

Ebert, Cara; Heesemann, Esther; Vollmer, Sebastian

Working Paper

Encouraging parents to invest: A randomized trial with two simple interventions in early childhood

Ruhr Economic Papers, No. 856

Provided in Cooperation with:

RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Ebert, Cara; Heesemann, Esther; Vollmer, Sebastian (2020) : Encouraging parents to invest: A randomized trial with two simple interventions in early childhood, Ruhr Economic Papers, No. 856, ISBN 978-3-86788-992-6, RWI - Leibniz-Institut für Wirtschaftsforschung, Essen, <https://doi.org/10.4419/86788992>

This Version is available at:

<https://hdl.handle.net/10419/228513>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Cara Ebert
Esther Heesemann
Sebastian Vollmer

Encouraging Parents to Invest: A Randomized Trial with two Simple Interventions in Early Childhood

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer

RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger

Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@tu-dortmund.de

Prof. Dr. Volker Clausen

University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Roland Döhrn, Prof. Dr. Manuel Frondel,
Prof. Dr. Ansgar Wübker

RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler

RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #856

Responsible Editor: Ronald Bachmann

All rights reserved. Essen, Germany, 2020

ISSN 1864-4872 (online) – ISBN 978-3-86788-992-6

The working papers published in the series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

Ruhr Economic Papers #856

Cara Ebert, Esther Heesemann, and Sebastian Vollmer

Encouraging Parents to Invest: A Randomized Trial with two Simple Interventions in Early Childhood

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

<http://dx.doi.org/10.4419/86788992>

ISSN 1864-4872 (online)

ISBN 978-3-86788-992-6

Cara Ebert, Esther Heesemann, and Sebastian Vollmer¹

Encouraging Parents to Invest: A Randomized Trial with two Simple Interventions in Early Childhood

Abstract

The lottery of birth draws some children into deprived environments and others into environments where they thrive. In a field experiment in rural India with 10-20 months old children we test two scalable interventions to reduce early disadvantages in health and mental development. We distribute a durable device for home iron fortification of meals, called the Lucky Iron Leaf, and picture books together with a training for caregivers in dialogic reading. We find no significant average impact of either intervention on anemia or mental development. However, we find a cross-productivity of children's baseline health and the interventions' effectiveness. Children, who are non-anemic at baseline, improve in receptive language skills by half a standard deviation one year after implementation.

JEL-Code: D04, I12, I15, J13

Keywords: Early childhood; parental investment; nutrition; health behavior; human capital

December 2020

¹ Cara Ebert, RWI; Esther Heesemann, University of Mannheim; Sebastian Vollmer, University of Göttingen. – We are grateful to Alastair Summerlee for the collaboration in the design of the Lucky Iron Leaf, its laboratory testing, and the provision of 1000 Lucky Iron Leafs by the Lucky Iron Fish Enterprise. Peter Cooper and the Mikhulu Child Development Trust 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 are grateful to Maximilian Sprengholz for the design of dialogic reading handouts. We are grateful for the support of Malavika Subramanyam and allowing us to connect to her ongoing field work. At last, we want to thank Abhijeet Kumar for field work coordination and his dedication to the project. Source of funding: The primary data collection and intervention implementation was funded by (i) State Ministry of Science and Culture, State of Lower Saxony, Germany (grant VWZN3060) and (ii) the DFG German Research Foundation (RTG1723: Globalization and Development and RTG 1666: Transformation of Global Agri-Food Systems). The trial is registered at the American Economic Association Registry www.socialscienceregistry.org/trials/2696. – All correspondence to: Cara Ebert, RWI - Büro Berlin, Invalidenstr. 112, 10115 Berlin, Germany, e-mail: cara.ebert@rwi-essen.de

I INTRODUCTION

The first 1000 days of life mark a critical period of human development. In those early years, the young brain forms millions of synapses to a dense network for the transmission and storage of information, and prunes network connections that are not used. The brain's plasticity during this phase, however, also implies its vulnerability to adverse environments. Children living in poverty face a multitude of circumstances, such as poor health and mental stimulation, which can impair skill formation (Walker et al., 2007, 2011*b*). Early disadvantages can affect children over a lifetime as the skill accumulation process is of dynamic and path-dependent nature (Cunha et al., 2006; Cunha and Heckman, 2007; Heckman, 2007; Cunha and Heckman, 2008; Cunha, Heckman and Schennach, 2010; Attanasio, 2015; Attanasio et al., 2020*a,b*; Attanasio, Meghir and Nix, 2020). Early skills and health as well as parental investment matter interdependently in reinforcing and complementing ways for human capital outcomes later in life. In consequence, health and mental development follow a steep socioeconomic gradient which widens with age (Taylor, Dearing and McCartney, 2004; Fernald et al., 2012; Hamadani et al., 2014; Rubio-Codina et al., 2015; Schady et al., 2015; Reynolds et al., 2017). The economic potential lost to reduced skill formation is substantial.¹ Hanushek and Woessmann (2012*a*) estimate that one standard deviation higher cognitive skills is associated with two percentage points higher annual national economic growth.

Early childhood development programs are salient policy interventions to avoid early disadvantages and their far-reaching consequences. Nutrition and stimulation interventions have shown to affect adult cognitive and psychosocial skills, test scores, educational attainment, earnings, crime, and even children's health and skills in the next generation (Alderman et al., 2001; Glewwe, Jacoby and King, 2001; Walker et al., 2005; Hoddinott et al., 2008; Behrman et al., 2009; Maluccio et al., 2009; Walker et al., 2011*a*; Hazarika and Viren, 2013; Behrman et al., 2014; Gertler et al., 2014; Walker et al., 2015; Thompson, 2018; Lazuka, 2020). An often-replicated two-year weekly stimulation program with one-year old children in Jamaica increased children's earnings by 25 percent 20 years later (Grantham-McGregor et al., 1991; Walker et al., 1991; Gertler et al., 2014).

Our study adds to an emerging literature investigating how to effectively put early childhood development interventions to scale. Previous studies have focused on integrating programs into existing infrastructure using, for example, community health workers and effectively improved children's mental development (Attanasio et al., 2014; Yousafzai et al., 2014, 2016; Attanasio et al., 2018; Andrew et al., 2020; Sylvia et al., 2020). However, in many places where early childhood interventions are needed, public infrastructure is dysfunctional. For example, Sylvia et al. (2020)

¹For examples on economic outcomes see Lee and Lee (1995), Hanushek and Kimko (2000), Denny, Harmon and O'Sullivan (2004), Heckman, Stixrud and Urzua (2006), Hanushek and Woessmann (2008), Hanushek and Woessmann (2012*b*), Hanushek (2013), and Atherton, Appleton and Bleaney (2013).

point out challenges in the supply-side compliance of a parenting program that uses existing public infrastructure in China. In rural Bihar (India), where this study is set, spot-checks at mother-child centers (Anganwadi centers) revealed that the centers are often overburdened and unreliable.² Therefore, we propose an alternative program design which may be suitable in contexts where service delivery via public infrastructure is not an option. Our approach intends to empower parents to improve nutrition and stimulation of their children with inexpensive and long lasting means.

We study two interventions that are remarkably simple in comparison to existing effective early skill development programs, such as the Jamaica-program which had subsequently been implemented, for example, in Bangladesh, China, Colombia, India and Pakistan (Hamadani et al., 2006; Tofail et al., 2013; Attanasio et al., 2014; Yousafzai et al., 2014; Andrew et al., 2020; Attanasio et al., 2020b; Grantham-McGregor et al., 2020; Sylvia et al., 2020). Our first intervention targets iron deficiency as a risk factor of skill development. In one home visit, we distribute an iron cooking utensil, called Lucky Iron LeafTM (hereafter Lucky Iron Leaf), for home iron fortification. The Lucky Iron Leaf leaches iron when placed in boiling water with some fruit acid and the enriched water then fortifies rice, lentils or alike during regular cooking. The intervention is simple in the delivery to households because the Lucky Iron Leaf is a durable source of iron, lasting for up to five years, and does not require replenishment as conventional fortification technologies do (e.g., multi-micronutrient powder). The second intervention targets the lack of psychosocial stimulation as a risk factor of skill development. During four home visits caregivers learn methods of dialogic reading, which is an interactive mode of sharing picture books to stimulate speaking and learning, and receive three durable picture books. 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.

The study follows a factorial design, in which we assigned 1,480 households with 10- to 20-months old children to one of four main experimental arms: one stand-alone Lucky Iron Leaf arm, one stand-alone dialogic reading arm, one combined interventions arm, which receives the Lucky Iron Leaf and dialogic reading treatment, and one no-intervention control arm. Among dialogic reading households we further cross-randomized the book sharing calendar. The factorial design allows us to capture potential interactions of the nutrition and psychosocial stimulation interventions that potentially stem from iron replenished children's improved mental capacity and enhanced learning potential during book sharing (Lozoff, 2007; Beard, 2008).

The study location is the rural district Madhepura in the northeast Indian state of Bihar. Chronic malnutrition and anemia are common in Madhepura and public child care services for under-three-

²During most spontaneous visits, Anganwadi workers and children were absent at arrival. Many Anganwadi workers complained about not receiving the funds they are entitled to. Similarly, Fraker, Shah and Abraham (2013) report about Anganwadi centers in Bihar: 23% of the centers were closed during opening hours, 59% provided meals to the children, only 53% of children attended the centers when meals were provided.

year olds do not exist (International Institute for Population Sciences, 2017).

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 cognitive, language, motor or socioemotional functions on average. However, we find cross-productivities of the interventions and children's health status at the beginning of the trial. Non-anemic children at baseline, who make up 30 percent of the sample, benefited from the combined intervention in receptive language skills by 0.54 standard deviations. Given the simplicity of our interventions, the effect size for non-anemic children is considerable. For example, an intensive 24 months weekly stimulation intervention in Colombia, following the Jamaica-program, improved receptive language and cognition by 0.22 and 0.26 standard deviations, respectively, in a study context where the anemia rate is about 19 percent (Attanasio et al., 2014).

The paper is organized as follows. Section II describes the interventions, section III the research design and data, and section IV the estimation strategy. Section V presents estimates of the program impact on children's development and anemia as well as effect heterogeneities. Section VI presents intervention take-up rates, treatment effects under consideration of spillovers and complier average causal effects. Section VII examines the interventions' impact on other domains of caregiving. Section VIII discusses contextual challenges for intervention take-up. Section IX concludes.

II THE INTERVENTIONS

II.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 2 million inhabitants and is almost entirely rural, with few exceptions such as its capital, also named Madhepura.

Chronic and acute malnutrition are common in Madhepura. About half of the children under the age of five are stunted and/or underweight, and 61 percent are 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 in 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. Yet, 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 consumed any iron supplements in the seven days prior to the interview (International

Institute for Population Sciences and ICF, 2017).

No public programs for psychosocial stimulation of under three-year old children exist in rural Bihar. The day care of the Integrated Child Development Services (ICDS) in Anganwadi centers exclusively addresses children from three years onwards.³ For children under the age of three the ICDS targets solely health aspects - i.e. macro-nutrient food rations, vaccinations, and health check-ups.

II.B Dialogic reading

Dialogic reading is a method of 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 in order to encourage children to think, speak and learn. In contrast to children listening passively to readings, they become story tellers themselves based on images in the book and assisted by the caregiver's questions and encouragements.

The effectiveness of dialogic reading is well studied in high-income countries.⁴ Evidence from low- and middle-income countries includes four small-scale, closely controlled trials in Mexico, Bangladesh, and South Africa (Valdez-Menchaca and Whitehurst, 1992; Opel, Ameer and Aboud, 2009; Cooper et al., 2014; Vally et al., 2015; Murray et al., 2016). After a four to eight week program with weekly meetings, children improved in receptive and expressive language, attention and socioemotional behavior. To our knowledge, this is the first trial which evaluates the impact of dialogic reading for implementation at large scale, in the Indian context and in a high anemia prevalence population (70%).

The training program of the dialogic reading intervention of this trial was developed by the Mikhulu Child Development Trust (Mikhulu) 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., 2016). During four home-visits of 45 to 60 minutes, female facilitators, trained by Mikhulu, taught and practiced methods of dialogic reading with the primary caregiver and the child. During each of the first three visits, a picture book was left with the families, resulting in a variety of three books in total to exercise book sharing throughout the year. The first two dialogic reading training sessions were delivered immediately after baseline data collection and one week apart from each other.

³Even 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).

⁴For 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).

Sessions three and four were delivered three months later, again one week apart from each other. Appendix IX details further aspects of the intervention.

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 basic aims in order to improve children's skill development (Fletcher and Reese, 2005). In particular, we expect benefits to emerge from increasing responsive parent-child interactions (Landry, Smith and Swank, 2006) and by exposing children to more spoken words (Weisleder and Fernald, 2013). To achieve optimal learning outcomes, children's interest in sharing picture books is key for them to engage with caregivers and books attentively (Fletcher and Reese, 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 in which books were shared. The calendar is intended to serve as a reminder and was inspired by the medical literature, in which calendar boxes or blisters are commonly used to improve adherence with medication (Zedler et al., 2011). The calendar was structured by the lunar months and included colorful graphics that indicate national holidays and festivities. The design was targeted to an illiterate population and it consisted of robust material to last throughout the intervention period (and beyond).

II.C The Lucky Iron LeafTM

The second intervention consists of 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 together with 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 water and regular food is added which will in turn be fortified while cooking it in the iron enriched water.⁵ 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 FishTM, a fish shaped iron ingot, which effectively reduced anemia among Cambodian women (Charles et al., 2011, 2015).⁶

⁵The safety of using such an iron ingot for cooking is documented by Armstrong, Dewey and Summerlee (2017).

⁶The leaf shape of the ingot was the result of focus group discussions we conducted in the study area prior to the baseline survey. The Lucky Iron Leaf and the Lucky Iron FishTM 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 FishTM

We altered the shape of the iron ingot to increase its acceptance in the Indian context. The Lucky Iron Leaf resembles a leaf of the *tulsi* tree which is a holy plant in Hinduism and commonly used for herbal tea against cough and colds. Male facilitators distributed the Lucky Iron Leaf to participants at no cost and provided instructions on its use during one home visit. The roll-out of the Lucky Iron Leaf closely followed the Lucky Iron FishTM implementation schedule of the trials in Cambodia (see appendix IX for further intervention details).

The iron ingot is a low dose fortification tool and, therefore, overdosing with the Lucky Iron Leaf is unlikely. Prior laboratory tests conducted at the University of Guelph 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 about half of an adult-equivalent portion. 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-years 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 through a reduction of iron deficiency and iron deficiency anemia. Iron deficiency can directly influence the neurodevelopment of children because iron is needed for myelination (white mass production) 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 thriving behavior and is, therefore, closely linked to motivation and learning (Flagel et al., 2011; Wise, 2004). Iron deficient children are also less attentive and more fearful than their peers and, hence, interact less with their environment, which can further inhibit learning and mental development (Lozoff et al., 1998). These biological and behavioral mechanisms put iron deficient children at risk of poorer cognitive, language, motor and socioemotional development (Lozoff, 2007; Beard, 2008).

A positive side effect of the Lucky Iron Leaf intervention might be general awareness raising of caregivers for child nutrition and anemia, which positively impacts complementary feeding practices (Dewey and Adu-Afarwuah, 2008).

Enterprise for the purpose of this study. For more information, see <https://luckyironfish.com>.

III EXPERIMENTAL DESIGN AND DATA

III.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.⁷ 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.

We randomly assigned the 1,995 households 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 ($N = 499$). The factorial design allows us to capture potential interactions of the nutrition and psychosocial stimulation interventions. We expect an increase in iron intake through the use of the Lucky Iron Leaf to improve children's brain growth and, thus, mental development (Lozoff, 2007; Beard, 2008). Additionally, children's concentration and participatory capacities to actively engage in book sharing may be improved, resulting in enhanced learning outcomes from dialogic reading (Flagel et al., 2001; Wise, 2004; Lozoff 2007; ortiz). As a result, treatment effects of the combined intervention may be a multiplicative product of both interventions, rather than the sum of each stand-alone intervention.

Randomization was conducted prior to the baseline survey because the intervention implementation was conducted subsequent to the survey. After the baseline survey and the first household visits, treatment groups (i) and (iii) were further randomized into a calendar ($N = 499$) and no-calendar group ($N = 498$). The randomization was stratified by caregiver attendance of dialogic reading training sessions 1 and 2.

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

The experimental design and outcome indicators are registered at the American Economic Association's registry for randomized controlled trials (www.socialscienceregistry.org/trials/2696).

⁷In 2015/16, 75 percent of all pregnant women in Madhepura had registered their pregnancies (International Institute for Population Sciences, 2017).

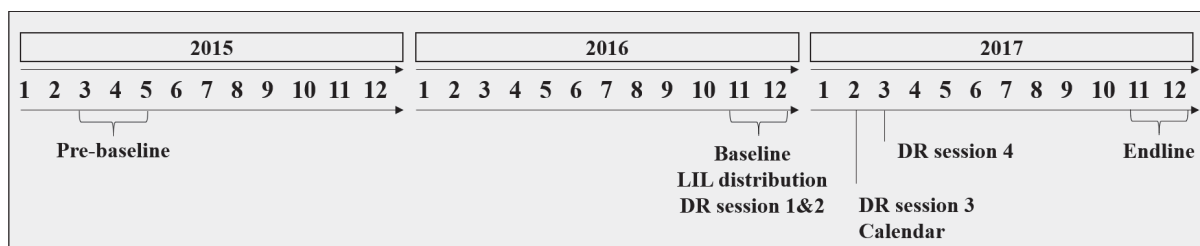


Figure 1: Timeline of data collections and intervention implementation

However, no pre-analysis plan, detailing the estimation strategy or heterogeneous treatment effect analyses, was included. Therefore, this paper thoroughly discusses the robustness of results to alternative estimation strategies or result interpretations.

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 women's permanent household location, resulting from the tradition of women moving to their parental home for pregnancy and child birth. Further households dropped out of the sample because not all conceived children had survived until baseline data collection in 2016.⁸ 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, biases from selective attrition at this stage are unlikely.

The final estimation sample consists of 1,164 households, corresponding to 78.5 percent of baseline households.⁹ 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). Additional but less common causes of attrition are migration of households across village borders, incorrect household identification information and child death. Ten households refused their participation.

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 approaches 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).

⁸The magnitude of child deaths is comparable to the state-wide statistics. The infant mortality rate in rural Bihar is 50 per 1,000 live birth and only 92 percent of all pregnancies lead to live birth in Bihar (Kochhar et al., 2014; International Institute for Population Sciences and ICF, 2017).

⁹We define the estimation sample as number of observations with at least one non-missing child development score.

III.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 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 to the Bihari context 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. 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 proxied by hemoglobin (Hb) levels and the presence of anemia.¹⁰ 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 months 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 caregivers to seek treatment at the local primary health center or with the community health worker and to change the diet towards more iron rich foods. Severely anemic children ($Hb < 7$ g/dl) were offered transport and cost coverage 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. Interaction effects of baseline testing and the Lucky Iron Leaf intervention can potentially occur. 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, 2019; Krämer, Kumar and Vollmer, 2020). Additionally, we test for heterogeneous take-up by baseline anemia status and do not find differential effects.

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

¹⁰While 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 2.01	283	5.77 2.30	-0.04 0.63	299	5.75 2.26	-0.03 0.71	294	5.75 2.34	-0.03 0.72	285
Hindu	0.87 0.34	283	0.82 0.38	0.13 0.13	300	0.85 0.36	0.05 0.52	295	0.86 0.34	0.02 0.84	286
Caste category:											
Scheduled caste	0.30 0.46	282	0.30 0.46	-0.01 0.89	300	0.31 0.46	-0.02 0.85	295	0.28 0.45	0.04 0.65	285
Scheduled tribe	0.04 0.19	282	0.04 0.20	-0.01 0.95	300	0.02 0.15	0.09 0.29	295	0.04 0.20	-0.02 0.85	285
Other backward class	0.59 0.49	282	0.58 0.49	0.02 0.77	300	0.58 0.49	0.02 0.82	295	0.61 0.49	-0.04 0.66	285
General category	0.07 0.26	282	0.08 0.27	-0.02 0.79	300	0.09 0.28	-0.06 0.45	295	0.07 0.25	0.02 0.84	285
Highest education in HH:											
No schooling	0.43 0.50	283	0.49 0.50	-0.10 0.21	300	0.47 0.50	-0.08 0.34	295	0.40 0.49	0.07 0.38	286
Primary	0.19 0.39	283	0.15 0.35	0.11 0.19	300	0.16 0.37	0.07 0.38	295	0.17 0.37	0.05 0.54	286
Middle school	0.13 0.34	283	0.13 0.34	0.00 0.98	300	0.15 0.35	-0.04 0.60	295	0.17 0.38	-0.11 0.18	286
≥ High school	0.25 0.43	283	0.24 0.43	0.02 0.76	300	0.22 0.42	0.06 0.44	295	0.26 0.44	-0.03 0.68	286
Asset quintile ^a	5.31 2.88	282	5.41 2.90	-0.04 0.66	300	5.01 2.88	0.10 0.22	293	5.23 2.94	0.03 0.75	285
Housing quintile ^b	3.76 2.77	283	3.70 2.73	0.02 0.79	300	3.48 2.75	0.10 0.23	295	3.62 2.82	0.05 0.55	286
Mother characteristics:											
Age in years	24.99 3.98	283	24.75 4.15	0.06 0.48	300	24.56 3.74	0.11 0.18	295	25.07 4.03	-0.02 0.80	286
Completed education:											
No schooling	0.74 0.44	283	0.76 0.43	-0.06 0.49	300	0.84 0.36	-0.26 0.00	295	0.73 0.44	0.01 0.91	286
Primary	0.09 0.29	283	0.05 0.22	0.16 0.05	300	0.03 0.18	0.24 0.00	295	0.07 0.25	0.09 0.26	286
Middle school	0.05 0.22	283	0.08 0.27	-0.10 0.25	300	0.04 0.19	0.08 0.36	295	0.10 0.31	-0.19 0.02	286
High school or higher	0.12 0.32	283	0.11 0.31	0.02 0.80	300	0.08 0.28	0.11 0.20	295	0.09 0.29	0.07 0.39	286
Can read SMS	0.27 0.44	283	0.28 0.45	-0.03 0.75	300	0.18 0.39	0.20 0.02	295	0.28 0.45	-0.03 0.69	286
Worked past 12 months	0.91 0.29	283	0.89 0.32	0.06 0.48	300	0.91 0.30	-0.00 0.99	295	0.91 0.30	0.01 0.92	286
Empowerment ^c	0.37 0.48	283	0.46 0.50	-0.20 0.02	300	0.38 0.49	-0.02 0.83	295	0.40 0.49	-0.06 0.45	286
Decides child nutrition	0.52 0.50	267	0.57 0.50	-0.11 0.21	279	0.54 0.50	-0.05 0.57	264	0.49 0.50	0.05 0.60	251

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 respective treatment group. p-values refer to a t-test of the equality of means of the control group and the respective treatment group and are presented below the Std. Diff.. ^a10 quintiles based on a durable asset index generated by factor analysis. ^b10 quintiles based on a housing quality index generated by factor analysis. ^cIndicator equals one if the mother is allowed to go alone to one of five places (market, health facility, neighbor's, 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 respective treatment group. p-values refer to a t-test of the equality of means of the control group and the respective treatment group and are presented below the Std. Diff. ^dSum of stimulating activities conducted with the child in the past 3 days. ^eEquals 1 if the caregiver explains wrong behavior to child, takes away privileges or gives child something else to do. ^fEquals 1 if the caregiver shouts, yells or screams at the child or spanks, hits, kicks or slaps the child.

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 received iron supplements in the last three months. Of the eight stimulating activities we enquired about, 5.6 activities were conducted with the 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 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.

IV ESTIMATION STRATEGY

The main analysis focuses on the evaluation of the four main treatment arms on children’s outcomes. The dialogic reading calendar will be considered with respect to intervention take-up in section VI.A. We estimate intention-to-treat effects using the following specification:

$$y_i = \alpha + \beta_1 DR_i + \beta_2 LIL_i + \beta_3 DR\&LIL_i + X_i' \gamma + u_i. \quad (1)$$

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 main effect estimation and four heterogeneous treatment effect estimations 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 hypotheses 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 hypotheses 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 hypotheses testing correction) and 11.9 percentage points in anemia prevalence (15.4 percentage points after multiple hypotheses testing correction), given a sample size of 260 children per treatment arm. The multiple hypotheses 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).

V RESULTS

VA 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 statistically insignificant across the three treatment arms, outcomes and specifications. The adjustment for covariates causes a slight and insignificant increase in the effect size relative to the simple model across all development scales. When we additionally control for baseline outcomes, the magnitude 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). Further, we exploit the panel structure of our data and estimate treatment effects using difference-in-differences, child fixed effects and inverse probability weighted estimations. At last, 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).¹¹ The robustness checks confirm the findings of Table II.

¹¹Head circumference was standardized according to WHO growth reference tables (World Health Organization,

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

	Model 1	Model 2	Model 3
Cognitive			
DR	0.029 (0.087)	0.107 (0.085)	0.082 (0.089)
LIL	-0.067 (0.088)	-0.001 (0.085)	-0.016 (0.089)
DR & LIL	0.037 (0.089)	0.064 (0.086)	0.083 (0.089)
Observations	1146	1136	1013
Adjusted R ²	-0.001	0.082	0.087
Control mean	-0.00	-0.00	-0.00
Receptive language			
DR	-0.046 (0.085)	-0.008 (0.085)	-0.027 (0.090)
LIL	-0.026 (0.085)	0.017 (0.085)	-0.022 (0.089)
DR & LIL	-0.008 (0.086)	0.018 (0.086)	0.001 (0.090)
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 (0.089)	0.035 (0.088)	0.040 (0.092)
LIL	-0.039 (0.089)	0.040 (0.088)	0.031 (0.092)
DR & LIL	0.020 (0.090)	0.036 (0.089)	0.056 (0.092)
Observations	1159	1148	1041
Adjusted R ²	-0.002	0.049	0.077
Control mean	-0.01	-0.01	0.00
Motor			
DR	-0.013 (0.085)	0.051 (0.083)	0.005 (0.087)
LIL	-0.022 (0.085)	0.020 (0.084)	-0.026 (0.087)
DR & LIL	0.029 (0.086)	0.059 (0.084)	0.015 (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 (0.081)	0.061 (0.080)	0.093 (0.086)
LIL	-0.022 (0.081)	0.040 (0.080)	0.085 (0.086)
DR & LIL	-0.017 (0.082)	0.012 (0.081)	0.019 (0.086)
Controls		✓	✓
Baseline outcome			✓
Observations	1151	1140	1017
Adjusted R ²	-0.002	0.039	0.044
Control mean	-0.01	-0.01	0.00

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.

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

The causal impact chain of the Lucky Iron Leaf on child development runs through iron deficiency. The null results of the Lucky Iron Leaf on child development suggest no improvements in hemoglobin levels. In Table III, we formally test the treatment impact on hemoglobin, any type of anemia, and moderate anemia. We use linear probability models to estimate the 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 than expected. This is a result of 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 by 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.

Failure to reduce iron deficiency implies early disruptions in the causal chain from intervention implementation to its impact on anemia and mental development. If there had been temporary improvements in iron deficiency, 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, known for its representation of nutrition history (Black et al., 2013), as well as on wasting and underweight.¹² The results show that the Lucky Iron Leaf intervention also had no impact on growth outcomes (see appendix Table A.III.2).¹³

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, statistically significant before multiple hypotheses 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

2007).

¹²Growth outcomes were derived in accordance with WHO growth reference tables (World Health Organization, 2007)

¹³The 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 (0.119)	-0.177 (0.119)	-0.121 (0.125)
LIL	0.003 (0.120)	-0.010 (0.120)	-0.132 (0.127)
DR & LIL	0.025 (0.119)	-0.012 (0.120)	-0.158 (0.125)
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 (0.043)	0.073* (0.043)	0.112** (0.048)
LIL	0.008 (0.044)	0.016 (0.044)	0.064 (0.049)
DR & LIL	-0.018 (0.043)	-0.004 (0.044)	0.045 (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 (0.039)	0.035 (0.040)	0.030 (0.046)
LIL	0.006 (0.040)	0.003 (0.040)	0.044 (0.047)
DR & LIL	-0.009 (0.040)	-0.001 (0.040)	0.054 (0.046)
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.

for by the positive effects of sharing books.

V.C Heterogeneous treatment effects

The effectiveness of the interventions potentially differs by participants' characteristics. We assess heterogeneous treatment effects by mothers' education and empowerment, and children's sex and anemia status at baseline. No heterogeneous tests were pre-registered in order not to miss important results ex-ante (Almås et al., 2018; Banerjee et al., 2020). 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., 2014; Ofosu and Posner, 2020). To attenuate such concerns, for each tested heterogeneity, we present theoretical underpinnings that directly emerge 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*).¹⁴ For India, Attanasio, Meghir and Nix (2020) find that parental cognition improves the productivity of investments on children's cognition, but not on health. 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*

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. Similarly, in the context of this study, women's empowerment may matter for the regular use of picture books and the Lucky Iron Leaf. More empowered women may choose to spend their time on book sharing or decide to alter cooking habits in order to incorporate the use of the Lucky Iron Leaf. Further, more empowered women can participate more confidently 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 that on children's outcomes

¹⁴In 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.

have been widely documented in India. Already early in life, daughters are breastfed for shorter time, receive less childcare time, vaccinations, and vitamin supplements, are less likely to be hospitalized, are shorter, have lower 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 Pande, 2017; Ebert and Vollmer, 2019). Therefore, caregivers may also choose to invest more or better quality book sharing into their sons than daughters or tend to rather feed the fortified meals to their sons than daughters.

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. Biologically, iron deficiency can have multiple effects that adversely alter brain growth and behavior (Lozoff, 2007; Beard, 2008). Iron deficient children are tiring easily, are more wary, hesitant, unhappy and have less social interactions (Lozoff, 2007). 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). Also economists studying the production function of child development, show that good health positively affects parental investment in children as well as later health and cognitive outcomes (Attanasio, Meghir and Nix, 2020; Attanasio et al., 2020a). The second dimension along which children's baseline anemia may affect children's hemoglobin level and, therefore, mental development, is parents' behavioral response to the results of their children's baseline anemia testing. The knowledge of, or being reminded about, your child's anemia status may evoke parents' motivation 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 insignificant (see appendix Table A.IV.1).¹⁵ 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 pro-

¹⁵The 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 hypotheses 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 hypotheses 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.

vides some evidence opposing our third hypothesis (see appendix Table A.IV.2). Although not significant after multiple hypotheses 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 than girls to improve in 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 hypotheses 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 hypotheses testing. The reason we find a salient effect in the combined intervention group might be that the additional home visit to distribute the Lucky Iron Leaf, which focused on children's health, reinforced the dialogic reading training.

Figure 2 visualizes the treatment effect heterogeneity by hemoglobin level of the combined intervention on receptive language skills. The marginal treatment effect at each hemoglobin level is estimated based on a model in which each treatment indicator is interacted with each hemoglobin level (rounded to the nearest integer). 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. Interestingly, the heterogeneity in treatment effects exactly corresponds to the WHO anemia cut-off for children below 59 months. Children with a hemoglobin level of 11 g/dl or above are considered non-anemic and show higher treatment effects on receptive language than anemic children do.

The results are in line with the hypothesis that non-anemic children 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 the subsequent section V.D.

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, though all coefficients are insignificant (see appendix Table A.IV.2).

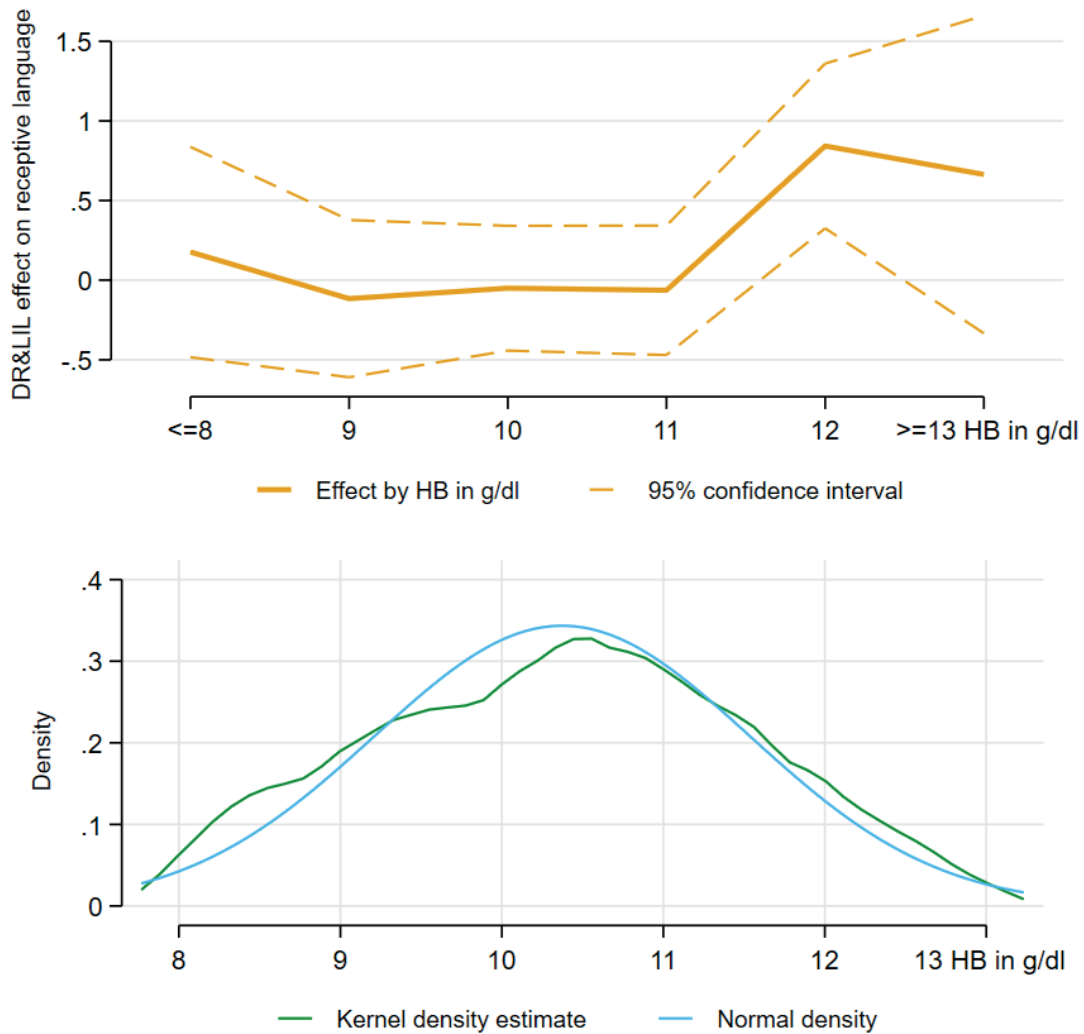


Figure 2: Effect of combined intervention on receptive language by baseline hemoglobin in g/dl

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.

V.D The role of anemia for intervention benefits

The heterogeneous treatment effects by baseline anemia status imply a cross-productivity of baseline health and intervention benefit. Yet, anemia status may correlate with a number of other characteristics, which could confound the heterogeneous treatment effect by anemia and, therefore, undermine our hypothesis on non-anemic children’s active engagement in book sharing. We test this hypothesis in three ways. First, we directly investigate whether non-anemic children en-

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)	0.042 (0.050)	-0.076* (0.042)	-0.016 (0.058)	0.117** (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: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

gage more actively in book sharing. Specifically, we regress baseline anemia on self-reported book sharing characteristics at endline in Table IV. Column (1) shows that 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 believe more often 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.

Second, we check the baseline balance of receptive language within the sample of anemic and non-anemic children, respectively. If non-anemic children in the treatment group have higher receptive language skills than control group children, 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 group 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 re-estimate 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. Therefore, we are not concerned that the heterogeneous treatment effects by anemia are relevantly driven by confounding factors.

VI UTILIZATION AND SPILLOVERS

VI.A *Utilization rates*

The tested interventions offer tools and knowledge to improve early childhood development, whereas strict compliance is not enforceable. 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 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 decent. Ninety-four percent of caregivers were able to freely name a specific content, character or page in her favorite book and 64 percent identified all intervention-book covers correctly when presented to them together with three random 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 used the Lucky Iron Leaf for cooking in the previous seven days and

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		All		Received	
	Mean	N	Mean	N	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 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)).

were able to present the Lucky Iron Leaf 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 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 prolonged cooking (11%), loss of the 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 treat-

ment arm. Utilization rates are similar to those in the individual treatment groups, presented in Panel 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 VIII.

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

We test the effectiveness of the dialogic reading calendar to increase book sharing utilization. Random assignment to the calendar was conditional on having participated in at least one dialogic reading training session and hence being revisited in March 2017 for training sessions 3 and 4. Therefore, the estimation sample is restricted to 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 the analysis, we pooled the pure dialogic reading and combined intervention group households to one calendar treatment arm and one no-calendar control arm. 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 dialog 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.

VI.C Spillovers

Because treatments were individually assigned, about 38 percent of treatment households have at least one control household in 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. Treatment contamination may occur, for example, when treatment households give away intervention picture books to control households. John Henry ef-

Table VI: Intervention take-up by calendar assignment

	≥ 4 times	Past 7 days
Calendar	-0.011 (0.042)	0.040 (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)), whether books were shared in the week prior to endline data collection (column (2)), and whether an intervention book was present at endline (column (3)). 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 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.

fects 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.

In order to quantify potential spillover effects, we rerun our intention-to-treat estimations and add the saturation rate in treated units within 100 meters distance to each household, following Baird et al. (2016). The saturation rate is the number of households within the specified distance that received the respective intervention divided by all study households in the same distance.¹⁶

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

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

For the sake of readability, we run the regression separately for each treatment group and present the spillover analysis for 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 in 100 meters distance and thus controls for the population density. The control vector X_i' includes the standard set of covariates.

¹⁶Ideally, 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.

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 deviations 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 increases from 0 to 1, the development score of children in families 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 statistically insignificant.

Additional descriptive statistics provide evidence on the pathways of spillovers. Whereas only four percent of dialogic reading households affirmed to have 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 to have 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 to identify in Table II.

VI.D Complier Average Causal Effects

Given imperfect utilization rates, intention-to-treat effects are not conclusive about the interventions' efficacy. The standard procedure to estimate improvements in child outcomes for intervention compliers would employ 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, assign-

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	0.266** (0.134)	0.103 (0.082)				
LIL			0.220 (0.138)	0.047 (0.085)		
DR&LIL					0.204 (0.143)	0.073 (0.087)
DR saturation	0.470*** (0.168)		0.446** (0.174)		0.489*** (0.176)	
DR sat. x DR	-0.596** (0.245)					
DR sat. x LIL			-0.537** (0.262)			
DR sat. x DR&LIL					-0.720*** (0.263)	
LIL saturation	-0.214 (0.175)		-0.236 (0.181)		-0.188 (0.184)	
LIL sat. x DR	0.139 (0.245)					
LIL sat. x LIL			0.022 (0.262)			
LIL sat. x DR&LIL					0.353 (0.255)	
Controls	✓	✓	✓	✓	✓	✓
Observations	547	547	541	541	531	531
Adjusted R ²	0.096	0.087	0.105	0.096	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.

ment 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 in 100 meters distance. 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 respective intervention in either the stand-alone or combined intervention groups.

We estimate two first stages:

$$DRcomp_i = \omega_0 + \delta_1 DRany_i + \delta_2 LILany_i + X_i' \lambda + v_i, \quad (3)$$

$$LILcomp_i = \mu_0 + \eta_1 DRany_i + \eta_2 LILany_i + X_i' \kappa + v_i, \quad (4)$$

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 any of the two potential treatment arms (stand-alone or combined), respectively.

The second stage is:

$$y_i = \theta_0 + \pi_1 \widehat{DRcomp}_i + \pi_2 \widehat{LILcomp}_i + X_i' \rho + \varepsilon_i, \quad (5)$$

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 to the child eating the fortified meals. The instruments' relevance is confirmed by the weak identification Cragg-Donald Wald F-statistics, which range between 25 and 28.

The treatment effects for compliers tend to be larger than the intention-to-treat effects in Tables II and III and are partially marginally significant before multiple hypotheses 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. Albeit consistent, these results should be taken cautiously 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.

VII IMPACT ON OTHER CAREGIVING DOMAINS

The dialogic reading training may affect caregiving beyond picture book reading in complementary or substituting ways.¹⁷ 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

¹⁷We 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.

Table VIII: Complier average causal effects

	Cognitive	Receptive	Expressive	Motor	Socio-emotional
DR ≥ 4 times per week	0.344* (0.175)	0.184 (0.181)	0.229 (0.189)	0.059 (0.175)	0.227 (0.177)
LIL ≥ 4 times per week	0.002 (0.440)	0.543 (0.445)	0.801* (0.473)	0.522 (0.438)	0.729* (0.437)
Controls	✓	✓	✓	✓	✓
Cragg-Donald Wald F statistic	28	28	28	27	28
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 (0.242)	-0.052 (0.090)			
LIL ≥ 4 times per week	1.010 (0.634)	-0.391* (0.235)			
Controls	✓	✓			
HemoCue machine fixed effects	✓	✓			
Cragg-Donald Wald F statistic	25	25			
Observations	423	423			
Control mean	10.58	0.57			

Note: Estimation sample is restricted to households that have no other study households in 100 meters distance. DR and LIL compliance are defined by having utilized the respective tool at least four times per week in the past 12 months and conditional to 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 sub-district 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: + $p < 0.0096$, ++ $p < 0.0048$, +++ $p < 0.0010$.

reading treatment groups. If dialogic reading substitutes 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 performed 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 summed these seven activities into an activity index ranging from zero to seven.¹⁸ 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

¹⁸The 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 (home-made) 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*** (0.042)	0.404*** (0.033)	0.229** (0.108)	0.021 (0.035)	0.121 (0.165)
LIL	-0.010 (0.042)	-0.005 (0.033)	-0.011 (0.109)	0.036 (0.035)	-0.060 (0.160)
DR & LIL	0.112*** (0.042)	0.378*** (0.034)	0.119 (0.110)	0.042 (0.035)	0.109 (0.164)
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 spans or shouts at child. Column (5) outcome: aspiration is measured by the highest level of education mother's are wishing 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.

households as non-reading when they report to have no books at home to avoid social desirability bias. The effects increase dramatically in size to about 40 percentage points and reassure 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 activities more 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 were explained to caregivers to be of harm 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 performance in school later on. 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 little evidence that the dialogic reading intervention affected caregiving positively or negatively beyond picture book sharing.

VIII CONTEXTUAL BARRIERS TO TAKE-UP

The effectiveness of the Lucky Iron Leaf intervention was essentially compromised by low utilization rates. In this section, we discuss how health knowledge, dietary habits and family structures

could have adversely influenced take-up of the Lucky Iron Leaf.

A common obstacle to 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 constitutes a key source of motivation for its (continual) utilization. However, only seven percent of Lucky Iron Leaf group respondents positively responded to the question “Have you heard of anemia?” twelve months after intervention roll-out (and four percent in the control group). Possibly, respondents did not know the term “anemia” but, nevertheless, were familiar with the concept of iron deficiency and its implications for children’s health and development. In light of low overall health knowledge in 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 successfully raise anemia awareness in a sustainable way and, therefore, to convey the need for increased iron intake. In contexts with low levels of education and health knowledge, more intensive or more frequent information sessions may be required to induce behavioral changes.

Furthermore, the use of acid food items in the daily meal preparation is a decisive factor for the effectiveness of the Lucky Iron Leaf. Fruit acid is essential for the iron to leach from the ingot into the water. While the majority of Lucky Iron Leaf households was aware of this requirement, only a quarter reported the consumption of any tomato or lemon in the week before the endline survey. This dietary pattern reflects the overall low consumption of fruits and vegetables in northeast India (Sharma et al., 2020). In Cambodia, where the same intervention (using a fish shaped iron ingot) led to sustained utilization by 90 percent of participants, fruit acids are regularly used for a variety of traditional meals such as soups soured with lemon juice (Charles et al., 2011, 2015). The required behavioral change to accommodate the Lucky Iron Leaf is therefore smaller in Cambodia than in India. However, the Cambodian trials also monitored utilization more closely, which might have affected utilization additionally.

Lastly, women’s empowerment potentially plays an important role for children’s nutrition. In our study sample, only one half of 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.

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 easily remove rust from the iron ingot. Further, the Lucky Iron Leaf manual instructs 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 looks rusty. Thus, we are not concerned that rust or alterations of taste and color of meals essentially deterred take-up.

IX 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 as a risk factor of early childhood development 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 as a risk factor of early childhood development and consists of four at home trainings of caregivers in dialogic reading methods of picture book sharing. 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.

One year after implementation, neither intervention nor their combination have improved anemia or mental functions of children on average. However, we find a cross-productivity of children's initial health status and psychosocial stimulation. Children in the combined intervention group, who were non-anemic at baseline (30% of study sample), improved their receptive language skills by 0.54 standard deviations. We conjecture that the impact in the combined intervention group was driven by the dialogic reading component, because the utilization of the Lucky Iron Leaf was low and we find no improvements in hemoglobin among non-anemic children (or any children). The home visit for the distribution of the Lucky Iron Leaf was focused on children's health and might have reinforced the dialogic reading trainings, resulting in a more pronounced cross-productivity of baseline health and intervention impact of the combined treatment. The relationship between anemia and dialogic reading impact does not seem to be driven by confounding health or socioeconomic indicators. Biologically, non-anemic children likely have an advantage in the efficacy of book sharing over their anemic peers, which is that they can engage more actively, whereas anemic children tend to be tired and less exploratory. Our results are in line with Tofail et al. (2013), who find that a psychosocial stimulation intervention in Bangladesh was effective for non-anemic children only.

Given the simplicity of our interventions, the effect size on non-anemic children is consider-

able. In Colombia, Attanasio et al. (2014) find improvements from an intensive 18 to 24 months weekly home-visit program in receptive language and cognition of 0.22 and 0.26 standard deviations, respectively. Assuming the Colombian iron deficiency anemia prevalence of 19 percent and the Indian coefficients, the hypothetical average effect is 0.44 standard deviations ($0.81 * 0.54 + 0.19 * 0$). While comparisons across the Colombian and Indian contexts are difficult to draw, the exercise demonstrates the magnitude of the identified effect given the simplicity of the interventions. Yet, the interventions tested in this study only significantly affected receptive language, whereas the comprehensive stimulation program in Columbia also improved cognition (0.26 SD) and a similar program in Bangladesh affected behavior (0.18 SD in vocalization to 0.45 SD in response to examiner).

The effectiveness of the Lucky Iron Leaf intervention was essentially compromised by low take-up rates. Utilization of the iron fortification device was reported to be 25 percent initially and decreased to 3 percent a year later. The local cultural and nutritional context likely mattered for intervention take-up. The main barriers reported by non-users were the dislike of the iron ingot among husbands and parents-in-law and the required use of fruit acid-rich items in the fortification process. The use of fruit acid-rich items is rare in rural Bihar and the intervention may have not sufficiently induced its use when cooking with the Lucky Iron Leaf.

In addition to the context-specificity of most randomized controlled trials, the main limitation of our study is the random assignment of households to treatment arms, which bears the risk of spillovers across treatment and control groups. We follow Baird et al. (2016) and disentangle the pure treatment effect from spillover effects by adding the saturation in opposite treatment households to the intention-to-treat estimation model. The pure intention-to-treat effect across treatment arms on children's mental functions increase 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. In future trials, cluster randomization or randomized treatment intensity variation may be used to control the contamination of untreated households.

Overall, our results imply that more short-lived psychosocial stimulation interventions can potentially be effective in populations where the prevalence of anemia among children is low. It thereby adds to a growing literature on bringing early childhood development programs to scale. Previous approaches successfully used existing infrastructure to scale long-lasting parenting programs (Attanasio et al., 2014; Yousafzai et al., 2014; Andrew et al., 2020; Attanasio et al., 2020b; Sylvia et al., 2020). Our approach may pose an alternative program design in contexts where supply-side compliance is expected to be low. Replications of this approach are required to further

explore this alternative way to scale.

REFERENCES

- 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*.
- Alderman, Harold, Jere R Behrman, Victor Lavy, and Rekha Menon.** 2001. “Child health and school enrollment: A longitudinal analysis.” *The Journal of Human Resources*, 36(1): 185–205.
- Almås, Ingvild, Orazio Attanasio, Jyotsna Jalan, Francisco Oteiza, and Marcella Vigneri.** 2018. “Using data differently and using different data.” *Journal of Development Effectiveness*, 10(4): 462–481.
- 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.
- Atherton, Paul, Simon Appleton, and Michael Bleaney.** 2013. “International school test scores and economic growth.” *Bulletin of Economic Research*, 65(1): 82–90.

- 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.** 2018. “Early stimulation and nutrition: The impacts of a scalable intervention.” National Bureau of Economic Research.
- 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.** 2020a. “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.** 2020b. “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.
- Banerjee, Abhijit, Esther Duflo, Amy Finkelstein, Lawrence F Katz, Benjamin A Olken, and Anja Sautmann.** 2020. “In praise of moderation: Suggestions for the scope and use of pre-analysis plans for RCTs in economics.” National Bureau of Economic Research.
- 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.

- Behrman, Jere R, John Hoddinott, John A Maluccio, Erica Soler-Hampejsek, Emily L Behrman, Reynaldo Martorell, Manuel Ramírez-Zea, and Aryeh D Stein.** 2014. “What determines adult cognitive skills? Influences of pre-school, school, and post-school experiences in Guatemala.” *Latin American Economic Review*, 23(1): 4.
- Behrman, Jere R, Maria C Calderon, Samuel H Preston, John Hoddinott, Reynaldo Martorell, and Aryeh D Stein.** 2009. “Nutritional supplementation in girls influences the growth of their children: Prospective study in Guatemala.” *The American Journal of Clinical Nutrition*, 90(5): 1372–1379.
- 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.
- 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.
- 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 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, and James J Heckman.** 2008. “Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation.” *The Journal of Human Resources*, 43(4): 738–782.
- Cunha, Flavio, James J Heckman, and Susanne M Schennach.** 2010. “Estimating the technology of cognitive and noncognitive skill formation.” *Econometrica*, 78(3): 883–931.
- Cunha, Flavio, James J Heckman, Lance Lochner, and Dimitriy V Masterov.** 2006. “Interpreting the evidence on life cycle skill formation.” *Handbook of the Economics of Education*, 1: 697–812.
- 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.
- Denny, Kevin, Colm Harmon, and Vincent O’Sullivan.** 2004. “Education, earning and skills: A multi-country comparison.” IFS Working Papers, Institute for Fiscal Studies (IFS).
- 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.** 2019. “Child-specific son preference, birth order and cognitive skills in early childhood.” Courant Research Centre: Poverty, Equity and Growth-Discussion Papers, No. 264.
- Fernald, Lia C H, Patricia Kariger, Melissa Hidrobo, and Paul J Gertler.** 2012. “Socioeconomic gradients in child development in very young children: Evidence from India, Indonesia, Peru, and Senegal.” *Proceedings of the National Academy of Sciences*, 109(Supplement 2): 17273 LP — 17280.
- 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*.
- 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.
- Glewwe, Paul, Hanan G Jacoby, and Elizabeth M King.** 2001. “Early childhood nutrition and academic achievement: A longitudinal analysis.” *Journal of Public Economics*, 81(3): 345–368.

- Grantham-McGregor, Sally, Akanksha Adya, Orazio Attanasio, Britta Augsborg, Jere Behrman, Bet Caeyers, Monimalika Day, Pamela Jarvis, Reema Kochar, Perna 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.
- Hamadani, Jena D, Fahmida Tofail, Syed N Huda, Dewan S Alam, Deborah A Ridout, Orazio Attanasio, and Sally M Grantham-McGregor.** 2014. “Cognitive deficit and poverty in the first 5 years of childhood in Bangladesh.” *Pediatrics*, 134(4): e1001 LP – e1008.
- Hamadani, Jena D, Syed N Huda, Fahmida Khatun, and Sally M Grantham-McGregor.** 2006. “Psychosocial stimulation improves the development of undernourished children in rural Bangladesh.” *The Journal of Nutrition*, 136(10): 2645–2652.
- Hanushek, Eric A.** 2013. “Economic growth in developing countries: The role of human capital.” *Economics of Education Review*, 37: 204–212.
- Hanushek, Eric A, and Dennis D Kimko.** 2000. “Schooling, labor-force quality, and the growth of nations.” *American Economic Review*, 90(5): 1184–1208.
- Hanushek, Eric A, and Ludger Woessmann.** 2008. “The role of cognitive skills in economic development.” *Journal of Economic Literature*, 46(3): 607–68.
- Hanushek, Eric A, and Ludger Woessmann.** 2012a. “Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation.” *Journal of Economic Growth*, 17(4): 267–321.
- Hanushek, Eric A, and Ludger Woessmann.** 2012b. “Schooling, educational achievement, and the Latin American growth puzzle.” *Journal of Development Economics*, 99(2): 497–512.
- 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.

- Hazarika, Gautam, and Vejoya Viren.** 2013. “The effect of early childhood developmental program attendance on future school enrollment in rural North India.” *Economics of Education Review*, 34: 146–161.
- Heckman, James J.** 2007. “The economics, technology, and neuroscience of human capability formation.” *Proceedings of the National Academy of Sciences*, 104(33): 13250–13255.
- Heckman, James J, Jora Stixrud, and Sergio Urzua.** 2006. “The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior.” *Journal of Labor Economics*, 24(3): 411–482.
- Hoddinott, John, John A Maluccio, Jere R Behrman, Rafael Flores, and Reynaldo Martorell.** 2008. “Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults.” *The Lancet*, 371(9610): 411–416.
- 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.
- 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.** 2019. “Anemia, diet, and cognitive development: Impact of health information on diet quality and child nutrition in rural India.” *Courant Research Centre 'Poverty, Equity and Growth Discussion Paper Series*.
- Krämer, Marion, Santosh Kumar, and Sebastian Vollmer.** 2020. “Improving child health and cognition: Evidence from a school-based nutrition intervention in India.” *The Review of Economics and Statistics*, forthcoming.
- 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.
- Lazuka, Volha.** 2020. “Infant health and later-life labor market outcomes: Evidence from the introduction of sulfa antibiotics in Sweden.” *Journal of Human Resources*, 55(2): 660–698.
- Lee, Doo Won, and Tong Hun Lee.** 1995. “Human capital and economic growth tests based on the international evaluation of educational achievement.” *Economics Letters*, 47(2): 219–225.
- 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. DPI3725*.
- 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.
- Maluccio, John A, John Hoddinott, Jere R Behrman, Reynaldo Martorell, Agnes R Quisumbing, and Aryeh D Stein.** 2009. "The Impact of improving nutrition during early childhood on education among Guatemalan adults." *The Economic Journal*, 119(537): 734–763.
- 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.
- Ofori, 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.

- Reynolds, Sarah A, Chris Andersen, Jere Behrman, Abhijeet Singh, Aryeh D Stein, Liza Benny, Benjamin T Crookston, Santiago Cueto, Kirk Dearden, Andreas Georgiadis, Sonya Krutikova, and Lia C H Fernald.** 2017. “Disparities in children’s vocabulary and height in relation to household wealth and parental schooling: A longitudinal study in four low- and middle-income countries.” *SSM - Population Health*, 3: 767–786.
- 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.
- Rubio-Codina, Marta, Orazio Attanasio, Costas Meghir, Natalia Varela, and Sally Grantham-McGregor.** 2015. “The socioeconomic gradient of child development: Cross-sectional evidence from children 6-42 months in Bogota.” *Journal of Human Resources*, 50(2): 464–483.
- 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.
- Schady, Norbert, Jere Behrman, Maria Caridad Araujo, Rodrigo Azuero, Raquel Bernal, David Bravo, Florencia Lopez-Boo, Karen Macours, Daniela Marshall, Christina Paxson, et al.** 2015. “Wealth gradients in early childhood cognitive development in five Latin American countries.” *Journal of Human Resources*, 50(2): 446–463.
- 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.
- 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.
- Taylor, Beck A, Eric Dearing, and Kathleen McCartney.** 2004. “Incomes and outcomes in early childhood.” *Journal of Human Resources*, 39(4): 980–1007.

- Thomas, Duncan.** 1993. “The distribution of income and expenditure within the household.” *Annals of Economics and Statistics*, , (29): 109–135.
- Thompson, Owen.** 2018. “Head start’s long-run impact: Evidence from the program’s introduction.” *Journal of Human Resources*, 53(4): 1100–1139.
- 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 Wright, Clive Osmond, and Sally M Grantham-McGregor.** 2015. “Early childhood stunting is associated with lower developmental levels in the subsequent generation of children.” *The Journal of Nutrition*, 145(4): 823–828.
- Walker, Susan P, Susan M Chang, C A Powell, and Sally M Grantham-McGregor.** 2005. “Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: Prospective cohort study.” *The Lancet*, 366(9499): 1804–1807.

- Walker, Susan P, Susan M Chang, Marcos Vera-Hernández, and Sally Grantham-McGregor.** 2011a. “Early childhood stimulation benefits adult competence and reduces violent behavior.” *Pediatrics*, 127(5): 849 LP — 857.
- 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.** 2011b. “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.
- 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.

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 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

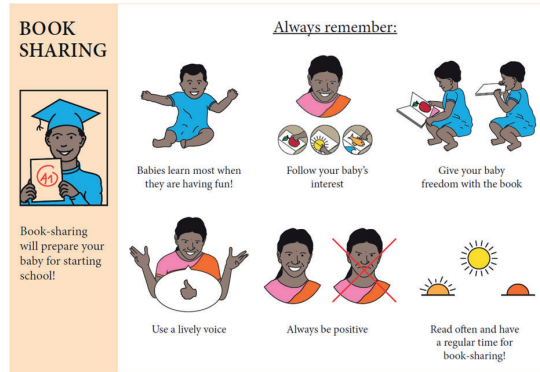
Session	Concepts
1	Having fun Follow babies interest Freedom with the book Using a lively voice Always be positive Practice regularly
2	Point and say Asking questions Repeating the word Elaborating Making and action
3	Making links Talking about experiences
4	Talking about feelings Making links about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

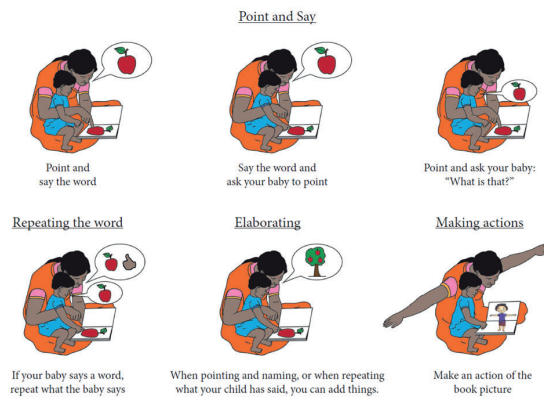
touch-and- feel book "The Very Hungry Caterpillar" produced by the Penguin publisher. The second book ("Little Painters") and third book ("Annual Haircut Day") were obtained from Pratham Books, a sub-organization of the education focused Indian NGO Pratham. All books were chosen based on their cultural 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-baseline - endline
DR	0.006 (0.032)	-0.035 (0.032)
LIL	0.019 (0.032)	-0.024 (0.033)
DR & LIL	0.018 (0.031)	-0.006 (0.030)
Observations	1480	1995
Adjusted R ²	0.00	0.00
F statistic	0.17	0.53

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. ^a10 quintiles based on a durable asset index generated by factor analysis. ^b10 quintiles based on a housing quality index generated by factor analysis. ^cIndicator 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. ^dThe sum of stimulating activities typically conducted with the child. ^eEquals 1 if caregiver explains wrong behavior to child, takes away privileges or gives child something else to do. ^fEquals 1 if caregiver shouts, yells or screams at child or spansks, 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 (0.078)	0.106 (0.085)	-0.016 (0.081)	-0.009 (0.085)	0.043 (0.085)	0.039 (0.088)	0.056 (0.078)	0.052 (0.082)	0.083 (0.071)	0.055 (0.080)
LIL	-0.010 (0.079)	0.022 (0.085)	0.012 (0.081)	0.029 (0.085)	0.059 (0.085)	0.060 (0.089)	0.007 (0.079)	0.038 (0.082)	0.054 (0.071)	0.044 (0.081)
DR & LIL	0.049 (0.079)	0.078 (0.086)	0.006 (0.082)	0.017 (0.087)	0.045 (0.086)	0.046 (0.090)	0.056 (0.079)	0.059 (0.083)	0.023 (0.072)	0.005 (0.082)
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 (0.097)	-0.009 (0.039)	-0.002 (0.031)	-0.034 (0.042)
LIL	-0.129 (0.097)	-0.044 (0.040)	0.024 (0.032)	-0.026 (0.043)
DR & LIL	0.062 (0.098)	-0.035 (0.039)	0.003 (0.031)	-0.070 (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 (0.144)	-0.140 (0.143)	-0.121 (0.125)	0.080 (0.052)	0.111** (0.052)	0.112** (0.048)	0.026 (0.048)	0.036 (0.049)	0.030 (0.046)
LIL	-0.118 (0.147)	-0.127 (0.145)	-0.132 (0.127)	0.060 (0.053)	0.060 (0.053)	0.064 (0.049)	0.050 (0.049)	0.051 (0.049)	0.044 (0.047)
DR & LIL	-0.074 (0.143)	-0.136 (0.143)	-0.158 (0.125)	-0.000 (0.052)	0.033 (0.052)	0.045 (0.048)	0.036 (0.048)	0.045 (0.048)	0.054 (0.046)
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 (0.137)	0.059 (0.050)	0.116 (0.097)	-0.063 (0.097)	0.052 (0.101)	0.067 (0.095)	0.083 (0.092)
DR x Primary	-0.028 (0.276)	0.035 (0.101)	-0.036 (0.196)	0.218 (0.195)	-0.080 (0.204)	-0.037 (0.191)	-0.101 (0.185)
LIL	0.099 (0.135)	-0.015 (0.049)	-0.065 (0.096)	-0.082 (0.095)	0.020 (0.099)	0.008 (0.094)	0.082 (0.090)
LIL x Primary	-0.504* (0.298)	0.137 (0.109)	0.400* (0.214)	0.512** (0.213)	0.169 (0.223)	0.147 (0.211)	-0.196 (0.203)
DR&LIL	-0.017 (0.138)	0.004 (0.051)	0.075 (0.099)	-0.068 (0.099)	0.065 (0.103)	0.133 (0.097)	0.050 (0.094)
DR&LIL x Primary	0.077 (0.274)	-0.049 (0.100)	-0.042 (0.195)	0.320 (0.194)	-0.142 (0.203)	-0.252 (0.191)	-0.177 (0.185)
Primary	0.167 (0.234)	0.002 (0.086)	0.325* (0.168)	-0.007 (0.168)	0.372** (0.176)	0.177 (0.165)	0.497*** (0.160)
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 (0.156)	0.096* (0.057)	0.185* (0.112)	0.075 (0.111)	-0.059 (0.116)	0.116 (0.110)	-0.053 (0.105)
DR x Empowered	0.140 (0.242)	-0.041 (0.088)	-0.179 (0.172)	-0.207 (0.172)	0.218 (0.178)	-0.144 (0.168)	0.277* (0.162)
LIL	0.093 (0.150)	-0.030 (0.055)	0.008 (0.108)	0.096 (0.107)	0.035 (0.111)	-0.001 (0.106)	-0.061 (0.101)
LIL x Empowered	-0.288 (0.248)	0.129 (0.091)	-0.023 (0.175)	-0.211 (0.175)	0.012 (0.182)	0.057 (0.171)	0.271 (0.165)
DR&LIL	0.073 (0.154)	-0.023 (0.056)	0.087 (0.110)	0.060 (0.110)	-0.004 (0.114)	0.076 (0.108)	-0.013 (0.104)
DR&LIL x Empowered	-0.211 (0.246)	0.049 (0.090)	-0.062 (0.177)	-0.114 (0.176)	0.105 (0.183)	-0.043 (0.173)	0.075 (0.167)
Empowered	0.092 (0.175)	-0.035 (0.064)	0.078 (0.125)	0.094 (0.125)	0.014 (0.130)	0.024 (0.122)	-0.103 (0.118)
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
Controls	✓	✓	✓	✓	✓	✓	✓
HemoCue machine fixed effects	✓	✓					

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 (0.173)	0.062 (0.063)	0.108 (0.123)	-0.026 (0.123)	-0.024 (0.128)	0.275** (0.120)	0.078 (0.116)
DR x Boy	-0.308 (0.236)	0.020 (0.086)	0.001 (0.169)	0.038 (0.168)	0.112 (0.175)	-0.427*** (0.165)	-0.028 (0.160)
LIL	0.069 (0.177)	-0.021 (0.065)	-0.034 (0.126)	-0.055 (0.126)	0.030 (0.131)	0.204* (0.124)	-0.019 (0.119)
LIL x Boy	-0.154 (0.241)	0.067 (0.088)	0.062 (0.171)	0.137 (0.170)	0.024 (0.178)	-0.347** (0.168)	0.108 (0.161)
DR&LIL	0.042 (0.169)	-0.007 (0.062)	0.022 (0.120)	-0.056 (0.120)	0.009 (0.125)	0.137 (0.118)	-0.037 (0.114)
DR&LIL x Boy	-0.104 (0.238)	0.003 (0.087)	0.086 (0.172)	0.151 (0.171)	0.055 (0.178)	-0.153 (0.168)	0.099 (0.162)
Boy	0.241 (0.169)	-0.030 (0.062)	-0.073 (0.121)	-0.089 (0.120)	-0.185 (0.125)	0.318*** (0.118)	-0.080 (0.114)
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 (0.240)	0.149* (0.088)	0.143 (0.193)	0.421** (0.195)	0.077 (0.204)	0.107 (0.189)	0.168 (0.187)
DR x Anemic	0.211 (0.284)	-0.052 (0.105)	0.012 (0.227)	-0.388* (0.230)	-0.075 (0.241)	0.017 (0.223)	-0.154 (0.220)
LIL	0.105 (0.244)	-0.012 (0.090)	0.301 (0.189)	0.330* (0.191)	0.207 (0.200)	0.147 (0.185)	0.074 (0.182)
LIL x Anemic	-0.344 (0.290)	0.107 (0.107)	-0.348 (0.226)	-0.418* (0.228)	-0.202 (0.240)	-0.028 (0.223)	0.026 (0.218)
DR&LIL	0.122 (0.233)	-0.032 (0.086)	0.451** (0.184)	0.538*** (0.185)	0.047 (0.195)	0.143 (0.181)	0.213 (0.177)
DR&LIL x Anemic	-0.433 (0.281)	0.112 (0.104)	-0.426* (0.222)	-0.596*** (0.223)	-0.039 (0.235)	-0.074 (0.218)	-0.288 (0.214)
Anemic	-1.052*** (0.203)	0.357*** (0.075)	0.238 (0.161)	0.326** (0.163)	-0.151 (0.171)	-0.060 (0.158)	-0.048 (0.156)
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)	0.60	0.09	0.21	0.79	0.99	0.31	0.91
p-value (LIL+interaction=0)	0.13	0.10	0.71	0.49	0.97	0.34	0.41
p-value (DR&LIL+int.=0)	0.05	0.17	0.85	0.64	0.95	0.58	0.54
Controls	✓	✓	✓	✓	✓	✓	✓
HemoCue machine fixed effects	✓	✓					

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 (0.008)						-0.007 (0.010)
Hindu	-0.017 (0.053)						0.041 (0.065)
Scheduled caste	0.120* (0.071)						0.099 (0.091)
Scheduled tribe	0.065 (0.108)						0.139 (0.138)
Other backward class	-0.003 (0.065)						0.004 (0.084)
Primary	-0.071 (0.051)						-0.075 (0.062)
Middle school	-0.041 (0.057)						-0.014 (0.070)
High school or higher	0.018 (0.060)						0.052 (0.074)
Asset index quintile	-0.001 (0.007)						0.001 (0.008)
Housing index quintile	0.003 (0.007)						-0.000 (0.008)
Mother characteristics:							
Age in years		-0.009** (0.004)					-0.012** (0.006)
Worked in past 12 months		-0.082 (0.058)					-0.063 (0.074)
Decides about child nutrition		0.054 (0.036)					0.069 (0.043)
Child characteristics:							
Currently breastfed			-0.021 (0.035)				-0.022 (0.042)
Vit-A past 6 months			-0.010 (0.038)				-0.014 (0.046)
Iron past 3 months			-0.018 (0.046)				-0.011 (0.055)
Age in months			-0.001 (0.007)				0.002 (0.009)
Home environment:							
Moderate stimulation				-0.082 (0.102)			-0.014 (0.128)
High stimulation				-0.127 (0.102)			-0.039 (0.128)
Good educational measures				-0.091 (0.060)			-0.160** (0.075)
Bad educational measures				0.140* (0.077)			0.073 (0.104)
Baseline outcomes:							
Cognitive					0.009 (0.021)		0.011 (0.024)
Receptive language					0.025 (0.018)		0.014 (0.021)
Expressive language					-0.023 (0.020)		-0.026 (0.024)
Motor					-0.011 (0.022)		0.006 (0.026)
Socioemotional					-0.027 (0.019)		-0.027 (0.023)
Child anthropometrics:							
Height (z-score)						-0.033** (0.014)	-0.035* (0.018)
Weight-for-age (z-score)						0.000 (0.013)	-0.002 (0.017)
Head circumference (z-score)						0.001 (0.016)	0.016 (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	1.56	1.92	1.28	1.51	1.39	1.85	1.05

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	0.421** (0.195)	0.430** (0.197)	0.440** (0.197)
DR x Anemic	-0.388* (0.230)	-0.369 (0.234)	-0.382 (0.233)
LIL	0.330* (0.191)	0.359* (0.196)	0.366* (0.195)
LIL x Anemic	-0.417* (0.228)	-0.441* (0.234)	-0.458** (0.233)
DR&LIL	0.538 ⁺ ** (0.185)	0.524 ⁺ ** (0.189)	0.493 ⁺ ** (0.189)
DR&LIL x Anemic	-0.596 ⁺ ** (0.223)	-0.556** (0.228)	-0.506** (0.228)
Anemic	0.326** (0.163)	0.315* (0.164)	0.321* (0.164)
Mother's age (years)			-0.011 (0.010)
Good educational measures			0.194 (0.125)
Bad educational measures			-0.176 (0.168)
Height-for-age (z-score)			0.051** (0.023)
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 ≥ 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.

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 (0.019)	0.105 (0.128)	-0.002 (0.026)	0.015 (0.042)	-0.001 (0.034)	0.042 (0.041)	-0.038 (0.039)
LIL	0.021 (0.019)	-0.073 (0.128)	-0.019 (0.026)	-0.059 (0.042)	0.007 (0.034)	0.059 (0.041)	-0.007 (0.040)
DR & LIL	0.028 (0.019)	0.001 (0.129)	-0.021 (0.026)	-0.015 (0.042)	-0.011 (0.034)	-0.001 (0.041)	-0.057 (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.