	0.46		0.46	0.89		0.46	0.85		0.45	0.65	
Scheduled tribe	0.04	282	0.04	-0.01	300	0.02	0.09	295	0.04	-0.02	285
	0.19		0.20	0.95		0.15	0.29		0.20	0.85	
Other backward class	0.59	282	0.58	0.02	300	0.58	0.02	295	0.61	-0.04	285
	0.49		0.49	0.77		0.49	0.82		0.49	0.66	
General category	0.07	282	0.08	-0.02	300	0.09	-0.06	295	0.07	0.02	285
	0.26		0.27	0.79		0.28	0.45		0.25	0.84	
Highest education in HH:											
No schooling	0.43	283	0.49	-0.10	300	0.47	-0.08	295	0.40	0.07	286
	0.50		0.50	0.21		0.50	0.34		0.49	0.38	
Primary	0.19	283	0.15	0.11	300	0.16	0.07	295	0.17	0.05	286
	0.39		0.35	0.19		0.37	0.38		0.37	0.54	
Middle school	0.13	283	0.13	0.00	300	0.15	-0.04	295	0.17	-0.11	286
	0.34		0.34	0.98		0.35	0.60		0.38	0.18	
$\geq$ High school	0.25	283	0.24	0.02	300	0.22	0.06	295	0.26	-0.03	286
	0.43		0.43	0.76		0.42	0.44		0.44	0.68	
Asset quintile[a]	5.31	282	5.41	-0.04	300	5.01	0.10	293	5.23	0.03	285
	2.88		2.90	0.66		2.88	0.22		2.94	0.75	
Housing quintile[b]	3.76	283	3.70	0.02	300	3.48	0.10	295	3.62	0.05	286
	2.77		2.73	0.79		2.75	0.23		2.82	0.55	
Mother characteristics:											
Age in years	24.99	283	24.75	0.06	300	24.56	0.11	295	25.07	-0.02	286
	3.98		4.15	0.48		3.74	0.18		4.03	0.80	
Completed education:											
No schooling	0.74	283	0.76	-0.06	300	0.84	-0.26	295	0.73	0.01	286
	0.44		0.43	0.49		0.36	0.00		0.44	0.91	
Primary	0.09	283	0.05	0.16	300	0.03	0.24	295	0.07	0.09	286
	0.29		0.22	0.05		0.18	0.00		0.25	0.26	
Middle school	0.05	283	0.08	-0.10	300	0.04	0.08	295	0.10	-0.19	286
	0.22		0.27	0.25		0.19	0.36		0.31	0.02	
High school or higher	0.12	283	0.11	0.02	300	0.08	0.11	295	0.09	0.07	286
	0.32		0.31	0.80		0.28	0.20		0.29	0.39	
Can read SMS	0.27	283	0.28	-0.03	300	0.18	0.20	295	0.28	-0.03	286
	0.44		0.45	0.75		0.39	0.02		0.45	0.69	
Worked past 12 months	0.91	283	0.89	0.06	300	0.91	-0.00	295	0.91	0.01	286
	0.29		0.32	0.48		0.30	0.99		0.30	0.92	
Empowerment[c]	0.37	283	0.46	-0.20	300	0.38	-0.02	295	0.40	-0.06	286
	0.48		0.50	0.02		0.49	0.83		0.49	0.45	
Decides child nutrition	0.52	267	0.57	-0.11	279	0.54	-0.05	264	0.49	0.05	251
	0.50		0.50	0.21		0.50	0.57		0.50	0.60	

Note: Table continues on next page. Standard deviations (SD) are presented directly below means. Std. Diff. refers to the standardized difference in means of the control group and the treatment group. p-values refer to a t-test of the equality of means of the control group and the treatment group and are presented below the Std. Diff. [a]10 quintiles based on a durable asset index generated by factor analysis. [b]10 quintiles based on a housing quality index generated by factor analysis. [c]Indicator equals one if the mother is allowed to visit alone one of five places (market, health facility, neighbors, relatives or friends outside the village, place of worship) and participates in one of four decisions (health investments, household purchases, family visits outside village, and farm).

Table I continued

	Control		Dialogic reading			Lucky Iron Leaf			Lucky Iron Leaf & Books		
	Mean SD	N	Mean SD	Std. Diff. p-value	N	Mean SD	Std. Diff. p-value	N	Mean SD	Std. Diff. p-value	N
Child characteristics:											
Sex of child	0.50	282	0.55	-0.10	298	0.58	-0.16	292	0.48	0.05	281
	0.50		0.50	0.23		0.49	0.06		0.50	0.53	
Currently breastfed	0.90	280	0.93	-0.10	294	0.90	-0.03	284	0.89	0.03	273
	0.31		0.26	0.23		0.29	0.74		0.32	0.71	
Vit-A past 6 months	0.72	180	0.73	-0.02	179	0.72	0.01	182	0.77	-0.11	180
	0.45		0.44	0.84		0.45	0.96		0.42	0.28	
Iron past 3 months	0.33	252	0.37	-0.09	258	0.34	-0.03	264	0.37	-0.09	253
	0.47		0.48	0.31		0.48	0.71		0.48	0.32	
Home environment:											
Stimulation index[d]	5.63	270	5.58	0.03	281	5.60	0.02	268	5.59	0.02	263
	1.73		1.68	0.70		1.75	0.85		1.79	0.77	
Good educat. measures[e]	0.81	267	0.79	0.05	277	0.78	0.06	267	0.79	0.04	257
	0.40		0.41	0.60		0.42	0.46		0.41	0.66	
Bad educat. measures[f]	0.73	274	0.68	0.12	284	0.74	-0.02	270	0.73	-0.01	264
	0.44		0.47	0.16		0.44	0.85		0.44	0.90	
Outcome measures:											
Cognitive	0.00	250	-0.11	0.11	261	-0.08	0.07	257	0.01	-0.01	255
	1.00		1.01	0.22		1.21	0.44		1.04	0.91	
Receptive language	-0.00	256	-0.19	0.18	267	-0.10	0.09	266	-0.18	0.17	260
	1.00		1.02	0.04		1.13	0.29		1.05	0.05	
Expressive language	0.00	255	-0.13	0.12	268	-0.11	0.10	265	-0.10	0.10	264
	1.00		1.16	0.19		1.10	0.24		1.01	0.27	
Motor	0.00	248	0.02	-0.03	255	0.08	-0.08	250	-0.01	0.01	251
	1.00		0.95	0.78		1.00	0.37		0.98	0.87	
Socioemotional	-0.00	250	-0.06	0.05	264	-0.11	0.10	258	0.11	-0.10	256
	1.00		1.17	0.54		1.16	0.26		1.13	0.26	
Hemoglobin g/dL	10.21	191	10.11	0.08	197	10.25	-0.03	185	10.32	-0.08	197
	1.34		1.32	0.43		1.41	0.78		1.40	0.44	
Anemia (any type)	0.71	191	0.72	-0.02	197	0.68	0.06	185	0.66	0.09	197
	0.46		0.45	0.85		0.47	0.59		0.47	0.38	
Moderate Anemia	0.39	191	0.43	-0.08	197	0.39	0.01	185	0.36	0.08	197
	0.49		0.50	0.44		0.49	0.95		0.48	0.45	
p-value of joint F-test				0.19			0.68			0.35	

Note: Standard deviations (SD) are presented directly below means. Std. Diff. refers to the standardized difference in means of the control group and the treatment group. p-values refer to a t-test of the equality of means of the control group and the treatment group and are presented below the Std. Diff. [d]Sum of stimulating activities conducted with the child in the past 3 days. [e]Equals 1 if the caregiver explains wrong behavior to child, takes away privileges or gives child something else to do. [f]Equals 1 if the caregiver shouts, yells or screams at the child or spanks, hits, kicks or slaps the child.

child in the past three days. Over 70 percent of caregivers used yelling or physical violence (e.g. spanking or slapping) as an educational measure in the past month ("bad educational measure" in Table I).

We use standardized differences in means, statistical significance of differences in means (p-value), and a joint F-test for orthogonality to evaluate sample balance in Table I. Overall, the joint F-test does not reject the equality of control group and treatment group means (see bottom of Table I). Few characteristics differ when judged by individual significant

differences or standardized differences of 0.2 or larger (Cohen, 1988; Imbens and Rubin, 2015). These include maternal education, literacy and empowerment, and children's sex. Notably, in the combined intervention and dialogic reading groups, children are worse off in receptive language skills. Other skill dimensions do not show statistically significant differences.

V ESTIMATION STRATEGY

Our main analysis focuses on the effects of the dialogic reading and Lucky Iron Leaf interventions and their interaction on children's outcomes. The dialogic reading calendar will be considered with respect to intervention take-up in section VII.A. We estimate intention-to-treat effects using the following specification:

$$y_i = \alpha + \beta_1 DR_i + \beta_2 LIL_i + \beta_3 DR\&LIL_i + X_i^{'}\gamma + u_i. \quad (3)$$

y_i is one of five child development scales, hemoglobin levels or an anemia dummy. DR_i, LIL_i and $DR\&LIL_i$ indicate treatment assignment to the dialogic reading, the Lucky Iron Leaf and the combined intervention group, respectively. X_i is a vector of imbalanced baseline covariates and includes mothers' education, reading abilities and empowerment, children's sex, and subdistrict fixed effects. In estimations on hemoglobin levels and anemia, we additionally control for measurement device fixed effects. β_1 and β_2 represent the intention-to-treat effects, i.e. disregarding non-compliance, of the respective stand-alone intervention, whereas β_3 measures the interaction effect of the combination of both interventions. Due to random treatment assignment at the household level, we do not cluster standard errors.

In addition to the covariate adjusted treatment effects, we present intention-to-treat effects without covariate adjustment and with baseline outcomes (ANCOVA). Controlling for baseline outcomes increases the estimates' precision, and overcomes a potential downward bias from imbalances in receptive language apparent in both dialogic reading groups; however, it also reduces the estimation sample by more than 100 observations. Therefore, we present the ANCOVA estimates as additional evidence only. Due to the dialogic reading groups' baseline disadvantages in receptive language, the no-covariate and covariate adjusted estimates on receptive language present lower bound estimates.

In total, we test 30 hypotheses in the main intention-to-treat analysis, following from six outcomes tested in five subgroups. The six outcomes refer to iron deficiency, proxied by Hb levels and anemia, and the five development scales. The five subgroups refer to the intervention impacts in the full estimation sample and intervention impacts by maternal education and empowerment, and children's sex and anemia status at baseline. We correct for multiple hypotheses testing using Bonferroni adjustment and taking into account the average correlation across outcomes of 0.31 (Sankoh, Huque and Dubey, 1997; Aker et al., 2012). The multiple-hypothesis-testing-corrected-equivalent 10 percent significance level is 0.0096 and the equivalent 5 percent significance level is 0.0048. We will use these thresholds to mark adjusted significance levels in regression tables.

Given a sample size of roughly 280 observations per main treatment arm, statistical power of 80 percent and a significance level of 0.05, we are able to detect effect sizes of 0.24 standard deviations in child development outcomes. Using the multiple-hypothesis-testing-corrected significance level of 0.0048 (equivalent to 0.05), the minimum detectable effect size rises to 0.31 standard deviations. The minimum detectable effect size in hemoglobin is 0.35 g/dl (0.46 g/dl after multiple hypothesis testing correction) and 11.9 percentage points in anemia prevalence (15.4 percentage points after multiple hypothesis testing correction), given a sample size of 260 children per treatment arm. The multiple-hypothesis-testing-adjusted minimum detectable effect size in hemoglobin is less than half the size of the impact (1.18 g/dl) the Lucky Iron Fish had on hemoglobin of Cambodian women (Charles et al., 2015).

VI RESULTS

VI.A Intention-to-treat effects on child development

Table II presents the intention-to-treat estimation results on cognitive development, receptive language, expressive language, motor skills, and socioemotional development. The different columns refer to a simple regression of the outcome on the three treatment group indicators (model 1), a covariate adjusted model (model 2), and a covariate and baseline outcome adjusted model (model 3). The intention-to-treat effects are small and not statistically significant across the three treatment arms, outcomes and specifications. The adjustment for covariates causes a slight but not significant increase in the effect size relative to the simple model across all development scales. When we add controls for baseline outcomes, the magnitudes of the effects remain similar and the sample size decreases by about 10 percent. For subsequent analyses, we use the covariate-adjusted specification (model 2) as our preferred specification because it potentially purges the coefficients of selection and improves efficiency, but maintains the larger sample size.

To test the results' robustness, we add children's age fixed effects (in months) and development test facilitator fixed effects to model 2 (see appendix Table A.III.1). We also exploit the panel structure of our data and estimate treatment effects using difference-in-differences, child fixed effects and inverse probability weighted estimations. Finally, we test whether dialogic reading affected head circumference, as an alternate measure of early mental functions (see appendix Table A.III.2; Dupont et al., 2018).[13] The robustness checks confirm the findings in Table II.

VI.B Intention-to-treat effects on hemoglobin and anemia

In Table III we present the intention-to-treat effect estimation results on hemoglobin, any type of anemia, and moderate anemia. We use linear probability models to estimate the

[13] Head circumference was standardized according to WHO growth reference tables (World Health Organization, 2007).

Table II: Intention-to-treat effects on child development

	Model 1	Model 2	Model 3
Cognitive			
DR	0.028	0.106	0.081
	(0.083)	(0.082)	(0.087)
LIL	-0.067	-0.001	-0.016
	(0.086)	(0.083)	(0.086)
DR & LIL	0.036	0.063	0.082
	(0.090)	(0.086)	(0.090)
Observations	1146	1136	1013
Adjusted R^2	-0.001	0.082	0.087
Control mean	-0.00	-0.00	-0.00
Receptive language			
DR	-0.047	-0.009	-0.028
	(0.083)	(0.084)	(0.090)
LIL	-0.027	0.017	-0.022
	(0.083)	(0.084)	(0.089)
DR & LIL	-0.008	0.018	0.000
	(0.087)	(0.087)	(0.091)
Observations	1159	1148	1038
Adjusted R^2	-0.002	0.029	0.030
Control mean	-0.02	-0.02	-0.03
Expressive language			
DR	-0.014	0.034	0.040
	(0.086)	(0.085)	(0.089)
LIL	-0.039	0.039	0.031
	(0.086)	(0.085)	(0.089)
DR & LIL	0.019	0.035	0.056
	(0.090)	(0.089)	(0.091)
Observations	1159	1148	1041
Adjusted R^2	-0.002	0.049	0.077
Control mean	-0.01	-0.01	0.00
Motor			
DR	-0.012	0.051	0.005
	(0.083)	(0.081)	(0.085)
LIL	-0.022	0.021	-0.026
	(0.086)	(0.083)	(0.085)
DR & LIL	0.029	0.060	0.015
	(0.086)	(0.084)	(0.087)
Observations	1123	1113	994
Adjusted R^2	-0.002	0.068	0.082
Control mean	-0.00	0.00	0.00
Sociomeotional			
DR	0.009	0.061	0.093
	(0.082)	(0.081)	(0.087)
LIL	-0.022	0.041	0.085
	(0.081)	(0.080)	(0.086)
DR & LIL	-0.017	0.012	0.019
	(0.084)	(0.083)	(0.088)
Observations	1151	1140	1017
Adjusted R^2	-0.002	0.040	0.044
Control mean	-0.01	-0.01	0.00
Controls		✓	✓
Baseline outcome			✓

Note: Dependent variables are standardized development scores and effect sizes are in standard deviations. Control variables are maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Baseline outcome indicates that the estimation controls for the baseline value of the model's dependent variable. Standard errors are in parentheses. Conventional significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Multiple testing corrected significance levels: $_{+}$ $p<0.0096$, $_{++}$ $p<0.0048$, $_{+++}$ $p<0.0010$.

impact on binary anemia outcomes. We find no effect of the Lucky Iron Leaf as a stand-alone intervention or in combination with dialogic reading on children's hemoglobin levels and anemia status. The coefficients are close to zero in the unadjusted model (model 1) and the covariate adjusted model (model 2). The baseline outcome adjusted model coefficients (model 3) are somewhat larger but have the opposite sign to what is expected, potentially due to sample selection caused by the reduction in sample size (see appendix Table A.III.3 for results from all models in the baseline outcome adjusted sample).

The results are robust to replacing the standard covariates with a set of covariates, which we identified in a balance analysis of a sample restricted to non-missing hemoglobin values at endline, rather than only non-missing child development scores as in Table I. Further, the results are robust to difference-in-differences and fixed effects estimations.

If iron deficiency is not overcome, then the causal chain from intervention implementation to the impact on anemia and mental development fails.[14] If there had been temporary improvements in iron levels, then they did not sustain until endline and did not affect development outcomes in the medium term. To further explore the effect of the Lucky Iron Leaf on outcomes that show lasting effects of temporary nutritional changes, we estimate its impact on stunting, reflecting a child's nutrition history (Black et al., 2013), as well as on wasting and underweight.[15] The results show that the Lucky Iron Leaf intervention also had no impact on growth outcomes (see appendix Table A.III.2).[16]

We presume zero-effects of the dialogic reading treatment on iron deficiency because there is no theoretical link from book reading to nutrition. Yet, in Table III, the intention-to-treat effects of the stand-alone dialogic reading treatment on hemoglobin are negative and, in models 2 and 3, they are statistically significant before multiple hypothesis testing. The effect in model 3 is driven by sample selection due to missing baseline anemia information (see appendix Table A.III.3). We do not find a statistically significant negative effect of the pure book sharing intervention on early skills. If the book sharing intervention indeed had a negative effect on iron deficiency, which also affected child development, then these negative effects on child development were compensated for by the positive effects of sharing books.

VI.C Heterogeneous treatment effects

We assess heterogeneous treatment effects by mothers' education and empowerment, and children's sex and anemia status at baseline. None of the heterogeneity tests were pre-registered. Naturally, the lack of pre-registration comes at a cost of credibility, among other important disadvantages (Humphreys, De la Sierra and Van der Windt, 2013; Miguel et al.,

[14]The theory of change breaks down at the point of utilization of the Lucky Iron Leaf as only three percent of households use it. In section VII.A, we describe utilization rates of the Lucky Iron Leaf in detail. In section IX we discuss potential reasons for under-utilization.

[15]Growth outcomes were derived in accordance with WHO growth reference tables (World Health Organization, 2007).

[16]The link between iron intake and growth outcomes is still disputed (Rivera et al., 2003; Iannotti et al., 2006). Therefore, the lack of treatment effects may also be interpreted as a missing link between the two in general.

Table III: Intention-to-treat effects on hemoglobin and anemia

	Model 1	Model 2	Model 3
Hemoglobin in g/dl			
DR	-0.152	-0.177	-0.121
	(0.119)	(0.119)	(0.122)
LIL	0.003	-0.010	-0.132
	(0.120)	(0.121)	(0.127)
DR & LIL	0.025	-0.012	-0.158
	(0.124)	(0.124)	(0.131)
Observations	1048	1039	710
Adjusted R^2	-0.000	0.033	0.269
Control mean	10.63	10.63	10.67
Any anemia			
DR	0.056	0.073*	0.112**
	(0.043)	(0.043)	(0.050)
LIL	0.008	0.016	0.064
	(0.044)	(0.043)	(0.048)
DR & LIL	-0.018	-0.004	0.045
	(0.044)	(0.044)	(0.048)
Observations	1048	1039	710
Adjusted R^2	0.000	0.025	0.176
Control mean	0.56	0.56	0.54
Moderate anemia			
DR	0.035	0.035	0.030
	(0.040)	(0.040)	(0.044)
LIL	0.006	0.003	0.044
	(0.040)	(0.040)	(0.046)
DR & LIL	-0.009	-0.001	0.054
	(0.039)	(0.040)	(0.045)
Observations	1048	1039	710
Adjusted R^2	-0.001	0.013	0.111
Control mean	0.28	0.28	0.26
Controls		✓	✓
HemoCue machine fixed effects		✓	✓
Baseline outcome			✓

Note: Hemoglobin is measured in g/dl. Any and moderate anemia are binary indicators. Control variables are maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. HemoCue machine fixed effects control for the hemoglobin measurement device and tester. Baseline outcome indicates that the estimation controls for the baseline value of the model's dependent variable. Standard errors are in parentheses. Conventional significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Multiple testing corrected significance levels: $_{+}$ $p<0.0096$, $_{++}$ $p<0.0048$, $_{+++}$ $p<0.0010$.

2014; Ofosu and Posner, 2020). To attenuate such concerns, for each tested heterogeneity, we explain why the heterogeneity is salient based on the study design or present theoretical underpinnings that emerge directly from the relevant literature:

1. *Maternal education*
   In a series of estimations of production functions for cognition and health of children in low income settings, Attanasio and coauthors investigate the relevance of parental cognition, typically proxied by education and literacy (Attanasio, Meghir and Nix,

2020; Attanasio et al., 2020*a*,*b*).[17] For India, Attanasio, Meghir and Nix (2020) find that parental cognition improves the productivity of investments in children's cognition, but not in health. Similarly, in the human capital production function described in section II, mother's education enters as parental characteristics P_t and is assumed to mediate investments into children I_t. In our estimation sample, in 86 percent of households the mother is reported as the person who usually shares books with the child, i.e., who conducts the investment.[18] Based on this, we hypothesize that the productivity of the dialogic reading investment is higher for better educated mothers. Educated mothers are likely to follow the training and internalize the methods of dialogic reading with more ease and perform higher quality book sharing. In line with Attanasio, Meghir and Nix (2020), we do not expect strong heterogeneities by mother's education for the nutrition intervention, because the cooking procedures are simple to implement.

2. *Maternal empowerment*
   Also women's empowerment enters the human capital production functions as P_t and is expected to mediate parental investments I_t. The literature on the influence of spouses' bargaining power on investments in children is abundant (e.g., Dyson and Moore, 1983; Thomas, 1993; Lundberg, Pollak and Wales, 1997; Haddad, Hoddinott and Alderman, 1997; Attanasio and Lechene, 2002; Duflo, 2003; Smith et al., 2003; Maitra, 2004; Qian, 2008; Bobonis, 2009; Jensen and Oster, 2009; Doepke and Tertilt, 2019; Kandpal and Baylis, 2019). In this literature, women are often found to invest more resources than men into children's health and education, provided they are sufficiently empowered to make these decisions. In the context of this study, more empowered women may choose to spend their time on book sharing or change the way they cook so that they can use the Lucky Iron Leaf. Further, more empowered women can be more confident and engaged in the dialogic reading trainings, which may improve their understanding of dialogic reading concepts and, thus, the quality of book sharing.

3. *Children's sex*
   Differential investments into children by sex and the impact of this on children's outcomes have been widely documented in India. Based on strong patriarchal gender norms, from the very first months daughters are breastfed for shorter time, receive less childcare time, vaccinations, and vitamin supplements, are less likely to be hospitalized, are shorter, have fewer cognitive skills and suffer excess mortality (e.g. Klasen and Wink, 2002, 2003; Jha et al., 2006; Oster, 2009; Asfaw, Lamanna and Klasen, 2010; Bhalotra and Cochrane, 2010; Jayachandran and Kuziemko, 2011; Barcellos, Carvalho and Lleras-Muney, 2014; Bongaarts and Guilmoto, 2015; Jayachandran and

[17] In the estimation of the production function in Colombia, Attanasio et al. (2020*b*) use mother's education, vocabulary, Raven's score, the number of adult books and the number of magazines and newspapers in the home to proxy for mother's cognition.

[18] Other responses include sibling in 37% of book sharing households, the father in 35%, and the mother-in-law in 21%.

Pande, 2017; Ebert and Vollmer, 2022). Therefore, caregivers may also choose to invest more or better quality book sharing with their sons than with their daughters or be more inclined to prepare fortified meals for their sons than for their daughters. We consider these gender norms, proxied by children's sex, as an environmental factor, V_t, in the human capital production function.

4. *Children's anemia status*
Children's anemia status, as a proxy for iron deficiency, can matter along two dimensions for the effectiveness of the interventions. The first dimension follows from the production technology for human capital. Biologically, iron deficiency can have adverse effects on children's health and behavior. Iron deficient children tire more easily, are more wary, more hesitant, less happy and have fewer social interactions (Lozoff et al., 1998; Lozoff, 2007, 2011). However, children's interest and active engagement during book sharing is key for children to learn and, thus, improvements in mental development from book sharing may be compromised among iron deficient children (Arnold et al., 1994; Jalongo, 2004; Ortiz, Stowe and Arnold, 2001). In the human capital production function of section II we describe this as the complementarity between children's inital health endowment, H_t, and parental investments, I_t. In a previous estimation of a production function of child development in India, it has been shown that good health complements parental investment in children (Attanasio, Meghir and Nix, 2020). The second dimension along which children's baseline anemia status may affect children's hemoglobin level and mental development at endline, is parents' behavioral response to the results of their children's baseline anemia testing.[19] Information about, or being reminded of, a child's anemia status may motivate parents to learn about and utilize the Lucky Iron Leaf.

The results by mothers' education and empowerment and children's sex do not confirm the hypotheses derived from the literature. The interaction coefficients of the dialogic reading intervention groups with maternal education are mostly negative and not significant (see appendix Table A.IV.1).[20] Similarly, the coefficients by maternal empowerment show no consistent patterns and are insignificant (see appendix Table A.IV.1). The heterogeneity analysis by children's sex provides some evidence opposing our third hypothesis (see appendix Table A.IV.2). Although not significant after multiple hypothesis testing, girls in the dialogic reading and Lucky Iron Leaf groups improve in motor development (0.28 SD and 0.20 SD respectively), whereas boys do not. One explanation may be that boys have less potential

[19]Every participant, irrespective of treatment assignment, was informed about the result of their or their children's hemoglobin measurement and anemia status.

[20]The coefficients on the interaction of the Lucky Iron Leaf intervention with maternal education are positive and sizeable for cognitive and receptive language skills (0.40 and 0.51 SD) and negative for hemoglobin (0.50 g/dl), but are insignificant after multiple hypothesis testing. The coefficients' size and significance decrease when the main effects are added such that the effect on hemoglobin is not significant even before multiple hypothesis testing. If we wanted to interpret the effects on cognition and receptive language, then these results could only stem from temporary improvements in iron deficiency, which later reversed. However, given the inconsistency of these results, we refrain from drawing any conclusions.

than girls to improve their motor skills, which is in line with the observed boy-premium in motor development (appendix Table A.IV.2). The same treatment intensity across sexes (or even lower treatment intensity among girls) could then lead to a larger increase in motor skills among girls than boys.

The heterogeneous treatment effects by children's baseline anemia status provide compelling insights. Non-anemic children, who make up 30 percent of the baseline sample, gain 0.54 standard deviations in receptive language skills from the combined intervention, whereas anemic children do not benefit at all (see appendix Table A.IV.2). The effect is significant at the five percent equivalent level after multiple hypothesis testing. The estimates of the stand-alone dialogic reading treatment on receptive language skills and the combined intervention on cognitive skills for non-anemic children are also considerable in size (0.42 SD and 0.45 SD, respectively), but are no longer statistically significant after multiple hypothesis testing.

Figure 2 is a visualization of the treatment effect heterogeneity of the combined intervention on receptive language skills for three ranges of hemoglobin levels: < 9 g/dl for moderate or severe anemia, 9 to < 11 g/dl for mild anemia and ≥ 11 g/dl for no anemia. The marginal treatment effects for these hemoglobin level ranges are estimated based on a model in which each treatment indicator is interacted with an indicator for the respective hemoglobin level range. The top panel of Figure 2 displays these treatment effects and the bottom panel shows the kernel density distribution of hemoglobin as well as the normal density distribution for comparison reasons.

The figure shows zero effects of the combined intervention for anemic children and significant large effects for non-anemic children. This result suggests that health endowment complements parental investment. The fact that we find a salient effect in the combined intervention group further suggests that different investments complement each other in a dynamic way: the Lucky Iron Leaf increases hemoglobin levels and avoids anemia which makes dialogic reading more productive. Non-anemic children likely gain more from dialogic reading because they can engage more actively in book sharing in comparison to anemic children who tend to be tired and less explorative. We examine potential mechanisms of the link between baseline anemia and intervention effectiveness in detail in section VI.D below. Another reason why we find a salient effect in the combined intervention group might be that the additional home visit to distribute the Lucky Iron Leaf, which was focused primarily on the child's well-being, reinforced the dialogic reading investment.

We do not observe heterogeneous treatment effects of the Lucky Iron Leaf by children's baseline anemia status on hemoglobin levels at endline. Therefore, we reject the hypothesis that caregivers of anemic children are more motivated to learn about and utilize the Lucky Iron Leaf. In fact, the coefficients rather suggest improvements in hemoglobin from iron fortification among non-anemic children and adverse effects among anemic children, consistent with a complementarity of health with iron fortified feeding, though none of the coefficients are significant (see appendix Table A.IV.2).

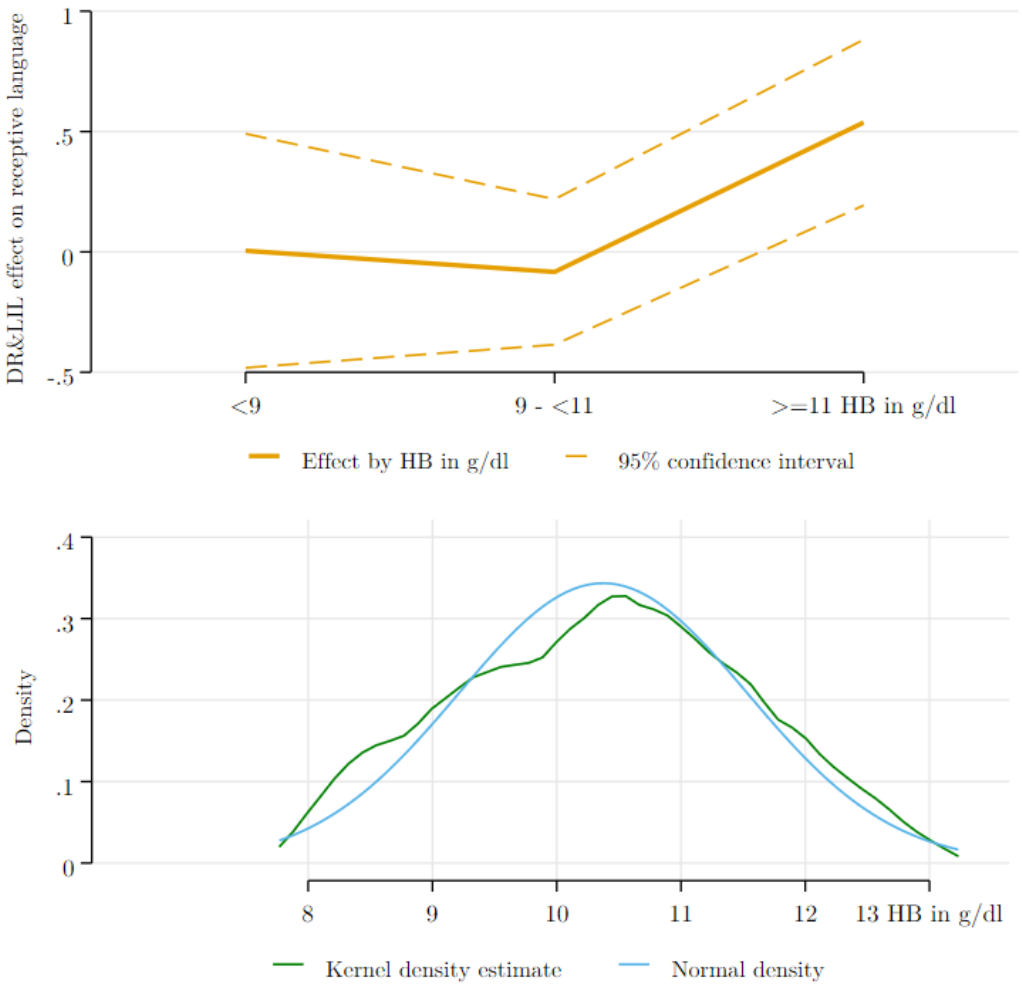


Figure 2: Effect of combined intervention on receptive language by baseline hemoglobin in g/dl

Note: Top graph shows marginal effects of combined intervention treatment on receptive language by hemoglobin levels in g/dl (HB) and 95% confidence interval (dashed line). The y-axis shows the effect on receptive language in standard deviations. HB levels on the x-axis were rounded to the nearest integer. HB levels smaller than 8 and larger than 13 were combined to ≥ 8 and ≤ 13, respectively. Bottom graph shows density distribution of HB levels based on Epanechnikov kernel with bandwidth 0.2331 and the normal density distribution.

VI.D The complementarity of children's health with parental investments

The heterogeneous treatment effect by children's anemia status at baseline implies that children's health endowment complements dialogic reading. However, anemia status may also correlate with a number of other characteristics, which could confound the heterogeneous treatment effect by anemia and, therefore, undermine the complementarity hypothesis. We test this hypothesis in three ways. First, we investigate directly whether non-anemic children engage more actively in book sharing. Specifically, we regress baseline anemia on self-reported book sharing characteristics at endline. Column (1) of Table IV shows that

Table IV: Anemia and book sharing characteristics

	Child takes initiative	BS $\geq$ 4 times past year	BS in past 7 days	Mother enjoys	Believe: beneficial
Anemic at baseline	-0.097**	0.042	-0.076*	-0.016	0.117**
	(0.038)	(0.049)	(0.043)	(0.057)	(0.058)
Controls	✓	✓	✓	✓	✓
Observations	396	442	440	317	317
Adjusted R^2	0.04	0.03	0.01	-0.00	0.04
Mean non-anemic	0.88	0.58	0.26	0.37	0.33

Note: Dependent variables are binary indicators reported by the caregiver at endline. Outcome (1) indicates whether child takes the initiative to share books. Outcome (2) indicates whether book sharing has been practiced at least four times per week in the past 12 months. Outcome (3) indicates whether books have been shared with the child in the past seven days. Outcome (4) indicates whether the caregiver enjoys book sharing and outcome (5) whether the caregiver believes book sharing is beneficial for the child. Control variables are maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. The estimation sample is restricted to book sharing households with non-missing anemia information. Standard errors are in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01.

non-anemic children tend to take the initiative for book sharing more often than anemic children. Whereas anemia at baseline does not affect book sharing utilization in the past year (column 2), column (3) suggests that anemia affects sustained utilization, i.e., whether book sharing was conducted in the past seven days. Further, mothers of anemic children enjoy book sharing to a similar extent as mothers of non-anemic children, but more of them believe that book sharing is beneficial for their children. These results suggest that mothers of non-anemic children do not behave in a way that would support the learning outcomes of their children, but that the engagement and behavior of their children themselves is relevant. These results, therefore, support the hypothesis of complementarity of baseline health and parental investments.

Second, we check the baseline balance of receptive language by treatment status within the sample of anemic and non-anemic children. If non-anemic children in the treatment group have higher receptive language skills than children in the control group, the heterogeneous effect might stem from that imbalance. However, the data shows the opposite: non-anemic children in the stand-alone and combined dialogic reading groups perform significantly worse than non-anemic control group children at baseline. There are no significant imbalances in the Lucky Iron Leaf group or among anemic children.

Third, we regress baseline anemia on a number of potential baseline correlates and then reestimate the heterogeneous treatment effect by anemia on receptive language, conditioning on the identified significant correlates of anemia. We run a linear probability model of anemia on (1) socioeconomic status variables, (2) maternal background characteristics, (3) child background characteristics, (4) home environment indicators, (5) baseline skill outcomes, and (6) baseline anthropometric measures, while holding constant our regular set of covariates. We find that maternal age, an indicator of good and bad educational measures, and the height-for-age z-score significantly correlate with anemia (see appendix Table

A.IV.3). When we reestimate the heterogeneous treatment effects by anemia on receptive language and add the identified correlates of anemia as covariates to the model, the effect for non-anemic children of the combined intervention group reduces from 0.54 to 0.49 standard deviations (p-value=0.009) (see appendix Table A.IV.4). Partially the reduction in the coefficient and precision is caused by a reduction in sample size by 34 observations. The effects for non-anemic children in other treatment groups increase slightly. Based on these results, we are not concerned that confounding factors are significant drivers of the heterogeneous treatment effect by anemia.

VII UTILIZATION AND SPILLOVERS

VII.A Utilization rates

The interventions offer tools and information to improve early childhood development but we cannot enforce strict compliance. Thus, the effectiveness of the interventions relies on the utilization frequency. Panel A of Table V presents implementation success and take-up rates of the dialogic reading intervention, which are similar across the stand-alone and combined intervention groups. Almost all households participated in at least one training session. The average total number of dialogic reading trainings received by households is three-and-a-half. With respect to book sharing frequency, about 60 percent of households reported having shared books at least four times per week since program implementation. A much smaller fraction (21%) reported to have shared books in the week prior to endline data collection and were able to present the book at endline. Further, almost 60 percent of households did not have any intervention book at home at endline and 66 percent of those reported that the books broke while sharing or playing. This suggests a considerable decline in utilization throughout the year. Among non-compliers, the two main barriers to take-up were insufficient time for sharing books (44%) and perceiving the child as too young (32%).

Among dialogic reading compliers the knowledge of contents of dialogic reading trainings and picture books was fair eight months after the last dialogic reading training was completed. Ninety-four percent of caregivers were able to freely name a specific content, character or page in their favorite book and 64 percent identified all intervention-book covers correctly when presented to them together with three unknown book covers. Only five percent of respondents could not name any of the twelve dialogic reading concepts taught.

Panel B of Table V presents take-up rates of the Lucky Iron Leaf intervention. Whereas almost all households received the fortification tool, its utilization was low. In one quarter of households the Lucky Iron Leaf was used for cooking and children also regularly ate the fortified meals. Only three percent of households had used the Lucky Iron Leaf for cooking in the previous seven days and were able to show the Lucky Iron Leaf to the enumerator at endline. The two main reasons for non-compliance were the husbands' or parents-in-law's dislike of the Lucky Iron Leaf (21%) and the unavailability of the fruit-acid-rich food items at home, which need to be added to the boiling water with the Lucky Iron Leaf (21%). Other common barriers to take-up were the necessity for prolonged cooking (11%), loss of the

Table V: Intervention take-up rates by experimental arm and intervention

	Dialogic reading				DR & LIL			
	All		≥1 training		All		≥1 training	
	Mean	N	Mean	N	Mean	N	Mean	N
Panel A: dialogic reading								
Received ≥1 training	0.93	379	1.00	354	0.95	367	1.00	350
No. of trainings (max=4)	3.42	379	3.67	354	3.50	367	3.67	350
Any book at endline	0.41	347	0.44	324	0.43	330	0.44	318
BS ≥4 per week past 12 mon.	0.59	335	0.63	314	0.58	330	0.60	317
BS past 7 days & book present	0.19	337	0.20	316	0.23	327	0.24	314
	Lucky Iron Leaf				DR & LIL			
	All		Received		All		Received	
	Mean	N	Mean	N	Mean	N	Mean	N
Panel B: Lucky Iron Leaf								
LIL received	0.95	365	1.00	347	0.97	348	1.00	338
LIL present at endline	0.35	342	0.37	312	0.33	335	0.34	311
LIL ≥4 per week past 12 mon.	0.37	333	0.38	305	0.35	324	0.37	301
LIL ≥4 and child eats food	0.27	334	0.28	306	0.22	327	0.24	304
LIL past 7 days & LIL present	0.03	340	0.03	310	0.03	335	0.03	311
					DR & LIL			
					All		Received	
					Mean	N	Mean	N
Panel C: DR and LIL								
≥1 training and LIL received					0.89	367	1.00	326
≥1 book and LIL present at endline					0.18	330	0.19	300
BS & LIL ≥4 per week and child eats food					0.14	326	0.16	293

Note: **Panel A** shows compliance with the dialogic reading intervention in a sample of all households in the dialogic reading arm (columns (1) and (2)), in a sample of households in the dialogic reading arm that participated in at least one dialogic reading training session (columns (3) and (4)), in a sample of all households in the combined treatment arm (columns (5) and (6)), and in a sample of households in the combined treatment arm that participated in at least one dialogic reading training session (columns (7) and (8)). **Panel B** shows compliance with the Lucky Iron Leaf intervention in a sample of all households in the Lucky Iron Leaf arm (columns (1) and (2)), in a sample of households in the Lucky Iron Leaf arm that received a Lucky Iron Leaf during intervention implementation (columns (3) and (4)), in a sample of all households in the combined treatment arm (columns (5) and (6)), and in a sample of households in the combined treatment arm that received a Lucky Iron Leaf during intervention implementation (columns (7) and (8)). **Panel C** shows compliance with with both interventions simultaneously in a sample of all households in the combined treatment arm (columns (5) and (6)), and in a sample of households in the combined treatment arm that participated in at least one dialogic reading training session and received a Lucky Iron Leaf during intervention implementation (columns (7) and (8)).

Lucky Iron Leaf (10%), and deficits in knowledge about its use and purpose (8%).

Panel C describes compliance with both interventions simultaneously in the combined treatment arm. Utilization rates are similar to those in the individual treatment groups, presented in Panels A and B. Further, in results not shown here, we find that across treatment

arms compliance does not differ systematically by socioeconomic characteristics or children's anemia status.

Based on the presented utilization rates, we conjecture that the effectiveness of the Lucky Iron Leaf intervention was essentially compromised by non-compliance with the intervention. Reasons and contextual barriers for non-compliance are discussed in detail in section IX.

VII.B Does the dialogic reading calendar increase book reading?

We test the effectiveness of the dialogic reading calendar in increasing book sharing utilization. The calendar was distributed during dialogic reading sessions 3 and 4 in March 2017 and participation in those sessions was conditional on having participated in at least one dialogic reading training session in December 2016. Therefore, the estimation sample is restricted to dialogic reading households that were scheduled for revisit in March 2017 – i.e. who had a chance to receive the calendar – and to the main intention-to-treat estimation sample of Table II. For reasons of statistical power, we pool the pure dialogic reading and combined intervention group households to one calendar treatment arm and one no-calendar control arm.[21] Therefore, the calendar treatment effect presents a mixed effect of the treatment effect for children in households that have received the Lucky Iron Leaf and those who have not. The F-test for group balance shows that the dialogic reading calendar arm is comparable to the dialogic reading no-calendar arm.

Columns (1) and (2) of Table VI show no effect of the calendar on having shared books at least four times per week in the past year or whether books were shared in the week prior to the endline data collection. The estimations control for dialogic reading group (i.e. pure vs. combined) and the number of training sessions attended. Adding covariates or removing the sample restriction does not change the results. We conclude that the calendar did not improve utilization of the dialogic reading intervention.

VII.C Spillovers

Because treatments were individually assigned, about 38 percent of treatment households have at least one control household within 100 meters distance. Thus, spillovers could possibly challenge the internal validity of our experiment. What we call spillovers in subsequent paragraphs may refer to externalities, contamination or John Henry effects. For example, caregivers exchanging information about techniques or the importance of picture book sharing may result in positive treatment externalities. Such externalities would bias the treatment effect if spillovers at the extensive margin are different to those at the intensive margin – e.g., the effect of exchanging dialogic reading knowledge would be larger for children of caregivers with no previous experience (control households) than among treatment peers.

[21] Results of a specification that separately includes terms for calendar without the Lucky Iron Leaf and with the Lucky Iron Leaf, i.e., interacting the calendar treatment indicator with the Lucky Iron Leaf treatment indicator, suggests no differential calendar-treatment effect by Lucky Iron Leaf intervention status.

Table VI: Intervention take-up by calendar assignment

	≥ 4 times	Past 7 days
Calendar	-0.011	0.040
	(0.042)	(0.036)
DR group	✓	✓
No. DR sessions	✓	✓
Observations	539	542
Adjusted R^2	-0.003	0.010
Control mean	0.63	0.20

Note: Dependent variables are binary, indicating whether books were shared at least 4 times per week in the past year (columns (1)) and whether books were shared in the week prior to endline data collection (column (2)). Control variables are whether the child's family was assigned to the pure dialogic reading or combined intervention group and the number of dialogic reading training sessions attended. The sample is restricted to dialogic reading households that received dialogic reading training 3 or 4 and the main estimation sample, i.e. non-missing child development test scores and covariates of the ITT estimation in Table II. Standard errors are in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Treatment contamination may occur, for example, when treatment households give away intervention picture books to control households. John Henry effects may arise when control group households attempt to compensate for the lack of treatment. Both contamination and John Henry effects would lead to downward bias in the treatment effect estimates.

To quantify potential spillover effects, we rerun our intention-to-treat estimations and add the saturation rate in treated units within 100 meters distance of each household, following Baird et al. (2016). The saturation rate is the number of households within the specified distance that received each intervention divided by all study households in the same distance.[22]

We estimate the pure treatment effect of each treatment arm $T = \{DR, LIL, DR\&LIL\}$ and spillover effects the following way:

$$y_i^{total} = \alpha_T + \beta_1^T Treated_i^T + \sum_{L=1}^{2} \beta_2^{L,T} Sat_i^L + \sum_{L=1}^{2} \beta_3^{L,T} Treated_i^T \times Sat_i^L + \delta^T Pop_i + X_i^{'}\gamma + u_{iT}. \quad (4)$$

For the sake of readability, we run the regression separately for each treatment group and on the total development score, y_i^{total} (standardized score of the sum of the five development dimensions). Sat_i^L refers to the saturation rate of household i with respect to exposure to intervention L, where $L = \{1,2\} = \{DR, LIL\}$. Pop_i presents the total number of households within 100 meters and thus controls for the population density. The control vector $X_i^{'}$ includes the standard set of covariates.

[22]Ideally, the denominator would be all households in the specified distance, but this information is not available to us. However, we expect the number of study households to be proportional to the total population.

Table VII: Intention-to-treat and spillover effects for total development z-scores

	DR subsample		LIL subsample		DR&LIL subsample	
	Total score	Total score	Total score	Total score	Total score	Total score
DR	0.265*	0.102				
	(0.141)	(0.082)				
LIL			0.219	0.047		
			(0.139)	(0.083)		
DR&LIL					0.203	0.073
					(0.144)	(0.086)
DR saturation	0.470‡‡**		0.447‡***		0.490‡‡*	
	(0.155)		(0.158)		(0.157)	
DR sat. x DR	-0.597**					
	(0.235)					
DR sat. x LIL			-0.537**			
			(0.247)			
DR sat. x DR&LIL					-0.719‡***	
					(0.252)	
LIL saturation	-0.215		-0.238		-0.189	
	(0.171)		(0.172)		(0.173)	
LIL sat. x DR	0.141					
	(0.236)					
LIL sat. x LIL			0.023			
			(0.258)			
LIL sat. x DR&LIL					0.355	
					(0.243)	
Controls	✓	✓	✓	✓	✓	✓
Observations	547	547	541	541	531	531
Adjusted R^2	0.096	0.087	0.105	0.095	0.079	0.068
Control mean	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00

Note: The DR subsample (columns (1) and (2)) consists of households of the stand-alone dialogic reading group and the control group. The LIL subsample (columns (3) and (4)) consists of households of the Lucky Iron Leaf group and the control group. The DR&LIL subsample (columns (5) and (6)) consists of households of the combined intervention group and the control group. Dependent variable is the total standardized development score (including cognitive, language, motor and socioemotional skills) and effect sizes are shown in standard deviations. Control variables are maternal education, reading ability and empowerment, the child's sex, population density and subdistrict fixed effects. Standard errors are in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01. Multiple testing corrected significance levels:$_{+}$ p<0.0096, $_{++}$ p<0.0048, $_{+++}$ p<0.0010.

β_1^T presents the treatment effect of intervention T for households without a nearby study neighbor and net of spillover effects. $\beta_2^{L,T}$ presents the spillover effect received by control households from intervention L. $\beta_3^{L,T}$ presents the additional spillover effect received by treatment arm T households from intervention L. If $\beta_3^{L,T}$ was equal to zero, control and treatment groups would be similarly affected by their exposure to intervention neighbors. δ_T presents the population density effect.

Table VII presents estimation results of the saturation model in columns (1), (3), and (5). For comparison, the even-numbered columns present the simple intention-to-treat effects, ignoring spillovers. The pure treatment effects increase to 0.20 to 0.27 standard de-

viations when we control for saturation effects. Further, the dialogic reading intervention shows consistently positive spillovers on control households. If the share of neighboring dialogic reading households among all neighboring study households within 100 meters of an untreated household increases from 0 to 1, i.e., in a 100 meter radius of an untreated household all study households are treated with dialogic reading, then the development score of children in households that were not treated increases by almost half a standard deviation. Interestingly, this effect is entirely offset for peer-dialogic reading households in the pure dialogic reading group and in the combined treatment group. The saturation effect of the Lucky Iron Leaf intervention on control households is of smaller magnitude, negative, and not statistically significant.

To examine whether the threshold of 100 meters distance is the relevant threshold, we extend equation (3) to include the saturation rates at 100 to 200 meters and 200 to 300 meters, respectively, their interaction with the treatment indicator, and the total number of households in these distances. The spillover effects for distances larger than 100 meters are small and not significant (see Appendix Figure A.V.1). Following Baird et al. (2016), we further estimate seemingly unrelated regressions (SUR) of this specification for all development outcomes. Based on F-tests of the joint significance of spillover effects received by untreated households being equal to zero, we confirm that the spillover effects for distances larger than 100 meters are zero across outcomes. The 100 meters distance threshold makes sense in the study context. Field visits showed that properties are small and children often spend time at their neighbors' or with neighboring children, however this is often focused on immediate neighbors as women's, and therefore mothers', movement is restricted by patriarchal social norms.

Additional descriptive statistics provide evidence on the pathways of spillovers. Whereas only four percent of dialogic reading households affirmed having lent books to neighbors or friends and book reading in control households increased by no more than four percentage points (10%) relative to baseline, a quarter of dialogic reading households report at endline having shared books with children of neighbors or friends (see appendix Table A.V.1). These numbers suggest that positive externalities, resulting from control household children spending time at dialogic reading households during book sharing, may have played an important role.

We conclude that children of untreated households benefited from the dialogic reading intervention implemented in neighboring households. Through this positive externality, the true intention-to-treat effect of the dialogic reading stand-alone and the combined intervention may be larger than what we are able identify in Table II.

VII.D Complier Average Causal Effects

Given imperfect utilization rates, intention-to-treat effects are not conclusive about the efficacy of the intervention. The standard procedure to estimate improvements in child outcomes for intervention compliers would be to use a two-stage least squares estimation with treatment assignment as an instrument for compliance. However, in the presence of spillovers the exclusion restriction of this estimation strategy becomes difficult to defend. If control

members compensate for the lack of treatment in ways that promote children's health and development, then the instrument, assignment to the experimental group, affects children's outcomes independent of actual compliance and, hence, violates the exclusion restriction. Therefore, we estimate complier average causal effects in a subsample of households that have no other study households within 100 meters. Because this reduces the sample size to about 430 households, we estimate the effect of dialogic reading compliance and Lucky Iron Leaf compliance, irrespective of whether participants are exposed to the intervention in either the stand-alone or combined intervention groups.

We estimate two first stages:

$$DRcomp_i = \omega_0 + \delta_1 DRany_i + \delta_2 LILany_i + X_i^{'}\lambda + \nu_i, \tag{5}$$

$$LILcomp_i = \mu_0 + \eta_1 DRany_i + \eta_2 LILany_i + X_i^{'}\kappa + \upsilon_i, \tag{6}$$

where $DRcomp_i$ and $LILcomp_i$ represent compliance with assignment to the dialogic reading intervention, $DRany_i$, and the Lucky Iron Leaf intervention, $LILany_i$, in either of the two potential treatment arms (stand-alone or combined).

The second stage is:

$$y_i = \theta_0 + \pi_1 \widehat{DRcomp_i} + \pi_2 \widehat{LILcomp_i} + X_i^{'}\rho + \varepsilon_i, \tag{7}$$

where the π-coefficients present the complier average causal effects.

In Table VIII, we define compliance among treatment group households as sharing books or using the Lucky Iron Leaf at least four times per week in the past year and conditional on the child eating the fortified meals. The instruments' relevance is confirmed by the weak identification Cragg-Donald Wald F-statistics, which range between 27 and 29.

The treatment effects for compliers in Table VIII tend to be larger than the intention-to-treat effects in Tables II and III and are partially marginally significant before multiple hypothesis testing. Dialogic reading improves cognitive skills by 0.34 standard deviations. The Lucky Iron Leaf shows large effects on hemoglobin levels (1.0 g/dl) and anemia (40 percentage points) as well as on expressive language and socioemotional skills. Although they are consistent, these results should be interpreted with caution because only 47 participants classify as Lucky Iron Leaf compliers and 125 participants as dialogic reading compliers. In a robustness check, we define compliance by the observed presence of an intervention book or Lucky Iron Leaf in homes at endline, for which we count 77 Lucky Iron Leaf compliers and 90 dialogic reading compliers. Using this observational measure of compliance avoids recall and desirability bias from self-reports. Reassuringly, employing the alternative compliance measure leads to very similar results.

Table VIII: Complier average causal effects

	Cognitive	Receptive	Expressive	Motor	Socio-emotional
DR ≥4 times per week	0.343**	0.184	0.228	0.058	0.227
	(0.173)	(0.180)	(0.185)	(0.175)	(0.178)
LIL ≥4 times per week	0.002	0.542	0.801*	0.524	0.729*
	(0.438)	(0.446)	(0.468)	(0.437)	(0.435)
Controls	✓	✓	✓	✓	✓
Cragg-Donald Wald F statistic	28	29	29	27	29
Observations	427	431	433	420	429
Control mean	-0.14	-0.19	-0.19	-0.07	-0.26
	Hemoglobin	Any anemia			
DR ≥4 times per week	0.265	-0.052			
	(0.247)	(0.088)			
LIL ≥4 times per week	1.010	-0.391*			
	(0.640)	(0.236)			
Controls	✓	✓			
HemoCue fixed effects	✓	✓			
Cragg-Donald Wald F statistic	27	27			
Observations	423	423			
Control mean	10.58	0.57			

Note: Estimation sample is restricted to households that have no other study households within 100 meters. DR and LIL compliance are defined by having used the intervention at least four times per week in the past 12 months and conditional on the child eating the fortified meals. Dependent variables are standardized development scores and effect sizes are shown in standard deviations. Control variables are maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. HemoCue machine fixed effects control for the hemoglobin measurement device and tester. Standard errors are in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01. Multiple testing corrected significance levels: $_{+}$ p<0.0096, $_{++}$ p<0.0048, $_{+++}$ p<0.0010.

VIII IMPACT ON OTHER CAREGIVING DOMAINS

The dialogic reading training may affect caregiving beyond picture book reading in complementary or substituting ways.[23] If dialogic reading and other caregiving domains are complements, we expect an increase in complementary activities or the quality of parenting measures in the dialogic reading treatment groups. If dialogic reading substitutes for other caregiver-child interactions, we expect a decrease in substituting activities in the dialogic reading treatment groups. In Table IX, we test the impact of the dialogic reading intervention on a number of caregiving activities and potential intervention mediators.

We asked all caregivers about activities they did with their children in the two weeks prior to the endline survey based on UNICEF's Multiple Indicator Cluster Survey 5. In addition to book reading, we asked about seven activities, such as playing and singing songs, and

[23]We also investigated nutritional mediators and substitution effects. However, because the Lucky Iron Leaf was essentially not utilized, we do not expect any impacts. The results in Appendix Table A.V.2 confirm the lack of effects.

summed these seven activities into an activity index ranging from zero to seven.[24] Column (1) of Table IX presents the impact on book reading (question was not specific to dialogic reading) and shows a significant increase in the probability of book reading of 14 and 11 percentage points in the stand-alone and combined dialogic reading treatment arms, respectively. In column (2), we re-code households as non-reading when they report having no books at home, to avoid social desirability bias. The effects increase dramatically in size to about 40 percentage points and provide further evidence that the intervention affected book sharing. In column (3), the dialogic reading stand-alone intervention shows a positive effect on the activity index, suggesting that caregivers perform 0.2 more activities with their children on average, in comparison to the 5.6 activities conducted in the control group. Given that the combined intervention does not show the same positive effects, the results in column (3) present weak evidence in favor of the tested activities being complementary to the dialogic reading intervention.

We further enquired about parenting measures, specifically shouting and spanking, which caregivers had been taught were harmful to children and should not be used during book sharing. Additionally, we asked about caregivers' aspirations with respect to their children's educational achievement, because the intervention trainings suggested potential gains from book sharing in children's later performance in school. However, columns (4) and (5) of Table IX show no effect of the interventions on the use of shouting or spanking and educational aspirations. Overall, these results provide limited evidence that the dialogic reading intervention affected caregiving positively or negatively beyond picture book sharing.

IX CONTEXTUAL BARRIERS TO TAKE-UP

The effectiveness of the Lucky Iron Leaf intervention was compromised by low utilization rates. Prior to the study, we conducted qualitative semi-structured interviews and a quantitative survey in the target region to investigate the likely acceptance of using a tulsi leaf-shaped iron ingot for cooking. The tulsi leaf was identified as an appropriate shape of the iron ingot as tulsi is commonly used for herbal tea – and therefore literally thrown into a pot of hot water – and 90 percent of respondents to the quantitative survey indicated that they would use the iron ingot for cooking. Previous smaller trials testing the effectiveness of the Lucky Iron Fish^{TM} (a fish-shaped iron ingot for meal fortification) in Cambodia and Peru had compliance rates of 80 percent or higher (Charles et al., 2010, 2015; Rappaport et al., 2017; Whitney et al., 2021). However, these trials had some form of follow-up either immediately after the distribution, to assist the incorporation of the Lucky Iron Fish^{TM} into daily cooking routines, or after three to six months to take blood samples and a short survey about the usage of the Lucky Iron Fish^{TM}. It could be that such follow-ups, even just for the purpose of data collection, were a factor in ensuring sustained compliance.

[24]The activity index includes the caregiver-child interactions story telling, singing songs, playing, taking the child outside of the compound, and naming objects. In addition, the index includes whether children played with (homemade) toys.

Table IX: Intention-to-treat effects on other caregiving domains

	(1) Read books	(2) Read books & has book	(3) No. of Activities	(4) Shout or spank	(5) Aspiration
DR	0.143***	0.404***	0.229**	0.021	0.121
	(0.041)	(0.033)	(0.108)	(0.035)	(0.164)
LIL	-0.010	-0.005	-0.011	0.036	-0.060
	(0.042)	(0.021)	(0.110)	(0.035)	(0.160)
DR & LIL	0.112***	0.378***	0.119	0.042	0.109
	(0.043)	(0.035)	(0.109)	(0.035)	(0.163)
Controls	✓	✓	✓	✓	✓
Observations	1114	1114	1100	1125	893
Adjusted R^2	0.04	0.23	0.03	0.00	0.12
Control mean	0.46	0.07	5.60	0.76	5.34

Note: Column (3) outcome: summation of indicators for 7 different activities recently conducted with the child. Column (4) outcome: dummy for whether mother spanks or shouts at child. Column (5) outcome: aspiration is measured by the highest level of education mothers are envisioning for their child, from 1=None to 8=Master degree or higher (5=Higher secondary). Control variables are maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Standard errors are in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01.

A common obstacle to the take-up of preventive health care products is the lack of information on products and underlying health issues (Dupas, 2011). In the case of the Lucky Iron Leaf intervention, knowledge about anemia would be a key motivation for its (continual) utilization. However, only seven percent of Lucky Iron Leaf group respondents answered positively to the question "Have you heard of anemia?" twelve months after intervention roll-out (and four percent in the control group). Respondents may not have known the term "anemia" but, nevertheless, were familiar with the concept of iron deficiency and its implications for children's health and development, as iron deficiency anemia education was part of the intervention. In light of low overall health knowledge among the study population – e.g., only five percent of the study sample had heard of under-nutrition in general and two-thirds of caregivers could not name any danger signs of malaria – one home visit may not have sufficed to raise anemia awareness in a sustainable way and, therefore, to convey the need for increased iron intake. We conclude that in contexts with low levels of education and health knowledge, more intensive or more frequent information sessions may be required to induce behavioral changes.

A frequently-discussed barrier to take-up of micronutrient fortification interventions is the alteration of taste and color of meals (de Barros and Cardoso, 2016). In addition, interventions that specifically use cast iron for fortification, e.g., fortification through iron pots, have discussed participants' dislike of rust and concerns about its health implications as potential barriers to take-up (Geerligs et al., 2002; Tripp et al., 2010; Charles et al., 2011, 2015; Alves, Saleh and Alaofè, 2019). We informed participants during the implementing home visits that rust is harmless to the body and explained how to remove rust easily from the iron ingot. Further, the Lucky Iron Leaf manual instructs users to clean and dry the iron ingot

after cooking to avoid rust. In line with these precautions, only 6.4 percent of non-compliers mentioned that they did not use the iron ingot because meals taste like iron or rust and only 3.4 percent mentioned that it looked rusty. Thus, we are not concerned that rust or alterations to the taste and color of meals deterred take-up.

Using an acidic food in the preparation of the daily fortified meal is essential for the Lucky Iron Leaf to be effective. The fruit acid causes the iron to leach from the ingot into the water. While the majority of Lucky Iron Leaf households were aware of this requirement, only a quarter reported the consumption of any tomato or lemon in the week before the end-line survey. This dietary pattern reflects the overall low consumption of fruits and vegetables in northeast India (Sharma et al., 2020). In Cambodia, where 90 percent of participants persistently used the Lucky Iron Fish^{TM}, fruit acids are regularly consumed in a variety of traditional meals such as soups soured with lemon juice (Charles et al., 2010, 2015; Rappaport et al., 2017). Therefore the behavior change required to utilize the Lucky Iron ingot is smaller in Cambodia than in India.

Lastly, women's empowerment potentially plays an important role in children's nutrition. In our study sample, only half of the mothers are involved in decisions about their children's diets. Although the intervention targeted specifically a household decision maker, in addition to the person responsible for cooking, 21 percent of caregivers did not use the Lucky Iron Leaf because their husbands or parents-in-law disliked it. Therefore, the intervention failed to convince key decision makers in a considerable number of households. In contexts where women are sufficiently empowered to make decisions, such barriers to take-up may be alleviated.

Social norms around gender roles might also have an impact with respect to the gender of the facilitator. Women as meal-makers, and therefore as main recipients of the intervention, might not have felt comfortable to talk to the male facilitator and, therefore, rejected the use of the Lucky Iron Leaf. Unfortunately, we did not ask women about the gender of the facilitator as a potential reason for non-compliance.

The discussion implies two key take-aways. First, additional follow-ups could potentially improve take-up if they include household decision makers and raise awareness about iron deficiency anemia. Follow-up visits could further support the incorporation of the Lucky Iron Leaf into daily cooking practices. If the distribution of the Lucky Iron Leaf is combined with home visits for a cognitive stimulation intervention, such as dialogic reading trainings, follow-ups could be integrated with the cognitive stimulation intervention. The distribution and follow-ups could also be integrated into the work of community health workers or with any other existing program. Second, for poor population groups, who rarely consume vegetables or fruits, the distribution of the Lucky Iron Fish may be accompanied by the distribution of powdered citric acid.

X Conclusion

We study whether two simple and short-lived interventions can reduce anemia and improve mental functions of one- to two-year old children in northeast India. The first intervention

targets iron deficiency and comprises one home visit to distribute a durable cooking tool for home iron fortification of meals, called Lucky Iron Leaf. The second intervention targets the lack of psychosocial stimulation and consists of the provision of picture books and four at-home trainings of caregivers in dialogic reading. Both interventions are remarkably simple in their delivery to households in comparison to existing effective early skill development programs. Children and their caregivers were randomly assigned to receive either, both or no intervention.

Children in the combined intervention group, who were non-anemic at baseline (30% of the baseline sample), improved their receptive language skills by 0.54 standard deviations. This result suggests that children's initial health endowment complements parental investments in the production technology of human capital. Biologically, non-anemic children likely have an advantage in the efficacy of book sharing over their anemic peers because they can engage actively and learn more, in comparison to anemic children who tend to be more tired and less exploratory.

The fact that we observe treatment effects for non-anemic children in the combined intervention provides evidence for parental investments complementing each other in a dynamic way: the Lucky Iron Leaf increases hemoglobin levels and avoids anemia which makes dialogic reading more productive. However, this result is limited by low utilitazion rates of the Lucky Iron Leaf. Utilization of the iron fortification device was reported to be 25 percent initially and decreased to 3 percent a year later. The complementarity of the two interventions, therefore, must stem from positive impacts among temporary Lucky Iron Leaf compliers. Children in families who used the Lucky Iron Leaf at least four times per week for cooking in the past year, show a reduction in the rate of anemia of 40 percent (the sample is small (N=423) and the effect is not significant anymore after multiple hypothesis testing). The home visit for the delivery of the Lucky Iron Leaf intervention, which was focused primarily on the child's well-being, may have also reinforced dialogic reading investments.

Interactions between children's health and parental investments have been previously documented. Tofail et al. (2013) find that a psychosocial stimulation intervention with 6 to 24 months old children in Bangladesh was effective for non-anemic children but not for anemic children. In the structural estimation of a human capital production function in India, Attanasio, Meghir and Nix (2020) find that good health positively affects parental investment in children as well as later health and cognitive outcomes.

Given the simplicity of our interventions, the effect size for non-anemic children is considerable. Comprehensive home-based parenting interventions patterned after the Jamaica-program with weekly home visits of 6 to 24 months improved receptive language by about 0.2 to 0.7 standard deviations and cognition by about 0.3 to 0.6 standard deviations (Attanasio et al., 2014; Yousafzai et al., 2014; Sylvia et al., 2020; Grantham-McGregor et al., 2020). The interventions tested in this study only significantly affected receptive language skills, whereas the more comprehensive home-visiting programs also improved cognition and, in some cases, expressive language and motor skills.

We further investigate spillovers of our nutrition and psychosocial stimulation interventions through neighborhood networks. We follow Baird et al. (2016) and use variation in the saturation in opposite treatment households to disentangle the pure treatment effect from

spillover effects. The pure intention-to-treat effect across treatment arms on children's mental functions increases in magnitude to 0.20 to 0.27 standard deviations. Further, we find large spillover effects of dialogic reading saturation in control households, whereas there are no spillovers of Lucky Iron Leaf saturation or among dialogic reading peers. The detected dialogic reading spillovers on control neighbors are in line with List, Momeni and Zenou (2019), who find large positive externalities of an early childhood cognitive stimulation intervention via neighborhood networks in the United States.

Our results add to a growing literature on delivering early childhood development programs at scale and suggest that more short-lived parenting interventions are potentially effective in healthier populations. Previous approaches successfully used existing infrastructure to scale intensive and long-lasting parenting programs (Attanasio et al., 2014; Yousafzai et al., 2014; Andrew et al., 2020; Attanasio et al., 2020*b*; Sylvia et al., 2020). Our approach offers an alternative program design in contexts where supply-side compliance is expected to be low. Replications of this approach are necessary to explore this alternative way to scale.

Funding

This work was supported by 3ie International Initiative for Impact Evaluation [grant number CPW.01.GV.IE]; Niedersächsisches Ministerium für Wissenschaft und Kultur [grant number VWZN3060]; Deutsche Forschungsgemeinschaft [grant numbers RTG1723, RTG1666]; and Stiftung fiat panis [grant number 38/2017].

Acknowledgements

We are grateful to Alastair Summerlee for the collaboration in the design of the Lucky Iron Leaf, its laboratory testing, and the provision of 1,000 Lucky Iron Leafs by the Lucky Iron Fish Enterprise. We are thankful to Peter Cooper, David Jeffery and the Mikhulu Child Development Trust who generously provided and adapted their dialogic reading program for this project. We thank Claudia Maehler, Marika Kisters, and Nadine Storch for adapting FREDI 0-3 to the Bihari context. We thank Maximilian Sprengholz for the design of dialogic reading handouts. We are grateful for the support and advice from Malavika Subramanyam. Finally, express our deep gratitude to Abhijeet Kumar for field work coordination and his dedication to the project. Esther Heeseman gratefully acknowledges the support of the United Nations University World Institute for Development Economics Research (UNU-WIDER).

Conflict of interest declaration

The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

REFERENCES

Afridi, Farzana. 2010. "Child welfare programs and child nutrition: Evidence from a mandated school meal program in India." *Journal of Development Economics*, 92(2): 152–165.

Aker, Jenny C, Rachid Boumnijel, Amanda McClelland, and Niall Tierney. 2012. "Zap it to me: The impacts of a mobile cash transfer program." *Unpublished manuscript*.

Alves, Clark, Ahlam Saleh, and Halimatou Alaofè. 2019. "Iron-containing cookware for the reduction of iron deficiency anemia among children and females of reproductive age in low-and middle-income countries: A systematic review." *PloS One*, 14(9): e0221094.

Andrew, Alison, Orazio Attanasio, Britta Augsburg, Monimalika Day, Sally Grantham-McGregor, Costas Meghir, Fardina Mehrin, Smriti Pahwa, and Marta Rubio-Codina. 2020. "Effects of a scalable home-visiting intervention on child development in slums of urban India: Evidence from a randomised controlled trial." *Journal of Child Psychology and Psychiatry*, 61(6): 644–652.

Armstrong, Gavin R, Cate E Dewey, and Alastair J S Summerlee. 2017. "Iron release from the Lucky Iron Fish®: safety considerations." *Asia Pacific Journal of Clinical Nutrition*, 26(1): 148–155.

Arnold, David H, Christopher J Lonigan, Grover J Whitehurst, and Jeffery N Epstein. 1994. "Accelerating language development through picture book reading: Replication and extension to a videotape training format." *Journal of Educational Psychology*, 86(2): 235–243.

Asfaw, Abay, Francesca Lamanna, and Stephan Klasen. 2010. "Gender gap in parents' financing strategy for hospitalization of their children: Evidence from India." *Health Economics*, 19(3): 265–279.

Attanasio, Orazio, and Valérie Lechene. 2002. "Tests of income pooling in household decisions." *Review of Economic Dynamics*, 5(4): 720 – 748.

Attanasio, Orazio, Camila Fernández, Emla O A Fitzsimons, Sally M Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina. 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.

Attanasio, Orazio, Costas Meghir, and Emily Nix. 2020. "Human capital development and parental investment in India." *The Review of Economic Studies*, 87(6): 2511–2541.

Attanasio, Orazio, Helen Baker-Henningham, Raquel Bernal, Costas Meghir, Diana Pineda, and Marta Rubio-Codina. 2022. "Early stimulation and nutrition: The impacts of a scalable intervention." *Journal of the European Economic Association*, 20(4): 1395–1432.

Attanasio, Orazio P. 2015. "The determinants of human capital formation during the early years of life: Theory, measurement, and policies." *Journal of the European Economic Association*, 13(6): 949–997.

Attanasio, Orazio, Raquel Bernal, Michele Giannola, and Milagros Nores. 2020*a*. "Child development in the early years: Parental investment and the changing dynamics of different dimensions." National Bureau of Economic Research.

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2020*b*. "Estimating the production function for human capital: Results from a randomized controlled trial in Colombia." *American Economic Review*, 110(1): 48–85.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. 2016. "Worms at work: Long-run impacts of a child health investment." *The Quarterly Journal of Economics*, 131(4): 1637–1680.

Barcellos, Silvia Helena, Leandro S Carvalho, and Adriana Lleras-Muney. 2014. "Child gender and parental investments in India: Are boys and girls treated differently?" *American Economic Journal: Applied Economics*, 6(1): 157–189.

Beard, John L. 2008. "Why Iron Deficiency Is Important in Infant Development." *The Journal of Nutrition*, 138(12): 2534–2536.

Bhalotra, Sonia, and Tom Cochrane. 2010. "Where have all the young girls gone? Identification of sex selection in India." Institute for the Study of Labor (IZA).

Black, Robert E, Cesar G Victora, Susan P Walker, Zulfiqar A Bhutta, Parul Christian, Mercedes De Onis, Majid Ezzati, Sally Grantham-McGregor, Joanne Katz, Reynaldo Martorell, et al. 2013. "Maternal and child undernutrition and overweight in low-income and middle-income countries." *The Lancet*, 382(9890): 427–451.

Bobonis, Gustavo J. 2009. "Is the allocation of resources within the household efficient? New evidence from a randomized experiment." *Journal of Political Economy*, 117(3): 453–503.

Bogler, Lisa, Christian Bommer, Cara Ebert, Abhijeet Kumar, SV Subramanian, Malavika A. Subramanyam, and Sebastian Vollmer. 2023. "Effects of a large-scale participatory learning and action programme in women's groups on health, nutrition, water, sanitation, and hygiene: a cluster-randomised controlled trial in Bihar, India." *Journal of Development Effectiveness*, 0(0): 1–18.

Bongaarts, John, and Christophe Z Guilmoto. 2015. "How many more missing women? Excess female mortality and prenatal sex selection, 1970–2050." *Population and Development Review*, 41(2): 241–269.

Bus, Adriana G, Marinus H van IJzendoorn, and Anthony D Pellegrini. 1995. "Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy." *Review of Educational Research*, 65(1): 1–21.

Census of India. 2011. "District Census Handbook Madhepura." Directorate of Census Operations, Bihar Series-11 Part XII-B.

Chakraborty, Tanika, and Rajshri Jayaraman. 2019. "School feeding and learning achievement: Evidence from India's midday meal program." *Journal of Development Economics*, 139: 249–265.

Charles, Christopher V, Cate E Dewey, Ann Hall, Chantharith Hak, Son Channary, and Alastair J S Summerlee. 2015. "A randomized control trial using a fish-shaped iron ingot for the amelioration of iron deficiency anemia in rural Cambodian women." *Tropical Medicine & Surgery*, 3(3): 195.

Charles, Christopher V, Cate E Dewey, William E Daniell, and Alastair JS Summerlee. 2010. "Iron-deficiency anaemia in rural Cambodia: Community trial of a novel iron supplementation technique." *European Journal of Public Health*, 21(1): 43–48.

Charles, Christopher V, Cate E Dewey, William E Daniell, and Alastair J S Summerlee. 2011. "Iron-deficiency anaemia in rural Cambodia: Community trial of a novel iron supplementation technique." *European Journal of Public Health*, 21(1): 43–48.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences 2nd edn.* Erlbaum Associates, Hillsdale.

Cooper, Peter J, Zahir Vally, Hallam Cooper, Theo Radford, Arthur Sharples, Mark Tomlinson, and Lynne Murray. 2014. "Promoting mother-infant book sharing and infant attention and language development in an impoverished South African population: A pilot study." *Early Childhood Education Journal*, 42(2): 143–152.

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

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

Dale, Philip S, Catherine Crain-Thoreson, Angela Notari-Syverson, and Kevin Cole. 1996. "Parent-child book reading as an intervention technique for young children with language delays." *Topics in Early Childhood Special Education*, 16(2): 213–235.

de Barros, Samara Fernandes, and Marly Augusto Cardoso. 2016. "Adherence to and acceptability of home fortification with vitamins and minerals in children aged 6 to 23 months: A systematic review." *BMC Public Health*, 16(1): 299.

Dewey, Kathryn G, and Seth Adu-Afarwuah. 2008. "Systematic review of the efficacy and effectiveness of complementary feeding interventions in developing countries." *Maternal & Child Nutrition*, 4: 24–85.

Doepke, Matthias, and Michele Tertilt. 2019. "Does female empowerment promote economic development?" Discussion Paper Series – CRC TR 224, No. 112.

Duflo, Esther. 2003. "Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in South Africa." *The World Bank Economic Review*, 17(1): 1–25.

Dupas, Pascaline. 2011. "Health behavior in developing countries." *Annual Review of Economics*, 3(1): 425–449.

Dupont, Caroline, Natalie Castellanos-Ryan, Jean R Séguin, Gina Muckle, Marie-Noëlle Simard, Gabriel D Shapiro, Catherine M Herba, William D Fraser, and Sarah Lippé. 2018. "The predictive value of head circumference growth during the first year of life on early child traits." *Scientific reports*, 8(1): 1–9.

Dyson, Tim, and Mick Moore. 1983. "On kinship structure, female autonomy, and demographic behavior in India." *Population and Development Review*, 35–60.

Ebert, Cara, and Sebastian Vollmer. 2022. "Girls unwanted–The role of parents' child-specific sex preference for children's early mental development." Elsevier.

Elmonayer, Randa Abdelaleem. 2013. "Promoting phonological awareness skills of Egyptian kindergarteners through dialogic reading." *Early Child Development and Care*, 183(9): 1229–1241.

Flagel, Shelly B, Jeremy J Clark, Terry E Robinson, Leah Mayo, Alayna Czuj, Ingo Willuhn, Christina A Akers, Sarah M Clinton, Paul EM Phillips, and Huda Akil. 2011. "A selective role for dopamine in stimulus–reward learning." *Nature*, 469(7328): 53–57.

Fletcher, Kathryn L, and Elaine Reese. 2005. "Picture book reading with young children: A conceptual framework." *Developmental Review*, 25(1): 64–103.

Fraker, A, N Shah, and R Abraham. 2013. "Quantitative assessment: beneficiary nutritional status & performance of ICDS Supplementary Nutrition Programme in Bihar." *International Growth Center*.

García, Jorge Luis, and James J Heckman. 2022. "Parenting promotes social mobility within and across generations." *Annual Review of Economics*, 15.

García, Jorge Luis, Frederik H Bennhoff, Duncan Ermini Leaf, and James J Heckman. 2021. "The dynastic benefits of early childhood education." National Bureau of Economic Research.

García, Jorge Luis, James J Heckman, Duncan Ermini Leaf, and María José Prados. 2020. "Quantifying the life-cycle benefits of an influential early-childhood program." *Journal of Political Economy*, 128(7): 2502–2541.

Geerligs, Paul Prinsen, Bernard Brabin, Albert Mkumbwa, Robin Broadhead, and Luis E Cuevas. 2002. "Acceptability of the use of iron cooking pots to reduce anaemia in developing countries." *Public Health Nutrition*, 5(5): 619–624.

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

Government of India and UNICEF. 2015. "Rapid Survey on Children (RSOC) 2013–14: National Report."

Grantham-McGregor, Sally, Akanksha Adya, Orazio Attanasio, Britta Augsburg, Jere Behrman, Bet Caeyers, Monimalika Day, Pamela Jervis, Reema Kochar, Prerna Makkar, Costas Meghir, Angus Phimister, Marta Rubio-Codina, and Karishma Vats. 2020. "Group sessions or home visits for early childhood development in India: A cluster RCT." *Pediatrics*.

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

Haddad, Lawrence James, John F. Hoddinott, and Harold Alderman. 1997. "Intrahousehold resource allocation in developing countries: Models, methods, and policy." International Food Policy Research Institute (IFPRI) and Johns Hopkins University Press.

Hargrave, Anne C, and Monique Sénéchal. 2000. "A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading." *Early Childhood Research Quarterly*, 15(1): 75–90.

Heckman, James J. 2008. "Schools, skills, and synapses." *Economic inquiry*, 46(3): 289–324.

Huebner, Colleen E. 2000. "Promoting toddlers' language development through community-based intervention." *Journal of Applied Developmental Psychology*, 21(5): 513–535.

Humphreys, Macartan, Raul Sanchez De la Sierra, and Peter Van der Windt. 2013. "Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration." *Political Analysis*, 1–20.

Iannotti, Lora L, James M Tielsch, Maureen M Black, and Robert E Black. 2006. "Iron supplementation in early childhood: Health benefits and risks." *The American Journal of Clinical Nutrition*, 84(6): 1261–1276.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

International Institute for Population Sciences. 2017. "National Family Health Survey (NFHS-4),2015-16: District Fact Sheet Madhepura Bihar."

International Institute for Population Sciences, and ICF. 2017. "National Family Health Survey (NFHS-4), 2015-16: India." IIPS, Mumbai.

Jalongo, Mary. 2004. *Young children and picture books.* ERIC.

Jayachandran, Seema, and Ilyana Kuziemko. 2011. "Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India." *The Quarterly Journal of Economics*, 126(3): 1485–1538.

Jayachandran, Seema, and Rohini Pande. 2017. "Why are Indian children so short? The role of birth order and son preference." *American Economic Review*, 107(9): 2600–2629.

Jensen, Robert, and Emily Oster. 2009. "The power of TV: Cable television and women's status in India." *The Quarterly Journal of Economics*, 124(3): 1057–1094.

Jha, Prabhat, Rajesh Kumar, Priya Vasa, Neeraj Dhingra, Deva Thiruchelvam, and Rahim Moineddin. 2006. "Low male-to-female sex ratio of children born in India: National survey of 1· 1 million households." *The Lancet*, 367(9506): 211–218.

Kandpal, Eeshani, and Kathy Baylis. 2019. "The social lives of married women: Peer effects in female autonomy and investments in children." *Journal of Development Economics*, 140: 26 – 43.

Kapil, Umesh, and A. Singh Bhadoria. 2014. "National Iron-plus initiative guidelines for control of iron deficiency anaemia in India, 2013." *The National Medical Journal of India*, 27(1): 27–9.

Kassebaum, Nicholas J, Rashmi Jasrasaria, Mohsen Naghavi, Sarah K Wulf, Nicole Johns, Rafael Lozano, Mathilda Regan, David Weatherall, David P Chou, Thomas P Eisele, et al. 2014. "A systematic analysis of global anemia burden from 1990 to 2010." *Blood*, 123(5): 615–624.

Kiese-Himmel, Christiane. 2013. "Elternfragebogen zur Wortschatzentwicklung im frühen Kindesalter: Eltern Antworten – Revision (ELAN-R)." In *Dorsch Lexikon der Psychologie*. , ed. Markus A. Wirtz, 364. Verlag Hans Huber.

Kiese-Himmel, Christiane. 2014. "Sprachentwicklungstest für zweijährige Kinder (SETK-2)." In *Dorsch Lexikon der Psychologie*. , ed. Markus A. Wirtz, 364. Verlag Hans Huber.

Klasen, Stephan, and Claudia Wink. 2002. "A turning point in gender bias in mortality? An update on the number of missing women." *Population and Development Review*, 28(2): 285–312.

Klasen, Stephan, and Claudia Wink. 2003. ""Missing women": Revisiting the debate." *Feminist Economics*, 9(2-3): 263–299.

Knauer, Heather A, Pamela Jakiela, Owen Ozier, Frances Aboud, and Lia CH Fernald. 2020. "Enhancing young children's language acquisition through parent–child book-sharing: A randomized trial in rural Kenya." *Early Childhood Research Quarterly*, 50: 179–190.

Kochar, Priyanka S, Rakhi Dandona, G Anil Kumar, and Lalit Dandona. 2014. "Population-based estimates of still birth, induced abortion and miscarriage in the Indian state of Bihar." *BMC Pregnancy and Childbirth*, 14(1): 413.

Krämer, Marion, Santosh Kumar, and Sebastian Vollmer. 2021*a*. "Anemia, diet, and cognitive development: Impact of health information on diet quality and child nutrition in rural India." *Journal of Economic Behavior Organization*, 190: 495–523.

Krämer, Marion, Santosh Kumar, and Sebastian Vollmer. 2021*b*. "Improving child health and cognition: Evidence from a school-based nutrition intervention in India." *The Review of Economics and Statistics*, 103(5): 818–834.

Landry, Susan H, Karen E Smith, and Paul R Swank. 2006. "Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills." *Developmental Psychology*, 42(4): 627.

Lever, Rosemary, and Monique Sénéchal. 2011. "Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction." *Journal of Experimental Child Psychology*, 108(1): 1–24.

List, John A, Fatemeh Momeni, and Yves Zenou. 2019. "Are estimates of early education programs too pessimistic? Evidence from a large-scale field experiment that causally measures neighbor effects." *CEPR Discussion Paper No. DP13725*.

Lonigan, Christopher J, and Grover J Whitehurst. 1998. "Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds." *Early Childhood Research Quarterly*, 13(2): 263–290.

Lozoff, Betsy. 2007. "Iron deficiency and child development." *Food and Nutrition Bulletin*, 28(4_suppl4): S560–S571.

Lozoff, Betsy. 2011. "Early iron deficiency has brain and behavior effects consistent with dopaminergic dysfunction." *The Journal of Nutrition*, 141(4): 740S–746S.

Lozoff, Betsy, Nancy K Klein, Edward C Nelson, Donna K McClish, Martin Manuel, and Maria Elena Chacon. 1998. "Behavior of infants with iron-deficiency anemia." *Child Development*, 69(1): 24–36.

Lundberg, Shelly J, Robert A Pollak, and Terence J Wales. 1997. "Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit." *Journal of Human Resources*, 463–480.

Macha, Thorsten, and Franz Petermann. 2017. "FREDI 0 – 3." *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 49(1): 50–56.

Mähler, Claudia, Friederike Cartschau, and Katharina Rohleder. 2016. *Frühkindliche Entwicklungsdiagnostik für Kinder von 0 - 3 Jahren (FREDI 0-3).* Göttingen:Hogrefe.

Maitra, Pushkar. 2004. "Parental bargaining, health inputs and child mortality in India." *Journal of Health Economics*, 23(2): 259–291.

Miguel, Edward, Colin Camerer, Katherine Casey, Joshua Cohen, Kevin M Esterling, Alan Gerber, Rachel Glennerster, Don P Green, Macartan Humphreys, Guido Imbens, et al. 2014. "Promoting transparency in social science research." *Science*, 343(6166): 30–31.

Murray, Lynne, Leonardo De Pascalis, Mark Tomlinson, Zahir Vally, Harold Dadomo, Brenda MacLachlan, Charlotte Woodward, and Peter J Cooper. 2016. "Randomized controlled trial of a book-sharing intervention in a deprived South African community: Effects on carer-infant interactions, and their relation to infant cognitive and socioemotional outcome." *Journal of Child Psychology and Psychiatry*, 57(12): 1370–1379.

Ofosu, George K, and Daniel N Posner. 2020. "Do pre-analysis plans hamper publication?" Vol. 110, 70–74.

Opel, Aftab, Syeda Saadia Ameer, and Frances E Aboud. 2009. "The effect of preschool dialogic reading on vocabulary among rural Bangladeshi children." *International Journal of Educational Research*, 48(1): 12–20.

Ortiz, Camilo, Rebecca M Stowe, and David H Arnold. 2001. "Parental influence on child interest in shared picture book reading." *Early Childhood Research Quarterly*, 16(2): 263–281.

Oster, Emily. 2009. "Proximate sources of population sex imbalance in India." *Demography*, 46(2): 325–339.

Qian, Nancy. 2008. "Missing women and the price of tea in China: The effect of sex-specific earnings on sex imbalance." *The Quarterly Journal of Economics*, 123(3): 1251–1285.

Rappaport, Aviva I, Kyly C Whitfield, Gwen E Chapman, Rickey Y Yada, Khin Meng Kheang, Jennie Louise, Alastair J Summerlee, Gavin R Armstrong, and Timothy J Green. 2017. "Randomized controlled trial assessing the efficacy of a reusable fish-shaped iron ingot to increase hemoglobin concentration in anemic, rural Cambodian women." *The American Journal of Clinical Nutrition*, 106(2): 667–674.

Richter, Linda M, Bernadette Daelmans, Joan Lombardi, Jody Heymann, Florencia Lopez Boo, Jere R Behrman, Chunling Lu, Jane E Lucas, Rafael Perez-Escamilla, Tarun Dua, et al. 2017. "Investing in the foundation of sustainable development: Pathways to scale up for early childhood development." *The Lancet*, 389(10064): 103–118.

Rivera, Juan A, Christine Hotz, Teresa González-Cossío, Lynnette Neufeld, and Armando García-Guerra. 2003. "The effect of micronutrient deficiencies on child growth: A review of results from community-based supplementation trials." *The Journal of Nutrition*, 133(11): 4010S–4020S.

Sankoh, Abdul J, Mohammad F Huque, and Satya D Dubey. 1997. "Some comments on frequently used multiple endpoint adjustment methods in clinical trials." *Statistics in Medicine*, 16(22): 2529–2542.

Sharma, Manika, Avinash Kishore, Devesh Roy, and Kuhu Joshi. 2020. "A comparison of the Indian diet with the EAT-Lancet reference diet." *BMC Public Health*, 20: 1–13.

Simsek, Zeynep Ceren, and Nesrin Isıkoglu Erdogan. 2015. "Effects of the dialogic and traditional reading techniques on children's language development." *Procedia-Social and Behavioral Sciences*, 197: 754–758.

Smith, Lisa C, Usha Ramakrishnan, Aida Ndiaye, Lawrence Haddad, and Reynaldo Martorell. 2003. "The importance of women's status for child nutrition in developing countries: International food policy research institute (IFPRI) research report abstract 131." *Food and Nutrition Bulletin*, 24(3): 287–288.

Sylvia, Sean, Nele Warrinnier, Renfu Luo, Ai Yue, Orazio Attanasio, Alexis Medina, and Scott Rozelle. 2020. "From quantity to quality: Delivering a home-based parenting intervention through China's Family Planning Cadres." *The Economic Journal*. ueaa114.

Thomas, Duncan. 1993. "The distribution of income and expenditure within the household." *Annals of Economics and Statistics*, , (29): 109–135.

Tofail, Fahmida, Jena D Hamadani, Fardina Mehrin, Deborah A Ridout, Syed N Huda, and Sally M Grantham-McGregor. 2013. "Psychosocial stimulation benefits development in nonanemic children but not in anemic, iron-deficient children." *The Journal of Nutrition*, 143(6): 885–893.

Tripp, Katherine, Nancy MacKeith, Bradley A Woodruff, Leisel Talley, Laurent Mselle, Zahra Mirghani, Fathia Abdalla, Rita Bhatia, and Andrew J Seal. 2010. "Acceptability and use of iron and iron-alloy cooking pots: Implications for anaemia control programmes." *Public Health Nutrition*, 13(1): 123–130.

Valdez-Menchaca, Marta C, and Grover J Whitehurst. 1992. "Accelerating language development through picture book reading: A systematic extension to Mexican day care." *Developmental Psychology*, 28(6): 1106.

Vally, Zahir, Lynne Murray, Mark Tomlinson, and Peter J Cooper. 2015. "The impact of dialogic book-sharing training on infant language and attention: A randomized controlled trial in a deprived South African community." *Journal of Child Psychology and Psychiatry*, 56(8): 865–873.

Walker, Susan P, Christine A Powell, Sally M Grantham-McGregor, John H Himes, and Susan M Chang. 1991. "Nutritional supplementation, psychosocial stimulation, and growth of stunted children: The Jamaican study." *The American Journal of Clinical Nutrition*, 54(4): 642–648.

Walker, Susan P, Susan M Chang, Amika S Wright, Rodrigo Pinto, James J Heckman, and Sally M Grantham-McGregor. 2022. "Cognitive, psychosocial, and behaviour gains at age 31 years from the Jamaica early childhood stimulation trial." *Journal of Child Psychology and Psychiatry*, 63(6): 626–635.

Walker, Susan P, Theodore D Wachs, Julie Meeks Gardner, Betsy Lozoff, Gail A Wasserman, Ernesto Pollitt, and Julie A Carter. 2007. "Child development: Risk factors for adverse outcomes in developing countries." *The Lancet*, 369(9556): 145–157.

Walker, Susan P, Theodore D Wachs, Sally Grantham-McGregor, Maureen M Black, Charles A Nelson, Sandra L Huffman, Helen Baker-Henningham, Susan M Chang, Jena D Hamadani, Betsy Lozoff, Julie M Meeks Gardner, C A Powell, Atif Rahman, and Linda Richter. 2011. "Inequality in early childhood: Risk and protective factors for early child development." *The Lancet*, 378(9799): 1325–1338.

Weisleder, Adriana, and Anne Fernald. 2013. "Talking to children matters: Early language experience strengthens processing and builds vocabulary." *Psychological Science*, 24(11): 2143–2152.

Whitehurst, Grover J, David H Arnold, Jeffery N Epstein, Andrea L Angell, Meagan Smith, and Janet E Fischel. 1994. "A picture book reading intervention in day care and home for children from low-income families." *Developmental Psychology*, 30(5): 679–689.

Whitehurst, Grover J, F L Falco, C J Lonigan, J E Fischel, B D DeBaryshe, and M C Valdez-Menchaca. 1988. "Accelerating language development through picture book reading." *Developmental Psychology*, 24(4): 552–559.

Whitney, Rachel, Wayne A Centrone, Hoover Supo Mamani, Karen Falkenstein, Robert S Levine, Jaden Harris, Melissa Hazlitt, and Czer Anthoney Lim. 2021. "Impact of a collaborative childhood anaemia intervention programme in Peru." *Tropical Medicine & International Health*.

Wise, Roy A. 2004. "Dopamine, learning and motivation." *Nature Reviews Neuroscience*, 5(6): 483–494.

World Health Organization. 2007. "WHO child growth standards: head circumference-for-age, arm circumference-for-age, triceps skinfold-for-age and subscapular skinfold-for-age: Methods and development." World Health Organization.

World Health Organization. 2011. "Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity." World Health Organization.

World Health Organization. 2016. "Guideline: Daily iron supplementation in infants and children." World Health Organization.

Yousafzai, Aisha K, Jelena Obradović, Muneera A Rasheed, Arjumand Rizvi, Ximena A Portilla, Nicole Tirado-Strayer, Saima Siyal, and Uzma Memon. 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, Aisha K, Muneera A Rasheed, Arjumand Rizvi, Robert Armstrong, and Zulfiqar A Bhutta. 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.

Zedler, Barbara K, Priyanka Kakad, Susan Colilla, Lenn Murrelle, and Nirav R Shah. 2011. "Does packaging with a calendar feature improve adherence to self-administered medication for long-term use? A systematic review." *Clinical Therapeutics*, 33(1): 62–73.

Two Scalable Interventions to Promote Health and Mental Development in Early Childhood:
A Randomized Controlled Trial in Rural India

ONLINE APPENDIX

Appendix I - The interventions

The dialogic reading intervention

The dialogic reading training was delivered by local female facilitators in Hindi or Maithili, the local language in Madhepura. Each training contained different contents, commencing with the benefits of dialogic reading and basic behavior guidelines when sharing a book, and advancing through the four home visits to topics of pointing and naming, evaluating, elaborating, talking about and relating to experiences, and talking about and relating to feelings. The content of each of the four training sessions and the procedure of each session are summarized in appendix Table A.I.1. The training content was developed by the Mikhulu Child Development Trust and was based on a program implemented and evaluated in South Africa before (Cooper et al., 2014; Murray et al., 2016; Vally et al., 2015).

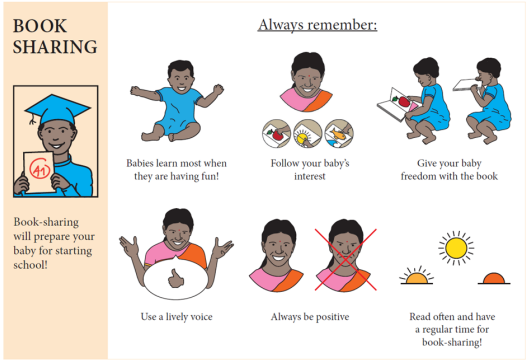
Table A.I.1: Topics of the dialogic reading training by session

Session	Concepts
1	Having fun Follow babies interest Freedom with the book Using a lively voice Always be positive Practice regularly
2	Point and say Asking questions Repeating the word Elaborating Making and action
3	Making links Talking about experiences
4	Talking about feelings Making links about feelings

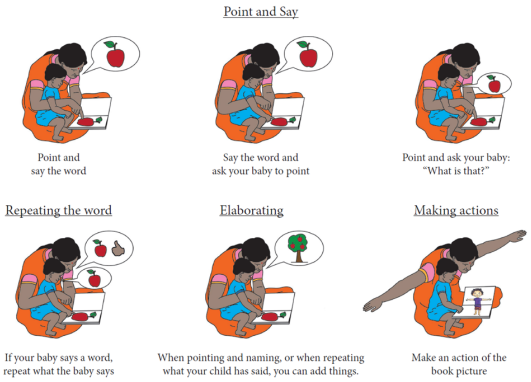
In the first three visits, the dialogic reading facilitators distributed one picture book per visit for the children and caregivers to keep. The colorful picture books contained familiar images, stimulating features of colors, shapes and materials and the content was suitable for practicing the dialogic reading concepts learned in the respective session. The book of the first session was the touch-and- feel book ”The Very Hungry Caterpillar” produced by the Penguin publisher. The second book (”Little Painters”) and third book (”Annual Haircut Day”) were obtained from Pratham Books, a sub-organization of the education focused Indian NGO Pratham. All books were chosen based on their cultural fit and children’s familiarity with images displayed and stories told. The books had a minimal amount of written words so the story could be followed by illiterate readers well.

In addition to the books, all treatment households received handouts which summarize the lessons learned by the training sessions in simple colorful pictures and simple instructions in Hindi language (see appendix Figure A.I.1 for the English version).

Home visits were conducted by local facilitators who were trained for six days by an experienced trainer of the Mikhulu Child Development Trust. In each household, the primary caregiver, typically the mother, and secondary caregivers such as grandparents, older siblings or fathers were encouraged to participate in the training. In case caregivers were not at home during the first visit, the facilitators revisited each house up to two times on the same day. Each home visit lasted 40 to 60 minutes. The first 25 to 45 minutes were spent on conveying the session's content using tablets for presentations, pictures and videos. Each session contained a review of previous sessions. In the subsequent 5 to 10 minutes, the facilitator introduced the new book to the caregiver and how the learned concepts can be applied using the book. At last, about 10 to 15 minutes were spent on the caregiver practicing book sharing with the child in the facilitator's presence using the new book. During that time the facilitators praised the caregiver for applying the concepts and advised on how to improve. At the end of each session, caregivers were encouraged to practice book sharing on a daily basis for the benefit of their child.



(a) Handout dialogic reading training 1



(b) Handout dialogic reading training 2

(c) Handout dialogic reading training 3 and 4

Figure A.I.1: Dialogic reading training material (English version)

The Lucky Iron Leaf intervention

The intervention addressed the main meal maker and at least one decision maker of the household, if available. However, often primary decision makers were not available and the intervention was carried out as long as the person responsible for cooking was present. The facilitators were male college students from Madhepura district who had received a five days training on the Lucky Iron Leaf utilization and iron deficiency anemia. Each intervention session started with a description of iron deficiency anemia, its symptoms, causes and consequences. A particular emphasis was put on the importance of sufficient iron intake for young children and women. Next, the facilitators introduced the Lucky Iron Leaf to the participants as a tool to avoid anemia. Using a small user manual suitable for an illiterate study population and depicted in Figure A.I.2, the facilitators explained the correct usage of the iron ingot. The facilitators were instructed to describe in detail how the Lucky Iron Leaf can be integrated in the preparation of context specific dishes such as daal, rice, cooked vegetables or lemon water. The face-to-face explanation was accompanied by the presentation of a short movie in which a woman from the study region demonstrates how she uses the Lucky Iron Leaf for cooking. The purpose of the movie was to deepen the understanding on how to use the Lucky Iron Leaf correctly and to overcome skepticism towards the product using social learning. At the end of each session, the household members were encouraged to ask questions and raise concerns. During some visits, participants were just preparing meals and, if appreciated by the household, the facilitator assisted the meal maker in using the Lucky Iron Leaf.

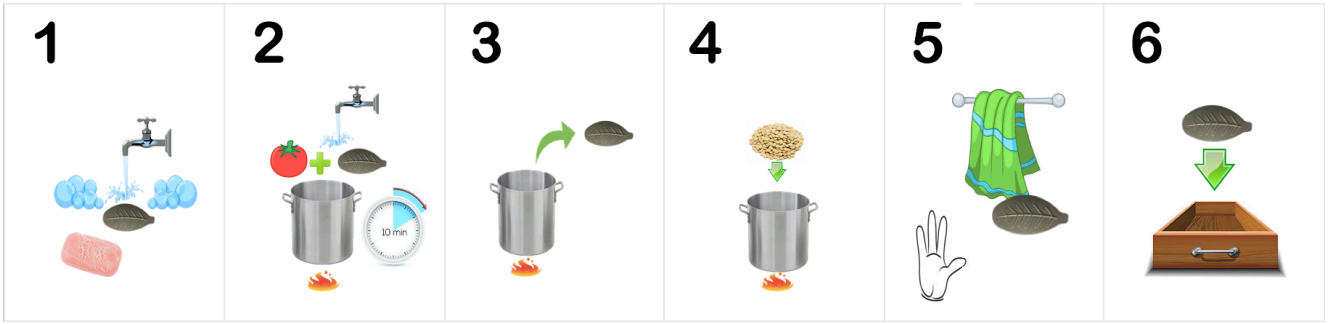


Figure A.I.2: Lucky Iron Leaf manual

Appendix II - Attrition and randomization

Table A.II.1: Selective attrition - Linear probability model results

	Baseline - endline	Pre-baleline - endline
DR	0.006	−0.035
	(0.030)	(0.031)
LIL	0.019	−0.024
	(0.030)	(0.031)
DR & LIL	0.018	−0.006
	(0.030)	(0.031)
Observations	1480	1995
Adjusted R^2	-0.00	-0.00
F statistic	0.19	0.54

Note: Outcome variables are binary indicators indicating the drop-out of households from the study sample between baseline and endline (column (1)) and pre-baseline and endline (column (2)). Standard errors in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A.II.2: Baseline balance in background characteristics of the baseline sample (not restricted to estimation sample)

	Control			Dialogic reading					Lucky Iron Leaf					Dialogic reading & Lucky Iron Leaf				
	Mean	SD	N	Mean	SD	N	Std. Diff.	p-value	Mean	SD	N	Std. Diff.	p-value	Mean	SD	N	Std. Diff.	p-value
Household characteristics:																		
Household size	5.66	2.04	354	5.69	2.30	377	−0.01	0.86	5.62	2.19	376	0.02	0.78	5.64	2.27	365	0.01	0.90
Hindu	0.87	0.34	355	0.83	0.37	379	0.10	0.16	0.85	0.36	379	0.05	0.48	0.87	0.34	367	0.01	0.88
Scheduled caste	0.29	0.46	353	0.31	0.46	379	−0.04	0.57	0.31	0.46	379	−0.04	0.62	0.32	0.47	365	−0.06	0.45
Scheduled tribe	0.04	0.20	353	0.04	0.20	379	0.00	0.99	0.03	0.16	379	0.09	0.23	0.04	0.20	365	0.01	0.93
Other backward class	0.59	0.49	353	0.58	0.49	379	0.04	0.59	0.57	0.50	379	0.05	0.49	0.58	0.49	365	0.03	0.65
General category	0.07	0.25	353	0.07	0.25	379	−0.00	0.97	0.09	0.29	379	−0.09	0.23	0.06	0.24	365	0.03	0.67
No schooling	0.43	0.50	355	0.49	0.50	379	−0.14	0.06	0.47	0.50	379	−0.10	0.18	0.42	0.49	367	0.02	0.82
Primary	0.17	0.38	355	0.14	0.35	379	0.09	0.23	0.15	0.35	379	0.07	0.32	0.16	0.37	367	0.03	0.69
Middle school	0.15	0.35	355	0.14	0.35	379	0.02	0.80	0.15	0.36	379	−0.00	0.96	0.17	0.38	367	−0.06	0.41
High school or higher	0.26	0.44	355	0.23	0.42	379	0.07	0.35	0.23	0.42	379	0.06	0.45	0.25	0.44	367	0.01	0.93
Asset index quintile[a]	5.46	2.91	354	5.34	2.99	379	0.04	0.57	5.05	2.91	376	0.14	0.06	5.16	2.94	366	0.10	0.16
Housing index quintile[b]	3.77	2.81	355	3.73	2.82	379	0.02	0.82	3.50	2.72	379	0.10	0.17	3.57	2.75	367	0.07	0.32
Mother characteristics:																		
Age in years	24.83	3.97	355	24.71	4.27	379	0.03	0.70	24.51	3.88	379	0.08	0.28	24.95	3.96	367	−0.03	0.68
No schooling	0.72	0.45	355	0.75	0.43	379	−0.08	0.30	0.82	0.38	379	−0.24	0.00	0.74	0.44	367	−0.05	0.49
Primary	0.10	0.29	355	0.05	0.22	379	0.16	0.03	0.04	0.20	379	0.21	0.00	0.07	0.25	367	0.10	0.18
Middle school	0.06	0.23	355	0.08	0.27	379	−0.09	0.22	0.04	0.21	379	0.05	0.48	0.10	0.29	367	−0.15	0.05
High school or higher	0.13	0.34	355	0.12	0.32	379	0.04	0.58	0.09	0.29	379	0.12	0.11	0.10	0.29	367	0.11	0.15
Can read SMS	0.28	0.45	355	0.28	0.45	379	−0.00	0.99	0.20	0.40	379	0.20	0.01	0.28	0.45	367	0.00	0.97
Worked past 12 months	0.92	0.29	355	0.89	0.32	379	0.09	0.24	0.91	0.30	379	0.05	0.54	0.91	0.30	367	0.05	0.53
Empowerment[c]	0.36	0.48	355	0.43	0.50	379	−0.14	0.05	0.39	0.49	379	−0.07	0.32	0.41	0.49	367	−0.11	0.14
Decides child nutrition	0.54	0.50	327	0.57	0.50	348	−0.08	0.30	0.53	0.50	337	0.01	0.92	0.49	0.50	322	0.10	0.23
Child characteristics:																		
Sex of child	0.50	0.50	348	0.54	0.50	373	−0.08	0.29	0.57	0.50	374	−0.15	0.04	0.46	0.50	361	0.08	0.29
Currently breastfed	0.89	0.31	345	0.91	0.29	368	−0.06	0.43	0.90	0.31	362	−0.01	0.92	0.89	0.31	349	−0.00	0.96
Iron past 3 months	0.33	0.47	310	0.37	0.48	326	−0.09	0.27	0.34	0.48	334	−0.03	0.68	0.37	0.48	320	−0.08	0.30

Note: Table continues on next page. Std. Diff. refers to the standardized difference in means of the control group and the respective treatment group. p-values refer to a t-test of the equality of means of the control group and the respective treatment group. [a]10 quintiles based on a durable asset index generated by factor analysis. [b]10 quintiles based on a housing quality index generated by factor analysis. [c]Indicator equals one if mother is allowed to go alone to at least one of five places (market, health facility, neighbor's, relatives or friends outside the village, place of worship) and participates in at least one of four decisions (health investments, household purchases, family visits outside village, and farm).

Table A.II.2 continued

	Control			Dialogic reading					Lucky Iron Leaf					Dialogic reading & Lucky Iron Leaf				
	Mean	SD	N	Mean	SD	N	Std. Diff.	p-value	Mean	SD	N	Std. Diff.	p-value	Mean	SD	N	Std. Diff.	p-value
Home environment:																		
Stimulation index[d]	5.66	1.74	332	5.61	1.70	349	0.03	0.74	5.61	1.75	344	0.03	0.73	5.61	1.80	336	0.03	0.73
Good educat. measures[e]	0.81	0.39	331	0.79	0.41	349	0.04	0.60	0.78	0.42	333	0.08	0.31	0.79	0.41	330	0.05	0.49
Bad educat. measures[f]	0.74	0.44	338	0.67	0.47	355	0.15	0.06	0.72	0.45	343	0.03	0.69	0.72	0.45	336	0.04	0.57
Outcome measures:																		
Cognitive	0.00	1.00	305	−0.08	1.03	328	0.08	0.34	−0.08	1.17	327	0.07	0.36	0.02	1.01	330	−0.02	0.82
Receptive language	0.00	1.00	312	−0.10	1.02	334	0.10	0.20	−0.07	1.10	335	0.07	0.41	−0.13	1.06	333	0.13	0.10
Expressive language	−0.00	1.00	313	−0.10	1.14	334	0.10	0.22	−0.12	1.09	335	0.11	0.16	−0.14	1.09	335	0.13	0.09
Motor	0.00	1.00	306	0.00	0.98	328	−0.00	1.00	0.06	1.01	327	−0.06	0.44	−0.05	0.97	332	0.05	0.50
Socioemotional	−0.00	1.00	307	−0.19	2.35	331	0.11	0.18	−0.11	1.12	327	0.11	0.17	0.14	2.17	333	−0.09	0.29
Hemoglobin g/dL	10.17	1.36	233	10.16	1.32	241	0.01	0.93	10.31	1.40	234	−0.10	0.26	10.31	1.41	251	−0.11	0.24
Anemia (any type)	0.71	0.46	233	0.71	0.45	241	−0.01	0.89	0.67	0.47	233	0.09	0.32	0.66	0.48	251	0.11	0.23
Moderate Anemia	0.40	0.49	233	0.41	0.49	241	−0.03	0.73	0.38	0.49	233	0.04	0.64	0.37	0.48	251	0.05	0.58

Note: Std. Diff. refers to the standardized difference in means of the control group and the respective treatment group. p-values refer to a t-test of the equality of means of the control group and the respective treatment group. [d]The sum of stimulating activities typically conducted with the child. [e]Equals 1 if caregiver explains wrong behavior to child, takes away privileges or gives child something else to do. [f]Equals 1 if caregiver shouts, yells or screams at child or spanks, hits, kicks or slaps child.

Appendix III - Additional intention-to-treat results

Table A.III.1: ITT effects on child development with tester fixed effects and children's age fixed effects

	Cognitive		Receptive language		Expressive language		Motor		Socioemotional	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
DR	0.097	0.106	−0.017	−0.009	0.042	0.038	0.056	0.053	0.083	0.055
	(0.077)	(0.082)	(0.081)	(0.085)	(0.081)	(0.086)	(0.077)	(0.081)	(0.071)	(0.082)
LIL	−0.010	0.022	0.012	0.029	0.059	0.060	0.007	0.038	0.054	0.044
	(0.075)	(0.082)	(0.080)	(0.085)	(0.082)	(0.086)	(0.077)	(0.082)	(0.071)	(0.080)
DR & LIL	0.049	0.078	0.006	0.017	0.044	0.045	0.057	0.059	0.023	0.005
	(0.079)	(0.086)	(0.082)	(0.088)	(0.086)	(0.089)	(0.078)	(0.084)	(0.072)	(0.085)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tester fixed effects	✓		✓		✓		✓		✓	
Age in months fixed effects		✓		✓		✓		✓		✓
Observations	1136	1136	1148	1148	1148	1148	1113	1113	1140	1140
Adjusted R^2	0.225	0.092	0.121	0.028	0.125	0.050	0.176	0.104	0.251	0.046

Note: Dependent variables are standardized development scores and effect sizes are in standard deviations. Control variables are maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Tester fixed effects are development test facilitator fixed effects. Standard errors in parentheses. Conventional significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Multiple testing corrected significance levels: $_{+}$ $p<0.0096$, $_{++}$ $p<0.0048$, $_{+++}$ $p<0.0010$.

Table A.III.2: Intention-to-treat effects on secondary health outcomes

	Head circumference	Stunted	Wasted	Underweight
DR	-0.126	-0.009	-0.002	-0.034
	(0.091)	(0.039)	(0.031)	(0.042)
LIL	-0.129	-0.044	0.024	-0.026
	(0.095)	(0.040)	(0.032)	(0.043)
DR & LIL	0.062	-0.035	0.003	-0.070*
	(0.092)	(0.040)	(0.030)	(0.042)
Controls	✓	✓	✓	✓
ECD outcome sample	✓			
Iron outcome sample		✓	✓	✓
Observations	1104	1119	1096	1125
Adjusted R^2	0.018	0.050	0.004	0.009
Control mean	-1.62	0.67	0.15	0.53

Note: Dependent variable in column (1) is standardized (z-scores) and effect sizes are in standard deviations. Dependent variables in column (2) to (4) are binary indicators. Control variables are maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Standard errors are in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A.III.3: ITT effects on hemoglobin and anemia in a sample restricted to the ANCOVA sample

	Hemoglobin			Any anemia			Moderate anemia		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
DR	–0.077	–0.140	–0.121	0.080	0.111**	0.112**	0.026	0.036	0.030
	(0.136)	(0.136)	(0.122)	(0.052)	(0.052)	(0.050)	(0.047)	(0.047)	(0.044)
LIL	–0.118	–0.127	–0.132	0.060	0.060	0.064	0.050	0.051	0.044
	(0.144)	(0.142)	(0.127)	(0.053)	(0.052)	(0.048)	(0.049)	(0.049)	(0.046)
DR & LIL	–0.074	–0.136	–0.158	–0.000	0.033	0.045	0.036	0.045	0.054
	(0.149)	(0.148)	(0.131)	(0.053)	(0.052)	(0.048)	(0.047)	(0.048)	(0.045)
Controls		✓	✓		✓	✓		✓	✓
HemoCue machine fixed effects		✓	✓		✓	✓		✓	✓
Baseline outcome			✓			✓			✓
Observations	710	710	710	710	710	710	710	710	710
Adjusted R^2	-0.003	0.038	0.269	0.001	0.039	0.176	-0.003	0.003	0.111

Note: Hemoglobin is measured in g/dl. Any anemia and moderate anemia are binary indicators. Control variables are maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Device fixed effects are HemoCue machine fixed effects. Baseline outcome indicates that the estimation controls for the baseline value of the model's dependent variable. Standard errors are in parentheses. Conventional significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Multiple testing corrected significance levels: $_{+}$ $p<0.0096$, $_{++}$ $p<0.0048$, $_{+++}$ $p<0.0010$.

Appendix IV - Additional results on heterogeneous treatment effects

Table A.IV.1: Heterogenous ITT effects by mothers' education and empowerment status

	Hemoglobin	Any anemia	Cognitive	Receptive	Expressive	Motor	Socio-emotional
Mother completed primary							
DR	–0.153	0.059	0.115	–0.063	0.052	0.067	0.084
	(0.141)	(0.049)	(0.096)	(0.097)	(0.099)	(0.093)	(0.097)
DR x Primary	–0.028	0.035	–0.037	0.218	–0.080	–0.037	–0.101
	(0.256)	(0.104)	(0.179)	(0.192)	(0.192)	(0.189)	(0.166)
LIL	0.099	–0.015	–0.065	–0.083	0.019	0.009	0.082
	(0.139)	(0.049)	(0.093)	(0.093)	(0.097)	(0.094)	(0.096)
LIL x Primary	–0.504*	0.137	0.400*	0.512**	0.169	0.146	–0.196
	(0.273)	(0.109)	(0.209)	(0.216)	(0.198)	(0.202)	(0.162)
DR&LIL	–0.017	0.004	0.075	–0.068	0.064	0.133	0.050
	(0.147)	(0.051)	(0.100)	(0.099)	(0.107)	(0.097)	(0.102)
DR&LIL x Primary	0.077	–0.049	–0.042	0.321	–0.142	–0.252	–0.177
	(0.265)	(0.102)	(0.200)	(0.202)	(0.195)	(0.195)	(0.165)
Primary	0.167	0.002	0.325**	–0.007	0.372**	0.177	0.497$^{***}_{+++}$
	(0.236)	(0.086)	(0.165)	(0.152)	(0.155)	(0.167)	(0.133)
Observations	1039	1039	1136	1148	1148	1113	1140
Adjusted R^2	0.03	0.02	0.08	0.03	0.05	0.07	0.04
Mother is empowered							
DR	–0.253	0.096*	0.184*	0.075	–0.059	0.116	–0.053
	(0.155)	(0.057)	(0.101)	(0.108)	(0.116)	(0.105)	(0.108)
DR x Empowered	0.140	–0.041	–0.180	–0.207	0.217	–0.144	0.276*
	(0.239)	(0.088)	(0.167)	(0.172)	(0.169)	(0.166)	(0.165)
LIL	0.093	–0.030	0.008	0.096	0.035	–0.002	–0.061
	(0.146)	(0.055)	(0.104)	(0.106)	(0.112)	(0.105)	(0.102)
LIL x Empowered	–0.288	0.129	–0.024	–0.211	0.011	0.058	0.272*
	(0.248)	(0.089)	(0.170)	(0.174)	(0.170)	(0.169)	(0.165)
DR&LIL	0.073	–0.023	0.086	0.059	–0.004	0.077	–0.013
	(0.161)	(0.056)	(0.111)	(0.109)	(0.120)	(0.111)	(0.106)
DR&LIL x Empowered	–0.211	0.049	–0.062	–0.115	0.104	–0.043	0.076
	(0.251)	(0.090)	(0.179)	(0.182)	(0.181)	(0.172)	(0.170)
Empowered	0.092	–0.035	0.079	0.094	0.015	0.024	–0.103
	(0.178)	(0.063)	(0.122)	(0.128)	(0.118)	(0.121)	(0.121)
Controls	✓	✓	✓	✓	✓	✓	✓
Device fixed effects	✓	✓					
Observations	1039	1039	1136	1148	1148	1113	1140
Adjusted R^2	0.03	0.03	0.08	0.03	0.05	0.07	0.04

Note: Hemoglobin is measured in g/dl. Any anemia is a binary indicator. Dependent variables in column (3) to (7) are standardized development scores and effect sizes are in standard deviations. Control variables in top panel are maternal reading ability and empowerment, the child's sex and subdistrict fixed effects. Note, results do not change when maternal reading ability is omitted. Control variables in bottom panel are maternal education and reading ability, the child's sex and subdistrict fixed effects. HemoCue machine fixed effects control for the hemoglobin measurement device and tester. Standard errors are in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01. Multiple testing corrected significance levels: $_{+}$ p<0.0096, $_{++}$ p<0.0048, $_{+++}$ p<0.0010.

Table A.IV.2: Heterogenous ITT effects by children's sex and anemia status at baseline

	Hemoglobin	Any anemia	Cognitive	Receptive	Expressive	Motor	Socio-emotional
Sex of child							
DR	–0.012	0.062	0.107	–0.026	–0.025	0.275**	0.078
	(0.174)	(0.062)	(0.118)	(0.124)	(0.118)	(0.114)	(0.117)
DR x Boy	–0.308	0.020	0.001	0.038	0.112	–0.427‡***	–0.028
	(0.233)	(0.085)	(0.163)	(0.166)	(0.169)	(0.160)	(0.160)
LIL	0.069	–0.021	–0.034	–0.056	0.030	0.204*	–0.019
	(0.182)	(0.064)	(0.117)	(0.124)	(0.111)	(0.122)	(0.121)
LIL x Boy	–0.154	0.067	0.062	0.137	0.024	–0.346**	0.108
	(0.241)	(0.088)	(0.166)	(0.168)	(0.167)	(0.165)	(0.161)
DR&LIL	0.042	–0.007	0.022	–0.057	0.008	0.138	–0.037
	(0.182)	(0.061)	(0.126)	(0.122)	(0.121)	(0.120)	(0.122)
DR&LIL x Boy	–0.104	0.003	0.085	0.152	0.055	–0.153	0.099
	(0.245)	(0.087)	(0.174)	(0.174)	(0.182)	(0.169)	(0.166)
Boy	0.241	–0.030	–0.073	–0.089	–0.186	0.317‡***	–0.079
	(0.173)	(0.061)	(0.116)	(0.120)	(0.119)	(0.115)	(0.117)
Observations	1039	1039	1136	1148	1148	1113	1140
Adjusted R^2	0.03	0.02	0.08	0.03	0.05	0.07	0.04
Children's baseline anemia status							
DR	–0.292	0.149	0.143	0.420**	0.075	0.107	0.168
	(0.199)	(0.090)	(0.179)	(0.179)	(0.183)	(0.172)	(0.174)
DR x Anemic	0.211	–0.052	0.012	–0.388*	–0.074	0.018	–0.154
	(0.254)	(0.108)	(0.215)	(0.216)	(0.224)	(0.212)	(0.214)
LIL	0.105	–0.012	0.300*	0.330*	0.206	0.146	0.074
	(0.186)	(0.084)	(0.177)	(0.170)	(0.157)	(0.168)	(0.164)
LIL x Anemic	–0.344	0.107	–0.347	–0.418**	–0.201	–0.026	0.025
	(0.247)	(0.102)	(0.221)	(0.213)	(0.201)	(0.215)	(0.204)
DR&LIL	0.122	–0.032	0.450**	0.538‡‡**	0.047	0.144	0.213
	(0.198)	(0.084)	(0.181)	(0.175)	(0.183)	(0.177)	(0.165)
DR&LIL x Anemic	–0.433*	0.112	–0.427*	–0.596‡***	–0.039	–0.075	–0.287
	(0.262)	(0.102)	(0.223)	(0.219)	(0.230)	(0.218)	(0.208)
Anemic	–1.052‡‡‡*	0.357‡‡‡*	0.238	0.327**	–0.151	–0.060	–0.048
	(0.177)	(0.073)	(0.151)	(0.143)	(0.147)	(0.149)	(0.151)
Controls	✓	✓	✓	✓	✓	✓	✓
HemoCue machine fixed effects	✓	✓					
Observations	710	710	758	766	766	742	760
Adjusted R^2	0.20	0.18	0.08	0.03	0.04	0.05	0.04
p-value (DR+interaction=0)	0.61	0.10	0.19	0.79	0.99	0.31	0.91
p-value (LIL+interaction=0)	0.15	0.10	0.71	0.49	0.97	0.36	0.42
p-value (DR&LIL+int.=0)	0.07	0.17	0.85	0.66	0.95	0.58	0.56

Note: Hemoglobin is measured in g/dl. Any anemia is a binary indicator. Dependent variables in column (3) to (7) are standardized development scores and effect sizes are in standard deviations. Control variables are maternal education, reading ability and empowerment, the child's sex (not in top panel) and subdistrict fixed effects. HemoCue machine fixed effects control for the hemoglobin measurement device and tester. Standard errors are in parentheses. Conventional significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Multiple testing corrected significance levels: $_{+}$ $p<0.0096$, $_{++}$ $p<0.0048$, $_{+++}$ $p<0.0010$.

Table A.IV.3: Linear probability model of children's anemia at baseline on family and children's background characteristics

	Anemia	Anemia	Anemia	Anemia	Anemia	Anemia	Anemia
Household characteristics:							
Household size	-0.003						-0.007
	(0.008)						(0.010)
Hindu	-0.017						0.041
	(0.054)						(0.069)
Scheduled caste	0.120						0.099
	(0.074)						(0.098)
Scheduled tribe	0.065						0.139
	(0.109)						(0.137)
Other backward class	-0.003						0.004
	(0.069)						(0.090)
Primary	-0.071						-0.075
	(0.053)						(0.063)
Middle school	-0.041						-0.014
	(0.058)						(0.074)
High school or higher	0.018						0.052
	(0.061)						(0.077)
Asset index quintile	-0.001						0.001
	(0.007)						(0.008)
Housing index quintile	0.003						-0.000
	(0.007)						(0.008)
Mother characteristics:							
Age in years		-0.009**					-0.012**
		(0.005)					(0.005)
Worked in past 12 months		-0.082					-0.063
		(0.052)					(0.069)
Decides about child nutrition		0.054					0.069
		(0.036)					(0.043)
Child characteristics:							
Currently breastfed			-0.021				-0.022
			(0.035)				(0.042)
Vit-A past 6 months			-0.010				-0.014
			(0.039)				(0.047)
Iron past 3 months			-0.018				-0.011
			(0.048)				(0.056)
ageLB			-0.001				0.002
			(0.007)				(0.009)
Home environment:							
Moderate stimulation				-0.082			-0.014
				(0.087)			(0.132)
High stimulation				-0.127			-0.039
				(0.087)			(0.131)
Good educational measures				-0.091			-0.160**
				(0.059)			(0.071)
Bad educational measures				0.140*			0.073
				(0.080)			(0.105)
Baseline outcomes:							
Cognitive					0.009		0.011
					(0.021)		(0.025)
Receptive language					0.025		0.014
					(0.019)		(0.021)
Expressive language					-0.023		-0.026
					(0.020)		(0.024)
Motor					-0.011		0.006
					(0.022)		(0.026)
Socioemotional					-0.027		-0.027
					(0.018)		(0.022)
Child anthropometrics:							
Height (z-score)						-0.033**	-0.035**
						(0.013)	(0.017)
Weight-for-age (z-score)						0.000	-0.002
						(0.013)	(0.017)
Head circumference (z-score)						0.001	0.016
						(0.017)	(0.022)
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	761	702	724	718	678	724	539
Adjusted R^2	0.02	0.02	0.01	0.01	0.01	0.02	0.00
F statistic	1.53	2.02	1.24	1.59	1.45	1.93	1.20

Note: Anemia is a binary indicator. Control variables are maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Standard errors in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A.IV.4: Heterogenous ITT effects on receptive language by children's anemia status and controlling for potential confounders

	Model 1	Model 2 Restricted to model 3 sample	Model 3
DR	0.420**	0.430**	0.439**
	(0.179)	(0.181)	(0.181)
DR x Anemic	-0.388*	-0.368*	-0.382*
	(0.216)	(0.219)	(0.220)
LIL	0.330*	0.358**	0.366**
	(0.170)	(0.176)	(0.180)
LIL x Anemic	-0.418**	-0.442**	-0.459**
	(0.213)	(0.219)	(0.222)
DR&LIL	0.538^{***}_{++}	0.524^{***}_{++}	0.493^{**}_{+}
	(0.175)	(0.178)	(0.180)
DR&LIL x Anemic	-0.596^{**}_{+}	-0.557**	-0.506**
	(0.219)	(0.224)	(0.224)
Anemic	0.327**	0.315**	0.322**
	(0.143)	(0.144)	(0.145)
Mother's age (years)			-0.011
			(0.009)
Good educational measures			0.194
			(0.125)
Bad educational measures			-0.175
			(0.169)
Height (z-score)			0.051**
			(0.024)
Controls	✓	✓	✓
Observations	766	732	732
Adjusted R^2	0.03	0.03	0.04

Note: Dependent variable is the standardized receptive language score and effect sizes are in standard deviations. Control variables are maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Standard errors in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01. Multiple testing corrected significance levels: $_{+}$ p<0.0096, $_{++}$ p<0.0048, $_{+++}$ p<0.0010.

Appendix V - Additional results on spillovers and mediators

Table A.V.1: Spillover statistics

	Panel A: Spill-overs from dialogic reading intervention					
	Dialogic reading households		Control households		Lucky Iron Leaf households	
	Lent books to neighbors	DR with neighbor/ friends	Shares books	Have/had intervention books	Shares books	Have/had intervention books
No	0.72	0.51	0.87	0.98	0.86	0.97
Yes	0.04	0.25	0.13	0.02	0.14	0.03
N/A	0.24	0.24				
Observations	666	666	326	326	337	337
	Panel B: Spill-overs from the Lucky Iron Leaf intervention			
	Lucky Iron Leaf households		Control households	Dialogic reading households
	LIL lent to s.o. else	Use LIL $\geq$4 times per week & Neighbor's children eat regularly in HH	Heard about LIL	Heard about LIL
No	0.88	0.92	0.93	0.92
Yes	0.06	0.06	0.02	0.02
N/A	0.06	0.02	0.05	0.05
Observations	677	665	329	353

Note: **Panel A**: Columns (1) to (2) refer to spillovers *sent* from the households in the pure dialogic reading treatment arm or the combined intervention arm, columns (3) to (4) refer to spillovers *received* by pure control group households and columns (5) to (6) refer to spillovers *received* by the pure Lucky Iron Leaf treatment arm households. **Panel B**: Columns (1) to (2) refer to spillovers *sent* from the households in the pure Lucky Iron Leaf treatment arm or the combined intervention arm, columns (3) refer to spillovers *received* by pure control group households and columns (4) refer to spillovers *received* by the pure dialogic reading treatment arm households. Share of households that fall in the N/A category is large for book lending and book sharing with other children. This is caused by a skip pattern resulting in treatment households not being asked these questions and retrospectively being coded as N/A.

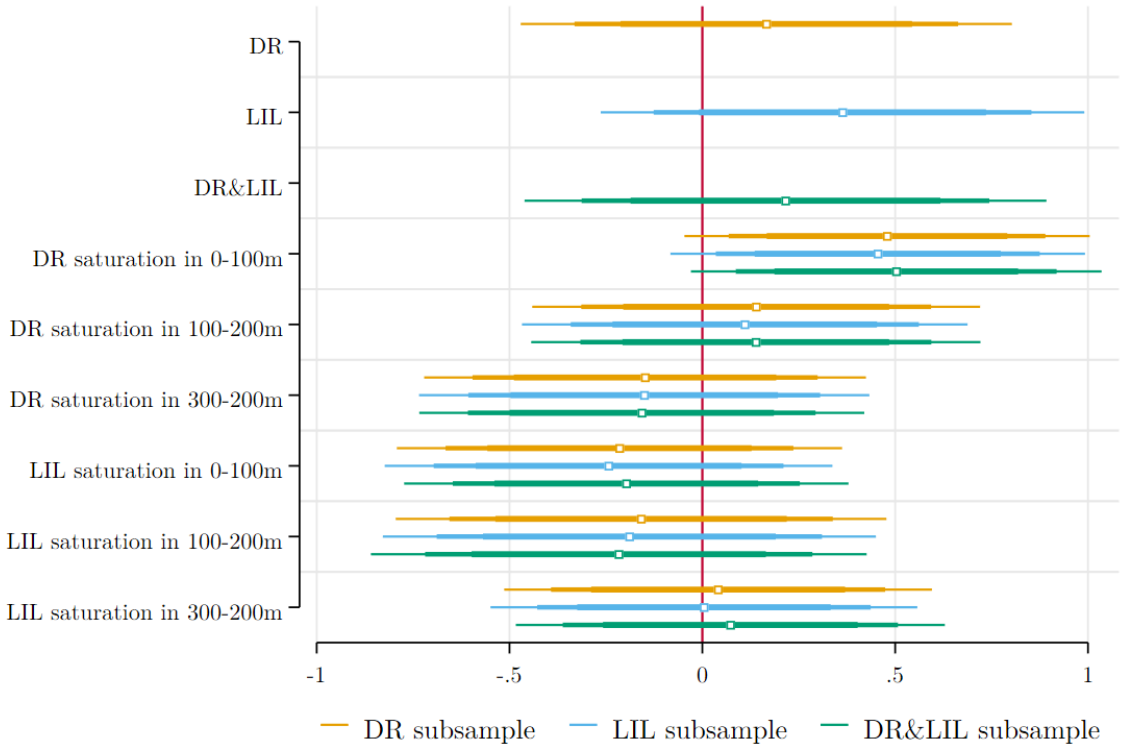


Figure A.V.1: Spillover results with additional distance thresholds

Note: The Figure shows results for a specification similar to equation (3) in section VII.C, extended by terms for the share of treated neighbors in the 100 to 200 meters distance and 200 to 300 meters distance, their interaction with the treatment status, and the total number of neighbors in these distances. The outcome is the overall development score. The specification is estimated separately for each intervention arm, similar to Table VII.

Table A.V.2: Lucky Iron Leaf mediators and substitution effects

	Anemia awareness	Food diversity	Vegetables	Meat	Iron supplements	Vitamin A supplements	Anganwadi visit
DR	0.006	0.105	-0.002	0.015	-0.001	0.042	-0.038
	(0.018)	(0.132)	(0.025)	(0.042)	(0.035)	(0.041)	(0.040)
LIL	0.021	-0.073	-0.019	-0.059	0.007	0.059	-0.007
	(0.018)	(0.128)	(0.026)	(0.042)	(0.034)	(0.041)	(0.041)
DR & LIL	0.028	0.001	-0.021	-0.015	-0.011	-0.001	-0.057
	(0.020)	(0.131)	(0.025)	(0.043)	(0.034)	(0.041)	(0.040)
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	1096	1133	1137	1136	1116	1096	1125
Adjusted R^2	0.11	0.01	0.05	0.00	0.01	0.03	0.00
Control mean	0.04	7.10	0.91	0.43	0.20	0.34	0.34

Note: Anemia awareness: has heard of anemia; Food diversity: summation of indicators for the consumption of 10 food items; Vegetables: indicator for the consumption of vegetables; Meat: indicator for the consumption of meat; Iron supplements: indicator for the child's consumption of iron supplements in the past 3 months; Vitamin A supplements: indicator for the child's consumption of vitamin A supplements in the past 6 months; Anganwadi visit: indicator for whether the Anganwadi center was visited with the child in the past 14 days. Control variables are maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Standard errors are in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.