RUHR
ECONOMIC PAPERS

Cara Ebert
Esther Heesemann
Sebastian Vollmer

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

RWI #856

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: Ruhr Bachmann

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

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

RWI

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 Vollmer[1]

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

July 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 impede skill formation (Walker et al., 2007, 2011*b*). Early mental disadvantages can affect a child over a lifetime as the skill accumulation process is of dynamic and path-dependent nature. Early skills and parental investment matter interdependently in reinforcing and complementing ways for human capital outcomes later in life (Cunha et al., 2006; Cunha and Heckman, 2007; Heckman, 2007; Cunha and Heckman, 2008; Cunha, Heckman and Schennach, 2010; Attanasio et al., 2014, 2015; Attanasio, Meghir and Nix, 2017; Attanasio et al., 2020). In consequence, health and mental development follow a steep socioeconomic gradient which widens with age (Fernald et al., 2012; Hamadani et al., 2014; Rubio-Codina 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 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, violent 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; Alderman, Hoddinott and Kinsey, 2006; 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). An often-replicated two-year weekly stimulation program with one-year old children in Jamaica increased these children's earnings by 25 percent 20 years later (Gertler et al., 2014).

We study whether two low-cost interventions can improve cognitive, language, motor and socioemotional skills of one- to two-year old children in rural Bihar, India. The two interventions 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 Colombia, India and Bangladesh (Grantham-McGregor et al., 1991; Walker et al., 1991; Hamadani et al., 2006; Tofail et al., 2013; Attanasio et al., 2014). 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

[1] For 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).

Iron Leaf™ (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 for up to five years and does not require replenishment as conventional fortification technologies do (e.g., multi-micronutrient powder). The second intervention targets 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 with young children and stimulates speaking and learning. In order to regularly exercise dialogic reading, we distributed three durable picture books to families.

To improve scalability, previous programs have been integrated into existing infrastructure of public programs, for example, using community workers (e.g. Attanasio et al., 2014; Andrew et al., 2020). The approach of this study is to empower parents to improve nutrition and stimulation of their children through a very short and focused program, which is therefore suitable in contexts where public infrastructure is not sufficiently functional. In the context of this study, for example, spot-checks of existing mother-child centers (Anganwadi centers) revealed that the centers are often overburdened and unreliable.[2]

The study is designed as a randomized controlled trial, in which we assigned 1,480 households with 10- to 20-months old children to one of four 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 trainings, and one no-intervention (control) arm. The study location is the rural district of Madhepura in the northeast Indian state of Bihar. Chronic malnutrition and anemia are common in Madhepura and stimulating public child care services for under-three-year olds do not exist (IIPS, 2017).

One year after implementation, we find no average impact of the Lucky Iron Leaf alone, the dialogic reading alone or the combination of both interventions on cognitive, language, motor or socioemotional functions. 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 on 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).

[2]During 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.

Due to randomization at the individual level, we carefully investigate the potential for treatment spillovers. Following Baird et al. (2016), we model spillovers in form of saturation rates in neighboring households of opposite treatment types. The pure intention-to-treat effect of each treatment arm 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. From a policy perspective, the detected positive treatment spillovers are relevant and should be taken into account in cost-effectiveness considerations of early childhood programs that intervene at the household level.

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. Section VI. presents intervention take-up rates, intention-to-treat effects under consideration of treatment spillovers, and complier average causal effects among households with no intervention neighbors. Section VII. briefly discusses potential intervention mediators. Section VIII. concludes.

II. THE EXPERIMENT

II.A. Background

The study is located in the district of Madhepura in Bihar, India. Bihar is the third largest state of India but has the lowest GDP per capita. With a population of 104 million about 88 percent of its population 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 (IIPS, 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. 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 (IIPS and ICF, 2017).

No public programs for psychosocial stimulation of under three-year olds exist in rural Bihar. The day care of the Integrated Child Development Services (ICDS) in so-called 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 emphasizes 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. Dialogic reading is solely based on having a conversation about images in the book and does not require literacy. Picture book reading is not practiced at all in the study population and the increase in reading frequency at the extensive margin is one of the basic aims of the dialogic reading intervention.

The effectiveness of dialogic reading in comparison to regular unspecified book sharing 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). They find gains in receptive and expressive language, child attention and socioemotional behavior immediately after the completion of a four to eight week program with weekly meetings. In addition to children, caregivers were also tested in South Africa and improved in book sharing interactions. This result is promising with respect to the sustainability of the gain in children's development and with respect to potential spillovers on other domains of caregiving. 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 content of the dialogic reading intervention 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 by Cooper et al. (2014), Vally et al. (2015), and Murray et al. (2016). During four home-visits female facilitators, trained by Mikhulu, explained and practiced methods of dialogic reading with the primary caregiver and the child.

The first two sessions were delivered immediately after baseline data collections and one week

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

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

apart from each other. Sessions three and four were delivered three months later and one week apart from each other. In total, three picture books were left with the families to practice. The intervention did not include any monitoring or encouragement through further home visits (see appendix VIII. for further intervention details).

II.C. The Lucky Iron Leaf™

The second intervention consists of the distribution of the Lucky Iron Leaf. 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 in children and in a large-scale intention-to-treat design.

The Lucky Iron Leaf is the Indian alternative to the Lucky Iron Fish™, a fish shaped iron ingot, which effectively reduced anemia among Cambodian women (Charles et al., 2011, 2015).[6] 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 Fish™ implementation schedule from the trials in Cambodia (see appendix VIII. 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 Guelp showed that the Lucky Iron Leaf enriches one serving of the typical lentils-based dish *daal* by 4.25-5 mg of bioavailable iron. For children who eat three times a day half of an adult-equivalent portion this amounts to 6-7 mg per day of supplementary iron. 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).

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

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

III. EXPERIMENTAL DESIGN AND DATA

III.A. Sampling and randomization

The study population is recruited from a listing of 2,000 households with pregnant women in 2015 who had children of ages between 10 to 20 months at the time of this study in 2016. The listing resulted from a random selection of six of Madhepura's thirteen blocks (sub-districts) and, in these six blocks, 68 from a total of 95 gram panchayats were sampled, comprising 180 villages. At village level, information on households with pregnant women were gathered from rural childcare center (Anganwadi center) registers.[7] 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 2,000 households to one of four treatment groups: (i) a dialogic reading group, (ii) a Lucky Iron Leaf group, (iii) a dialogic reading and Lucky Iron Leaf group, and (iv) a no intervention control group. The randomization was conducted prior to the baseline survey, as the treatment implementation started subsequent to the survey. Due to migration, non-response and child death, the listed 2,000 households reduced to 1,483 households at baseline.

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. We collected endline data exactly one year after the baseline survey in December 2017.

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 ELAN-R and SETK-2 tests (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

[7] In 2015/16, 76% of all pregnant women in Madhepura had registered their pregnancies (IIPS, 2017).

age in months trend, and a heteroskedastic residual variance.

We proxy iron deficiency by hemoglobin (Hb) levels and the presence of anemia.[8] Our anemia cut-offs follow the WHO recommendations, according to which 6-59 months old children are classified as anemic if Hb$<$11 g/dl (World Health Organization, 2011). We used HemoCue 301® machines to determine the Hb concentration from capillary blood samples at the point of care and informed all patients about the results directly after testing. In case of mild (9-10.9 g/dl) or moderate anemia (7-8.9 g/dl), we recommended caregivers to seek treatment at the local primary health center or with the community health worker and to change the diet towards more iron rich foods. Severely anemic children (Hb$<$7 g/dl) were offered transport and cost coverage of treatment in the local hospital or health centers. We followed this action plan for ethical reasons. At baseline, about 70 percent of children in the sample were anemic and 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, Krämer, Kumar and Vollmer (2019) find that the behavior and health of participants did not change in response to recommendations and referrals. We additionally test for heterogeneous take-up by baseline anemia and do not find differential effects.

Sample characteristics and balance

Our estimation sample consists of 1,164 households at endline, which is 78.5% of the 1,483 baseline households.[9] The main reasons for attrition at endline are the absence of the household or the child at the time of data collection (N=152) and incomplete child development tests (N=83). Ten households refused their participation. Additional but less common causes of attrition are migration of households across village borders, incorrect household identification information and child death. To check for selective attrition from baseline to endline, we regress an indicator for attrition on the three treatment group indicators. All coefficients are below two percentage points and none of them approaches statistical significance (appendix Table A.II.1, column 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).

Table I presents means of baseline background characteristics and outcomes in our estimation sample of all four experimental groups. Most participants are Hindus and members of the caste category "other backward class". The highest education level in most households and among mothers is uncompleted primary school. To evaluate the wealth status of households, we created an asset index and a housing index, which we divided into 10 asset and housing quintiles, respectively.

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

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

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

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

Note: Table continues on next page. Std. Diff. refers to the standardized difference in means of the control group and the respective treatment group. p-values refer to a t-test of the equality of means of the control group and the respective treatment group. [a] 10 quintiles based on a durable asset index generated by factor analysis. [b] 10 quintiles based on a housing quality index generated by factor analysis. [c] Indicator equals one if mother is allowed to go alone to 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: Std. Diff. refers to the standardized difference in means of the control group and the respective treatment group. p-values refer to a t-test of the equality of means of the control group and the respective treatment group. [d]Sum of stimulating activities conducted with the child in the past 3 days. [e]Equals 1 if caregiver explains wrong behavior to child, takes away priviliges or gives child something else to do. [f]Equals 1 if mother shouts, yells or screams at child or spanks, hits, kicks or slaps child.

About 90 percent of mothers were breastfeeding their child 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 families 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 (combined intervention: standardized difference=0.17, p-value=0.05; dialogic reading: standardized difference=0.18, p-value=0.04).

IV. ESTIMATION STRATEGY

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 treatment group, respectively. X_i is a vector of imbalanced baseline covariates and includes mothers' education, reading abilites and empowerment, children's sex, and subdistrict fixed effects. In estimations on hemoglobin levels and anemia, we additionally control for measurement device fixed effects. The β-coefficients represent the intention-to-treat effects of the respective intervention, disregarding non-compliance. 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 36 hypotheses, following from six outcomes tested in six subgroups. The six outcomes refer to iron deficiency proxied by Hb and anemia and the five development scales. The six subgroups refer to the main effect estimation and five heterogeneous treatment effect estimations by maternal education, maternal empowerment, children's sex, age and anemia status at baseline. We correct for multiple hypotheses testing using Bonferroni adjustment and taking into account an average correlation across outcomes of 0.31 (Sankoh, Huque and Dubey, 1997; Aker et al., 2012).[10] The multiple hypotheses testing corrected equivalent 10 percent significance level is 0.0084 and the equivalent 5 percent significance level is 0.0042. We will use these p-values to

[10]The minimum correlation of outcome variables is 0.07 between hemoglobin and cognitive or motor skills. The maximum correlation is 0.64 between cognitive and motor skills. The median correlation is 0.30, the 25th percentile correlation is 0.12, and the 75th percentile correlation is 0.41.

mark corrected significance levels in regression Tables. Given a sample size of 1,164, roughly 290 observations per treatment arm, and statistical power of 80 percent, we can detect effect sizes of 0.22 standard deviations in child development outcomes, 0.35 g/dl in Hb levels, and 12 percentage points in anemia prevalence.

V. RESULTS

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

Table II presents the intention-to-treat estimation results on cognitive development, receptive language, expressive language, motor skills, and socioemotional development. The different columns refer to a simple regression of the outcome on the three treatment group indicators (model 1), a covariate adjusted model (model 2) and a covariate and baseline outcome adjusted model (model 3). The intention-to-treat effects are small and 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 analysis, 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 tester 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. The results confirm the findings of Table II.

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

The causal impact 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 the 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

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

	Model 1	Model 2	Model 3
Cognitive			
DR	0.028	0.106	0.082
	(0.087)	(0.085)	(0.089)
LIL	-0.067	-0.001	-0.016
	(0.088)	(0.085)	(0.089)
DR & LIL	0.037	0.064	0.083
	(0.089)	(0.086)	(0.089)
Observations	1146	1136	1013
Adjusted R^2	-0.001	0.082	0.087
Receptive language			
DR	-0.046	-0.009	-0.027
	(0.085)	(0.085)	(0.090)
LIL	-0.027	0.017	-0.022
	(0.085)	(0.085)	(0.089)
DR & LIL	-0.008	0.018	0.000
	(0.086)	(0.086)	(0.090)
Observations	1159	1148	1038
Adjusted R^2	-0.002	0.029	0.030
Expressive language			
DR	-0.014	0.035	0.040
	(0.089)	(0.088)	(0.092)
LIL	-0.039	0.040	0.031
	(0.089)	(0.088)	(0.092)
DR & LIL	0.020	0.036	0.056
	(0.090)	(0.089)	(0.092)
Observations	1159	1148	1041
Adjusted R^2	-0.002	0.049	0.077
Motor			
DR	-0.013	0.051	0.005
	(0.085)	(0.083)	(0.087)
LIL	-0.022	0.021	-0.026
	(0.085)	(0.084)	(0.087)
DR & LIL	0.029	0.059	0.015
	(0.086)	(0.084)	(0.087)
Observations	1123	1113	994
Adjusted R^2	-0.002	0.068	0.082
Sociomeotional			
DR	0.009	0.061	0.093
	(0.081)	(0.080)	(0.086)
LIL	-0.022	0.041	0.085
	(0.081)	(0.080)	(0.086)
DR & LIL	-0.017	0.012	0.019
	(0.082)	(0.081)	(0.086)
Observations	1151	1140	1017
Adjusted R^2	-0.002	0.039	0.044
Controls		✓	✓
Baseline outcome			✓

Note: Control variables include maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Baseline outcome controls for the baseline value of the respective model's outcome. Standard errors in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01. Multiple testing corrected significance levels: + p<0.1, ++ p<0.05, +++ p<0.01.

Table III: Intention-to-treat effects on iron deficiency

	Model 1	Model 2	Model 3
Hemoglobin			
DR	-0.152	-0.177	-0.121
	(0.119)	(0.119)	(0.125)
LIL	0.003	-0.010	-0.132
	(0.120)	(0.120)	(0.127)
DR & LIL	0.025	-0.012	-0.158
	(0.119)	(0.120)	(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	0.073*	0.112**
	(0.043)	(0.043)	(0.048)
LIL	0.008	0.016	0.064
	(0.044)	(0.044)	(0.049)
DR & LIL	-0.018	-0.004	0.045
	(0.043)	(0.044)	(0.048)
Observations	1048	1039	710
Adjusted R^2	0.000	0.025	0.176
Control mean	0.56	0.56	0.54
Moderate anemia			
DR	0.035	0.035	0.030
	(0.039)	(0.040)	(0.046)
LIL	0.006	0.003	0.044
	(0.040)	(0.040)	(0.047)
DR & LIL	-0.009	-0.001	0.054
	(0.040)	(0.040)	(0.046)
Observations	1048	1039	710
Adjusted R^2	-0.001	0.013	0.111
Control mean	0.28	0.28	0.26
Controls		✓	✓
Device fixed effects		✓	✓
Baseline outcome			✓

Note: Control variables include maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Device fixed effects are hemoglobin measurement device fixed effects. Baseline outcome controls for the baseline value of the respective model's outcome. Standard errors are in parentheses. Conventional significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Multiple testing corrected significance levels: $_{+}$ $p<0.0084$, $_{++}$ $p<0.0042$, $_{+++}$ $p<0.0008$.

expected. This is a result of sample selection caused by the reduction in sample size (see appendix Table A.III.2 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.

A failure to reduce iron deficiency implies early interruptions 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.

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.2). 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 effects on child development were compensated for by positive effects of sharing books.

V.C. Heterogeneous treatment effect analysis

The effectiveness of the interventions potentially differs by participants' characteristics. We assess heterogeneous treatment effects by mothers' education and empowerment status, and children's sex, age, and anemia status at baseline. The rationale for estimating heterogeneous treatment effects by these characteristics is as follows:

1. *Maternal education and empowerment* are likely to matter for intervention quality. Educated mothers are likely to follow the training and internalize the procedures of both interventions with more ease and perform higher quality book sharing. Similarly, empowerment matters for women to attend the implementing trainings more confidently and engaged.

2. *Children's sex* can affect the caregivers' motivation to improve the health and development of their child when sons are preferentially treated in comparison to girls, which is common in India and our study location (e.g. Barcellos, Carvalho and Lleras-Muney, 2014; Jayachandran and Pande, 2017; Ebert and Vollmer, 2019).

3. *Children's age* might correlate with the quality of book sharing, because high quality dialogic reading is performed more easily with older children. Additionally, older children are more likely to be weaned and hence to consume the fortified food more often.

4. *Children's anemia status* matters for active participation in sharing books. Anemic children tend to be tired and less inquiring. Therefore, non-anemic children can gain more from dialogic reading. On the other hand, children's anemia status can evoke caregivers' motivation to learn about and utilize the Lucky Iron Leaf; i.e., caregivers, who learned about their child being anemic during our baseline testing, are potentially more attentive during the intervention implementation.

The results by mothers' baseline characteristics are mixed (see appendix Table A.IV.1). Children of educated mothers in the Lucky Iron Leaf group gain 0.51 standard deviations in receptive language and 0.40 standard deviations in cognitive skills. However, children of educated mothers in the Lucky Iron Leaf group also have lower hemoglobin levels at endline. Similarly by mothers' empowerment, children of empowered women gain 0.27 standard deviations in socioemotional development from the dialogic reading and Lucky Iron Leaf interventions, whereas children of mothers with low empowerment gain in cognitive skills from dialogic reading (0.18 SD). However, none of the results by mothers' characteristics are significant after multiple hypotheses testing.

The heterogeneity analysis by children's characteristics shows suggestive evidence for differential treatment effects by children's sex and age (see appendix Table A.IV.2), and strong evidence for heterogeneous treatment effects by children's baseline anemia status (Table IV). Opposed to hypothesis 2, girls in the dialogic reading and Lucky Iron Leaf groups improve in motor development (0.28 SD and 0.35 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 can then lead to a larger increase in motor skills among girls than boys. Further, supporting hypothesis 3, children of the older age group (15-20 months at baseline) show improvements in motor skills (0.39 SD) from the dialogic reading stand-alone intervention and in cognitive and receptive language skills (0.41 SD and 0.38 SD, respectively) from the combined intervention. While the results by children's sex and age are suggestive, the effects do not remain significant after multiple hypotheses testing.

Table IV presents heterogeneous treatment effects by children's baseline anemia status. Non-anemic children, who make up 30 percent of the sample, gain 0.54 standard deviations in receptive language skills from the combined intervention, whereas anemic children do not benefit at all. 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 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.

Table IV: Heterogeneous intention-to-treat effects on hemoglobin and child development

	Hemoglobin	Any anemia	Cognitive	Receptive	Expressive	Motor	Socio-emotional
DR	-0.292	0.149*	0.143	0.421**	0.077	0.107	0.168
	(0.240)	(0.088)	(0.193)	(0.195)	(0.204)	(0.189)	(0.187)
DR x Anemic	0.211	-0.052	0.012	-0.388*	-0.075	0.017	-0.154
	(0.284)	(0.105)	(0.227)	(0.230)	(0.241)	(0.223)	(0.220)
LIL	0.105	-0.012	0.300	0.330*	0.208	0.148	0.074
	(0.244)	(0.090)	(0.189)	(0.191)	(0.200)	(0.185)	(0.182)
LIL x Anemic	-0.344	0.107	-0.348	-0.417*	-0.203	-0.028	0.025
	(0.290)	(0.107)	(0.226)	(0.228)	(0.240)	(0.223)	(0.218)
DR&LIL	0.122	-0.032	0.451**	0.538‡‡**	0.049	0.144	0.213
	(0.233)	(0.086)	(0.184)	(0.185)	(0.195)	(0.181)	(0.177)
DR&LIL x Anemic	-0.433	0.112	-0.427*	-0.596‡**	-0.040	-0.075	-0.287
	(0.281)	(0.104)	(0.222)	(0.223)	(0.235)	(0.218)	(0.214)
Anemic	-1.052‡‡‡*	0.357‡‡‡*	0.237	0.326**	-0.150	-0.059	-0.048
	(0.203)	(0.075)	(0.161)	(0.163)	(0.171)	(0.158)	(0.156)
Controls	✓	✓	✓	✓	✓	✓	✓
Device fixed effects	✓	✓					
Observations	710	710	758	766	766	742	760
Adjusted R^2	0.20	0.18	0.08	0.03	0.04	0.05	0.04
p-value (DR+interaction=0)	0.60	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

Note: Control variables include maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Device fixed effects are hemoglobin meaurement device 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<0.0084, $_{++}$ p<0.0042, $_{+++}$ p<0.0008.

We do not observe the same heterogeneity in treatment effects on hemoglobin by children's baseline anemia status. 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. 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.

The heterogeneous treatment effects by baseline anemia status imply a cross-productivity of baseline health and intervention benefit. Yet, the anemia status may correlate with a number of other characteristics which could confound the heterogeneous treatment effect by anemia. We test this hypotheses in two ways. First, 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. In fact, non-anemic children in the stand-alone and combined dialogic reading group perform significantly worse than non-anemic control group children. There are no significant imbalances in the Lucky Iron Leaf group or among anemic children. Second, 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, (6) and baseline anthropometric measures, while holding constant our regular set of covariates which we condition on in Table IV already. 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 the 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 treatment effects by anemia are relevantly driven by other factors.

The identified heterogeneous treatment effect is in line with Tofail et al. (2013)'s finding of improved mental development (their index includes cognitive and language items) from a nine months weekly stimulation program for non-iron deficient non-anemic children only. Tofail et al. (2013) recognize the endogeneity concern and argue that similar improvements in home stimula-

tion in that population support the relevance of children's iron deficiency and anemia status for the intervention's effectiveness.

Given that the combined intervention only comprised five home visits, the effect size on receptive language for non-anemic children is very large. Intensive 18 to 24 months weekly stimulation interventions increased receptive language by 0.22 standard deviations in Colombia and mental development (cognitive skills and language) by 0.27 standard deviations in Bangladesh. In Colombia, the combined effect of stimulation with biweekly micronutrient supplement replenishments was insignificantly lower than the single intervention effect size (Attanasio et al., 2014). Assuming the Colombian iron deficiency anemia prevalence of 19 percent and the Indian coefficients, the average effect is 0.44 standard deviations $(0.81 * 0.54 + 0.19 * 0)$.[11] This hypothetical effect size is higher than the actual average effect on receptive language found in Colombia. While it is futile to compare the Colombian to the Indian context, the exercise demonstrates the magnitude of the identified effect given the simplicity of the interventions. Yet, we only observe improvements in receptive language, while the comprehensive stimulation program in Columbia also improves cognition (0.26 SD) and the Bangladesh program improves behavior (0.18 SD in vocalization to 0.45 SD in response to examiner).

VI. Compliance and spillovers

VI.A. Take-up 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 eligible for the dialogic reading training participated in at least one training session (93% in the dialogic reading group and 95% in the combined group). 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 reported to have shared books in the week prior to endline data collections and were able to present the book at endline (19% in the dialogic reading group and 23% in the combined group). 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

[11] Attanasio et al. (2014) report an iron deficiency anemia prevalence of 19 percent for Colombia. The Indian coefficients are 0.54 for non-anemic children and 0 for anemic children.

Table V: Intervention take-up rates

	Dialogic reading				DR & LIL			
	Eligible		≥1 training		Eligible		≥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			
	Eligible		Received		Eligible		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			
					Eligible		Received	
					Mean	N	Mean	N
Panel C: DR and LIL								
≥1 training and LIL received					0.89	367	1.00	326
≥1 book and LIL present at endline					0.18	330	0.19	300
BS & LIL ≥4 per week and child eats food					0.14	326	0.16	293

throughout the year.

Panel B of Table V presents take-up rates of the Lucky Iron Leaf intervention. Treatment delivery rates are 95 and 97 percent in the stand-alone and combined intervention groups, respectively. Utilization of the Lucky Iron Leaf was substantially lower than dialogic reading utilization. One third of households cooked with the Lucky Iron Leaf at least four times per week in the past year; and in only a quarter of households, 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 were able to present the Lucky Iron Leaf at endline. These numbers suggest that the intervention was not successful in initiating a continual utilization of the Lucky Iron Leaf.

Panel C presents rates of compliance with both interventions simultaneously in the combined treatment group, of which 89 percent received both interventions. In line with panels A and B, take-up rates lie somewhat below those of the Lucky Iron Leaf. Eighteen percent of households are able to present the Lucky Iron Leaf and at least one book at endline. Fourteen percent utilize both interventions at least four times per week in the past 12 months and the child also consumes

the fortified food.

We do not find systematic differences in compliance by socioeconomic characteristics or children's anemia status. Therefore, the identified heterogeneous treatment effects by children's anemia status are not attributable to different utilization frequencies, but rather due to different utilization efficacies.

There are two major barriers to compliance with the dialogic reading intervention: having no time for sharing books (44%) and perceiving the child as too young (32%). The two main reasons for non-compliance with the Lucky Iron Leaf are 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%). Indeed, only 23.7 percent of households reported consumption of tomato or lemon in the seven days prior to the endline survey. 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%).

VI.B. Spillovers

Because treatments were individually assigned, about 38 percent of treatment households have at least one control household in 100 meters distance. Spillovers could hence possibly challenge the internal validity of our experiment. In 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 reading 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 effects arise when control group households attempt to compensate for the lack of treatment. Both 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.[12]

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

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

spillover effects the following way:

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

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}, the z-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.

$\beta_{1,T}$ presents the treatment effect of intervention T for households without a nearby study neighbor and net of spillover effects. $\beta_{2,T}^L$ presents the spillover effect received by control households from intervention L. $\beta_{3,T}^L$ presents the additional spillover effect received by treatment arm T households from intervention L. If $\beta_{3,T}^L$ 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 VI 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 dialogic reading 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 suggest that different mechanisms may be at work causing dialogic reading 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 percent) 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

Table VI: 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**	0.103				
	(0.134)	(0.082)				
LIL			0.220	0.047		
			(0.138)	(0.085)		
DR&LIL					0.204	0.074
					(0.143)	(0.087)
DR saturation	0.470***		0.447**		0.490***	
	(0.168)		(0.174)		(0.176)	
DR sat. x DR	-0.597**					
	(0.245)					
DR sat. x LIL			-0.537**			
			(0.262)			
DR sat. x DR&LIL					-0.721***	
					(0.263)	
LIL saturation	-0.213		-0.236		-0.187	
	(0.175)		(0.181)		(0.184)	
LIL sat. x DR	0.138					
	(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

Note: Control variables include maternal education, reading ability and empowerment, the child's sex, population density 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<0.0084, $_{++}$ p<0.0042, $_{+++}$ p<0.0008.

played an important role. In line with Table VI, descriptive evidence for spillovers from the Lucky Iron Leaf intervention seems not discernable; only 2% of control households report to have heard about the cooking device at endline.

VI.C. 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 parents of the control group start compensating the lack of treatment through other stimulating or health promoting activities, then the instrument, assignment to experimental groups, 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.

For dialogic reading compliance and Lucky Iron Leaf compliance, we estimate two first stages:

$$DRcomp_i = \omega_0 + \delta_1 DRany_i + \delta_2 LILany_i + X_i^{'}\lambda + \nu_i, \tag{3}$$

$$LILcomp_i = \mu_0 + \eta_1 DRany_i + \eta_2 LILany_i + X_i^{'}\kappa + \upsilon_i, \tag{4}$$

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

The second stage is:

$$y_i = \theta_0 + \pi_1 \widehat{DRcomp_i} + \pi_2 \widehat{LILcomp_i} + X_i^{'}\rho + \varepsilon_i, \tag{5}$$

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

In Table VII, 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. In line with the theory, 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 a large effect on hemoglobin levels (1.0 g/dl) and anemia (40 percentage points) and, following this, on expressive language and socioemotional skills. Albeit consistent, these results should be taken with a grain of salt, because only 47 participants classify as Lucky Iron Leaf compliers and 125 participants as dialogic reading compliers. When we define compliance by the observed presence of an intervention book or Lucky Iron Leaf in homes at endline, we count 77 Lucky Iron Leaf compliers and 90 dialogic reading compliers. Further, compliance measured by the observed presence of a book or the iron ingot avoids recall and desirability bias from self-reports. Reassuringly, employing this alternative compliance measure leads to very similar results.

Table VII: Complier average causal effects

	Cognitive	Receptive	Expressive	Motor	Socio-emotional
DR ≥4 times per week	0.344*	0.184	0.230	0.059	0.227
	(0.176)	(0.181)	(0.189)	(0.175)	(0.177)
LIL ≥4 times per week	0.002	0.543	0.803*	0.525	0.729*
	(0.440)	(0.445)	(0.474)	(0.438)	(0.437)
Controls	✓	✓	✓	✓	✓
Cragg-Donald Wald F statistic	28	28	28	27	28
Observations	427	431	433	420	429
	Hemoglobin	Any anemia			
DR ≥4 times per week	0.265	-0.052			
	(0.242)	(0.090)			
LIL ≥4 times per week	1.010	-0.391*			
	(0.634)	(0.235)			
Controls	✓	✓			
Device fixed effects	✓	✓			
Cragg-Donald Wald F statistic	25	25			
Observations	423	423			

Note: 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. Control variables include the standard set of maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Device fixed effects are hemoglobin meaurement device fixed effects. Standard errors are in parentheses. Conventional significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Multiple testing corrected significance levels: $_{+}$ $p<0.0084$, $_{++}$ $p<0.0042$, $_{+++}$ $p<0.0008$.

VII. Analysis of intervention mechanisms

VII.A. Anemia knowledge and dietary substitution

The success of the interventions hinges on changing mediators in the causal chain from intervention roll-out to improvements in early childhood development. An explicit mediator of the Lucky Iron Leaf intervention is iron deficiency, tested in section V.B.. Knowledge about anemia constitutes the key source of motivation for intervention uptake, but 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). Either people do not know the term "anemia" but know about iron deficiency, they forgot about anemia due to the disuse of the Lucky Iron Leaf, or the intervention was not successful in conveying anemia knowledge and therefore the need for increased iron intake. The latter two explanations are in line with low take-up rates and imply the intervention's ineffectiveness in inducing the required behavioral change.

We further explore offsetting substitution effects away from food diversity, vegetables, meat,

iron or vitamin A supplements as alternative explanations and find no evidence for the substitution hypothesis (appendix Table A.V.2).

VII.B. Impact on other caregiving domains

The dialogig reading training may affect caregiving beyond picture book reading. We asked all caregivers - independent of treatment assignment - about activities they performed with their children in the two weeks prior to the endline survey.[13] We enquired about book reading and seven other activities, such as playing, singing songs, going outside the compount etc., and summed these other activities into an activity index. 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. At last, the dialogic reading intervention may have affected caregivers aspirations with respect to their children's educational achievement, because the intervention trainings implied potential gains from book sharing in children's performance in school later on.

Table VIII: Mediators of the dialogic reading intervention

	(1) Read books	(2) Read books & has book	(3) No. of Activities	(4) Shout or spank	(5) Aspiration
Pure DR	0.168***	0.392‡‡‡	0.034	0.029	0.012
	(0.065)	(0.051)	(0.175)	(0.056)	(0.244)
Observations	629	629	616	636	502
Adjusted R^2	0.04	0.23	0.04	0.00	0.15
Pure LIL	0.005	0.032	0.007	0.059	-0.230
	(0.064)	(0.031)	(0.172)	(0.053)	(0.242)
Observations	619	619	611	631	528
Adjusted R^2	0.02	0.03	0.03	0.01	0.12
Pure DR & LIL	0.157**	0.399‡‡‡	-0.034	-0.045	-0.176
	(0.066)	(0.052)	(0.173)	(0.055)	(0.240)
Observations	610	610	598	625	505
Adjusted R^2	0.02	0.20	0.02	0.01	0.16
Controls	✓	✓	✓	✓	✓
Control mean	0.46	0.07	5.60	0.76	5.34

Note: Column (3) outcome: summation of indicators for 7 different activities recently conducted with the child. Column (4) outcome: dummy for whether mother spanks or shouts at child. Column (5) outcome: dummy for whether Anganwadi center was visited in the pas 14 days. 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 include 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. Multiple testing corrected significance levels: $_{+}$ p<0.0084, $_{++}$ p<0.0042, $_{+++}$ p<0.0008.

[13] Partially these indicators are based on UNICEF's Multiple Indicator Cluster Survey.

To disentangle pure treatment effects from spillover effects, we use the spillover specification in equation (2), to explore the interventions' impacts on other domains of caregiving in Table VIII. Column 1 of Table VIII shows a significant increase in the probability of book reading by about 17 percentage points in both dialogic reading treatment groups. This relationship traces the first stage of the functional chain and is thus reassuring. However, because this result relies on self-reported book reading, we might confuse real increases with caregivers responding in the desired and anticipated way. In column 2, we re-code households as non-reading when they report to have no books at home. The effects increase dramatically in size to about 40 percentage points and support that desirability bias does not upward bias the finding in column 1. However the results on other caregiving domains are sobering. Columns 3 to 5 show that dialogic reading does not affect other activities conducted with the child, the caregivers shouting or spanking behavior, or educational aspirations.

VII.C. Quality of sharing books

Based on Figure 1, we evaluate the quality of book sharing, proxied by maternal knowledge of the trainings' and books' contents, as reasonable. Ninety-four percent of caregivers were able to freely name a specific content, character or page in her favorite book (top left). And, 64 percent of caregivers identify all intervention-book covers correctly from a selection of three intervention and three random book covers shown to them (top right).[14] Of the twelve dialogic reading concepts trained, only 5 percent of respondents cannot name any concept (bottom left).

At last, 85 percent of caregivers report that children take the initiative to share books (bottom right), which implies positive book sharing experiences and a decent quality of dialogic reading. Interestingly, non-anemic children are 5 percentage points more likely to have taken the initiative to share books than anemic children, suggesting that non-anemic children are more actively engaged in sharing books.

VIII. CONCLUSION

We study whether two simple and short-lived interventions can improve anemia and mental functions, specifically cognitive, language, motor, and psychosocial skills, 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 the Lucky Iron Leaf. The second intervention targets psychosocial stimulation as a risk factor of early childhood development and consists of four at home trainings of

[14] 91 percent of respondents identify no non-intervention book.

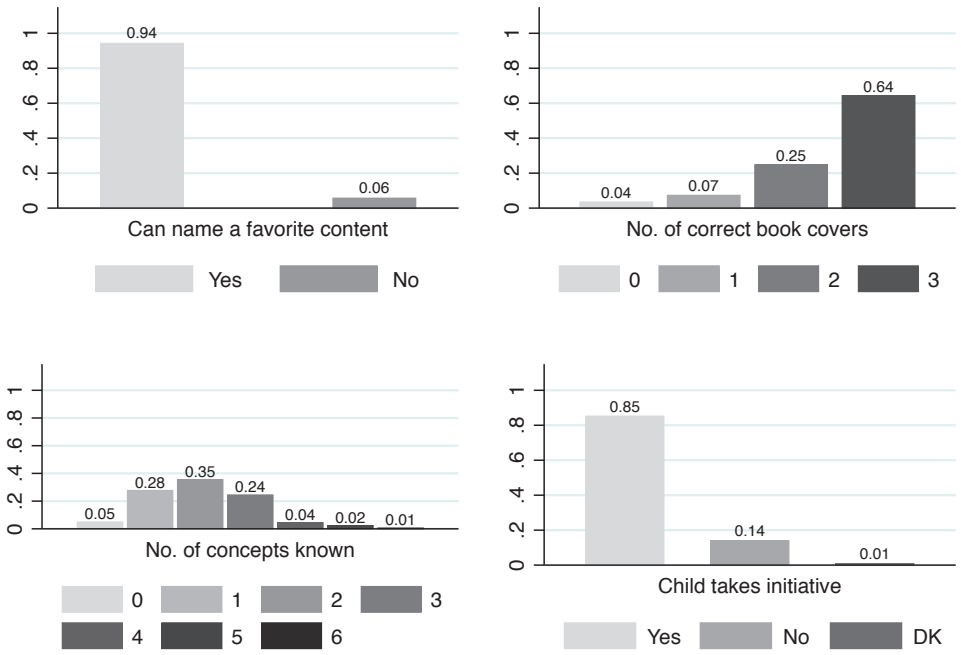


Figure 1: Quality of book sharing among regular users (shared books at least four times per week in the past year)

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. We randomly allocate children in the targeted age range to a no-intervention control group, a Lucky Iron Leaf treatment group, a dialogic reading treatment group, and a treatment group that receives both interventions.

One year after implementation, neither intervention nor their combination have improved anemia or mental functions of children. 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 and make 30 percent of the 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 heterogeneous treatment effects by anemia do not seem to be driven by other health or socioeconomic indicators. Biologically, non-anemic children likely have an advantage in the efficacy of book shar-

ing over their anemic peers – they can engage more actively, whereas anemic children tend to be tired and less exploratory. Given the simplicity of our interventions, the effect size on non-anemic children is considerable. In the Colombian context, where children's iron deficiency anemia prevalence is 19 percent, Attanasio et al. (2014) find improvements from an intensive weekly home-visit program in receptive language and cognition by 0.22 and 0.26 standard deviations, respectively.

The intention-to-treat results need to be evaluated against the backdrops of (i) low utilization of the Lucky Iron Leaf and (ii) potential spillovers arising from treatment assignment at the individual level. Utilization of the iron fortification device was reported to be at 25 percent initially and decreased to 3 percent a year later. In contrast, Charles et al. (2015) report daily utilization rates of 80 to 90 percent throughout 12 months in Cambodia; but, in their study compliance was closely monitored. In this study, common reasons for non-compliance were the dislike of the iron ingot among husbands and parents-in-law, and the required use of fruit-acid items in the fortification process, which is naturally integrated in Cambodian but not northeast Indian dishes. Our results imply that a single home visit is not sufficient to convince households of a new cooking utensil and to stimulate the required behavioral changes for the device's consistent usage.

The second consideration regards our choice of evaluation design. We randomly assigned 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. We find a positive saturation effect of dialogic reading on mental functions of almost half a standard deviation for control households that switch from no surrounding dialogic reading household to all surrounding households being dialogic reading ones. However, we do not find saturation effects among peer dialogic reading households or of the Lucky Iron Leaf intervention. Further, the pure intention-to-treat effects increase in magnitude to 0.20 to 0.27 standard deviations in all three treatment arms. From a policy perspective, positive treatment spillovers on neighboring untreated households are desirable and can improve the cost-effectiveness of the interventions. However, a different study design, randomizing communities into different levels of treatment intensities, would be needed to fully understand the benefits from treatment externalities.

In light of imperfect compliance and treatment spillovers, we estimate the interventions' impact for treatment compliers in a sub-sample of households with zero dialogic reading saturation and using a two-stage least squares estimation strategy (CACE). Albeit imprecisely measured, we find that compliance with the Lucky Iron Leaf reduces anemia and translates into sizeable improvements in expressive language functions and socioemotional behavior. Further, the dialogic reading effects increase moderately in size. We interpret these results as suggestive evidence for the efficacy of the interventions, but remain cautious due to the small and selective sample underlying this exercise.

A relevant concern lingering our results is the lack of effectiveness of the interventions for anemic children. It prompts to reconsider the suitability of the dialogic reading and Lucky Iron Leaf interventions as implemented here to cater the most disadvantaged children in rural Bihar. At the same time, these short-lived interventions can potentially be suitable where anemia rates are low, psychosocial stimulation is lacking and acid food items are commonly used. Replications of the study in somewhat better-off contexts will increase our understanding to which extent the program constitutes an adequate alternative to the comprehensive programs tested in Bangladesh, Colombia and Jamaica.

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.

Alderman, Harold, John Hoddinott, and Bill Kinsey. 2006. "Long term consequences of early childhood malnutrition." *Oxford Economic Papers*, 58(3): 450–474.

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.

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, 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. 2017. "Human Capital Development and Parental Investment in India." National Bureau of Economic Research Working Paper 21740.

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2015. "Estimating the Production Function for Human Capital: Results from a Randomized Control Trial in Colombia." National Bureau of Economic Research Working Paper 20965.

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

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

Barcellos, Silvia Helena, Leandro S Carvalho, and Adriana Lleras-Muney. 2014. "Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently?" *American Economic Journal: Applied Economics*, 6(1): 157–189.

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.

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.

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).

Ebert, Cara, and Sebastian Vollmer. 2019. "Early childhood development and heterogeneity in idiosyncratic son preferences for the next born child in India."

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.

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

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, 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.

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. 2012*a*. "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. 2012*b*. "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.

IIPS. 2017. "National Family Health Survey (NFHS-4),2015-16: District Fact Sheet Madhepura Bihar."

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

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

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.

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.

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 Papers.

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.

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.

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.

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.

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.

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.

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.

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.

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–3." *The Journal of nutrition*, 143(6): 885–893.

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. 2011*a*. "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. 2011*b*. "Inequality in early childhood: risk and protective factors for early child development." *The Lancet*, 378(9799): 1325–1338.

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.

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

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

APPENDIX

Appendix I - The interventions

The dialogic reading intervention

The training was delivered in Hindi or Maithili, the local language in Madhepura, by local female facilitators. 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).

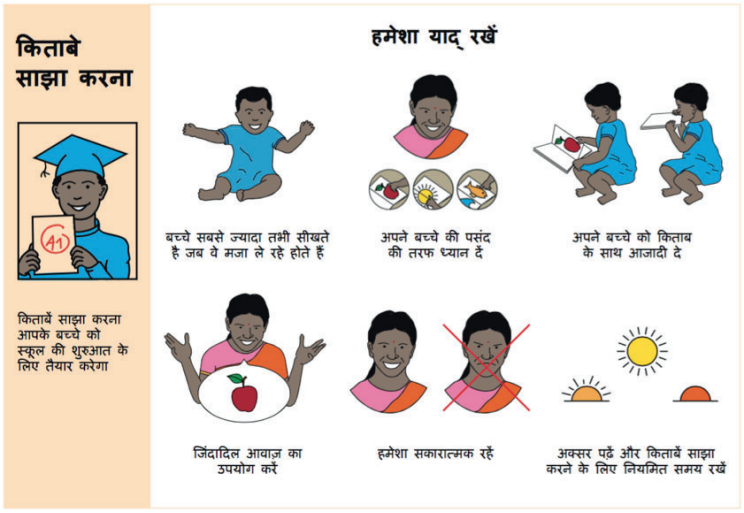
Table A.I.1: Topics of the dialogic reading training by session

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 we 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. Two of the books were obtained from the Pratham Books publisher, a sub-organization of the education focused Indian NGO Pratham, to ensure cultural appropriateness

and familiarity of the stories displayed. The books had a minimal amount of written words so the story could be followed by illiterate readers well. 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).

Home visits were conducted by local facilitators who were trained for one week by experienced personnel of the Mikhulu Trust. Each home visit lasted for about 40 to 60 minutes. Twenty-five to 45 minutes were spent on conveying the content using tablets for presentations, pictures and videos. Each session contained a review of previous sessions. Another 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 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 daily book sharing 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

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.

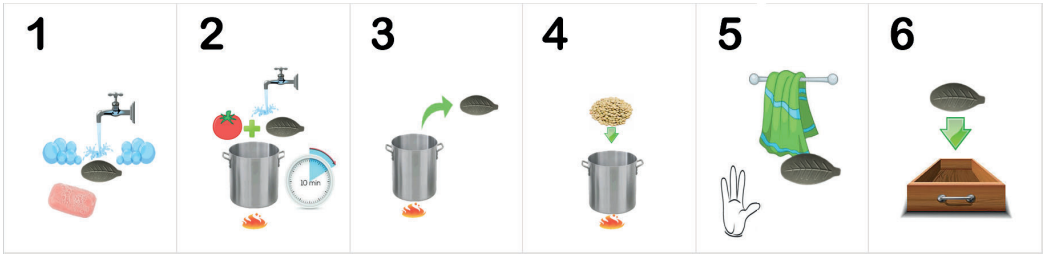


Figure A.I.2: Lucky Iron Leaf manual

Appendix II - Attrition and randomization

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

	Baseline - endline
DR	0.006
	(0.032)
LIL	0.019
	(0.032)
DR & LIL	0.018
	(0.031)
Observations	1480
Adjusted R^2	0.00
F statistic	0.17

Note: Standard errors in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01. Multiple testing corrected significance levels: $_{+}$ p<0.0084, $_{++}$ p<0.0042, $_{+++}$ p<0.0008.

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

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

Note: Table continues on next page. Std. Diff. refers to the standardized difference in means of the control group and the respective treatment group. p-values refer to a t-test of the equality of means of the control group and the respective treatment group. [a]10 quintiles based on a durable asset index generated by factor analysis. [b]10 quintiles based on a housing quality index generated by factor analysis. [c]Indicator equals one if mother is allowed to go alone to 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 A.II.2 continued

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

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

Appendix III - Additional intention-to-treat results

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

	Cognitive		Receptive language		Expressive language		Motor		Socioemotional	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
DR	0.097	0.106	–0.016	–0.009	0.043	0.039	0.056	0.052	0.083	0.055
	(0.078)	(0.085)	(0.081)	(0.085)	(0.085)	(0.088)	(0.078)	(0.082)	(0.071)	(0.080)
LIL	–0.010	0.022	0.012	0.029	0.059	0.060	0.007	0.038	0.054	0.044
	(0.079)	(0.085)	(0.081)	(0.085)	(0.085)	(0.089)	(0.079)	(0.082)	(0.071)	(0.081)
DR & LIL	0.049	0.078	0.006	0.017	0.045	0.046	0.056	0.059	0.023	0.005
	(0.079)	(0.086)	(0.082)	(0.087)	(0.086)	(0.090)	(0.079)	(0.083)	(0.072)	(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: Standard errors in parentheses. Conventional significance levels: * p<0.1, ** p<0.05, *** p<0.01. Multiple testing corrected significance levels: $_{+}$ p<0.0084, $_{++}$ p<0.0042, $_{+++}$ p<0.0008.

Table A.III.2: ITT effects on hemoglobin and anemia in a sample restricted to the ANCOVA sample

	Hemoglobin			Any anemia			Moderate anemia		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
DR	−0.077	−0.140	−0.121	0.080	0.111**	0.112**	0.026	0.036	0.030
	(0.144)	(0.143)	(0.125)	(0.052)	(0.052)	(0.048)	(0.048)	(0.049)	(0.046)
LIL	−0.118	−0.127	−0.132	0.060	0.060	0.064	0.050	0.051	0.044
	(0.147)	(0.145)	(0.127)	(0.053)	(0.053)	(0.049)	(0.049)	(0.049)	(0.047)
DR & LIL	−0.074	−0.136	−0.158	−0.000	0.033	0.045	0.036	0.045	0.054
	(0.143)	(0.143)	(0.125)	(0.052)	(0.052)	(0.048)	(0.048)	(0.048)	(0.046)
Device fixed effects		✓	✓		✓	✓		✓	✓
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
Control mean									

Note: Control variables include maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Device fixed effects are hemoglobin meaurement device fixed effects. Baseline outcome controls for the baseline value of the respective model's outcome. Standard errors are in parentheses. Conventional significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Multiple testing corrected significance levels: $_{+}$ $p<0.0084$, $_{++}$ $p<0.0042$, $_{+++}$ $p<0.0008$.

Appendix IV - Additional results on heterogeneous treatment effects

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

	Hemoglobin	Any anemia	Cognitive	Receptive	Expressive	Motor	Socio-emotional
Mother completed primary							
DR	−0.153	0.059	0.115	−0.063	0.052	0.067	0.084
	(0.137)	(0.050)	(0.097)	(0.097)	(0.101)	(0.095)	(0.092)
DR x Primary	−0.028	0.035	−0.036	0.218	−0.079	−0.037	−0.101
	(0.276)	(0.101)	(0.196)	(0.195)	(0.204)	(0.191)	(0.185)
LIL	0.099	−0.015	−0.065	−0.082	0.019	0.009	0.082
	(0.135)	(0.049)	(0.096)	(0.095)	(0.099)	(0.094)	(0.090)
LIL x Primary	−0.504*	0.137	0.401*	0.512**	0.169	0.146	−0.196
	(0.298)	(0.109)	(0.214)	(0.213)	(0.223)	(0.211)	(0.203)
DR&LIL	−0.017	0.004	0.075	−0.068	0.065	0.133	0.050
	(0.138)	(0.051)	(0.099)	(0.099)	(0.103)	(0.097)	(0.094)
DR&LIL x Primary	0.077	−0.049	−0.042	0.320*	−0.141	−0.252	−0.176
	(0.274)	(0.100)	(0.195)	(0.194)	(0.203)	(0.191)	(0.185)
Primary	0.167	0.002	0.325*	−0.007	0.371**	0.177	0.497***
	(0.234)	(0.086)	(0.168)	(0.168)	(0.176)	(0.165)	(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	0.096*	0.184*	0.075	−0.059	0.116	−0.053
	(0.156)	(0.057)	(0.112)	(0.111)	(0.116)	(0.110)	(0.105)
DR x Empowered	0.140	−0.041	−0.179	−0.207	0.217	−0.144	0.277*
	(0.242)	(0.088)	(0.172)	(0.172)	(0.178)	(0.168)	(0.162)
LIL	0.093	−0.030	0.008	0.096	0.036	−0.001	−0.061
	(0.150)	(0.055)	(0.108)	(0.107)	(0.111)	(0.106)	(0.101)
LIL x Empowered	−0.288	0.129	−0.024	−0.211	0.011	0.057	0.273*
	(0.248)	(0.091)	(0.175)	(0.175)	(0.182)	(0.171)	(0.165)
DR&LIL	0.073	−0.023	0.087	0.059	−0.004	0.076	−0.013
	(0.154)	(0.056)	(0.110)	(0.110)	(0.114)	(0.108)	(0.104)
DR&LIL x Empowered	−0.211	0.049	−0.062	−0.115	0.104	−0.043	0.076
	(0.246)	(0.090)	(0.177)	(0.176)	(0.183)	(0.173)	(0.167)
Empowered	0.092	−0.035	0.078	0.094	0.014	0.024	−0.104
	(0.175)	(0.064)	(0.125)	(0.125)	(0.130)	(0.122)	(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	✓	✓	✓	✓	✓	✓	✓
Device fixed effects	✓	✓					

Note: Control variables include maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Device fixed effects are hemoglobin measurement device fixed effects. Standard errors are in parentheses. Conventional significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Multiple testing corrected significance levels: $_{+}$ $p<0.0084$, $_{++}$ $p<0.0042$, $_{+++}$ $p<0.0008$.

Table A.IV.2: Heterogenous ITT effects by the child's sex and age group

	Hemoglobin	Any anemia	Cognitive	Receptive	Expressive	Motor	Socio-emotional
Sex of child							
DR	−0.012	0.062	0.108	−0.026	−0.024	0.275**	0.078
	(0.173)	(0.063)	(0.123)	(0.123)	(0.128)	(0.120)	(0.116)
DR x Boy	−0.308	0.020	0.001	0.038	0.112	−0.428***	−0.029
	(0.236)	(0.086)	(0.169)	(0.168)	(0.175)	(0.165)	(0.160)
LIL	0.069	−0.021	−0.034	−0.056	0.030	0.205*	−0.019
	(0.177)	(0.065)	(0.126)	(0.126)	(0.131)	(0.124)	(0.119)
LIL x Boy	−0.154	0.067	0.062	0.136	0.024	−0.347**	0.108
	(0.241)	(0.088)	(0.171)	(0.170)	(0.178)	(0.168)	(0.161)
DR&LIL	0.042	−0.007	0.022	−0.057	0.009	0.137	−0.037
	(0.169)	(0.062)	(0.120)	(0.120)	(0.125)	(0.118)	(0.114)
DR&LIL x Boy	−0.104	0.003	0.086	0.151	0.055	−0.154	0.099
	(0.238)	(0.087)	(0.172)	(0.171)	(0.178)	(0.168)	(0.162)
Boy	0.241	−0.030	−0.073	−0.089	−0.185	0.318***	−0.079
	(0.169)	(0.062)	(0.121)	(0.120)	(0.125)	(0.118)	(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
Age group of child							
DR	−0.421*	0.098	0.013	−0.137	−0.047	−0.210	−0.003
	(0.215)	(0.079)	(0.153)	(0.154)	(0.160)	(0.149)	(0.146)
DR x Older	0.379	−0.048	0.130	0.187	0.123	0.386**	0.094
	(0.256)	(0.094)	(0.183)	(0.183)	(0.191)	(0.179)	(0.174)
LIL	−0.016	0.003	−0.209	−0.107	0.103	−0.177	0.023
	(0.232)	(0.085)	(0.161)	(0.161)	(0.168)	(0.156)	(0.153)
LIL x Older	0.007	0.018	0.291	0.172	−0.090	0.276	0.023
	(0.271)	(0.099)	(0.189)	(0.189)	(0.197)	(0.184)	(0.179)
DR&LIL	0.090	−0.106	−0.209	−0.235	−0.025	−0.122	−0.016
	(0.217)	(0.079)	(0.155)	(0.156)	(0.161)	(0.151)	(0.148)
DR&LIL x Older	−0.146	0.144	0.405**	0.375**	0.090	0.265	0.039
	(0.260)	(0.095)	(0.186)	(0.187)	(0.194)	(0.182)	(0.177)
Older	0.106	−0.108	−0.104	−0.070	−0.034	−0.102	−0.028
	(0.191)	(0.070)	(0.136)	(0.136)	(0.142)	(0.133)	(0.129)
Observations	1039	1039	1136	1148	1148	1113	1140
Adjusted R^2	0.04	0.03	0.09	0.03	0.05	0.07	0.04
Controls	✓	✓	✓	✓	✓	✓	✓
Device fixed effects	✓	✓					

Note: Control variables include maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Device fixed effects are hemoglobin measurement device 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<0.0084, $_{++}$ p<0.0042, $_{+++}$ p<0.0008.

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

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Household characteristics:							
Household size	-0.003						-0.007
	(0.008)						(0.010)
Hindu	-0.017						0.039
	(0.053)						(0.065)
Scheduled caste	0.120*						0.100
	(0.071)						(0.091)
Scheduled tribe	0.065						0.137
	(0.108)						(0.138)
Other backward class	-0.003						0.004
	(0.065)						(0.083)
Primary	-0.071						-0.075
	(0.051)						(0.062)
Middle school	-0.041						-0.014
	(0.057)						(0.070)
High school or higher	0.018						0.051
	(0.060)						(0.074)
Asset index quintile	-0.001						0.001
	(0.007)						(0.008)
Housing index quintile	0.003						-0.000
	(0.007)						(0.008)
Mother characteristics:							
Age in years		-0.009**					-0.011**
		(0.004)					(0.006)
Worked in past 12 months		-0.082					-0.062
		(0.058)					(0.074)
Decides about child nutrition		0.054					0.069
		(0.036)					(0.043)
Currently breastfed			-0.020				-0.023
			(0.035)				(0.042)
Vit-A past 6 months			-0.010				-0.014
			(0.038)				(0.046)
Iron past 3 months			-0.018				-0.012
			(0.046)				(0.055)
Home environment:							
Moderate stimulation				-0.082			-0.015
				(0.102)			(0.128)
High stimulation				-0.127			-0.040
				(0.102)			(0.127)
Good educational measures				-0.091			-0.158**
				(0.060)			(0.075)
Bad educational measures				0.140*			0.073
				(0.077)			(0.104)
Baseline outcomes:							
Cognitive					0.009		0.011
					(0.021)		(0.024)
Receptive language					0.025		0.014
					(0.018)		(0.021)
Expressive language					-0.023		-0.025
					(0.020)		(0.024)
Motor					-0.011		0.007
					(0.022)		(0.026)
Socioemotional					-0.027		-0.027
					(0.019)		(0.023)
Child anthropometrics:							
Height (z-score)						-0.033**	-0.036**
						(0.014)	(0.018)
Weight-for-age (z-score)						0.000	-0.002
						(0.013)	(0.017)
Head circumference (z-score)						0.001	0.015
						(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	0.01
F statistic	1.56	1.92	1.37	1.51	1.39	1.85	1.08

Note: Control variables include maternal education, reading ability and empowerment, the child's sex and sub-district fixed effects. Standard errors in parentheses. Conventional significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Multiple testing corrected significance levels: $_{+}$ $p<0.0084$, $_{++}$ $p<0.0042$, $_{+++}$ $p<0.0008$.

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

	Model 1	Model 2 Restricted to model 3 sample	Model 3
DR	0.421**	0.431**	0.440**
	(0.195)	(0.197)	(0.197)
DR x Anemic	-0.388*	-0.369	-0.382
	(0.230)	(0.234)	(0.233)
LIL	0.330*	0.359*	0.366*
	(0.191)	(0.196)	(0.195)
LIL x Anemic	-0.417*	-0.441*	-0.458**
	(0.228)	(0.234)	(0.233)
DR&LIL	0.538‡‡**	0.524‡***	0.492***
	(0.185)	(0.189)	(0.189)
DR&LIL x Anemic	-0.596‡***	-0.556**	-0.505**
	(0.223)	(0.228)	(0.228)
Anemic	0.327**	0.315*	0.321*
	(0.163)	(0.164)	(0.164)
Mother's age (years)			-0.011
			(0.010)
Good educational measures			0.194
			(0.125)
Bad educational measures			-0.175
			(0.168)
Height (z-score)			0.051**
			(0.023)
Controls	✓	✓	✓
Observations	766	732	732
Adjusted R^2	0.03	0.03	0.04

Note: Control variables include 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<0.0084, $_{++}$ p<0.0042, $_{+++}$ p<0.0008.

Appendix V - Additional results on spillovers and mediators

Table A.V.1: Spillover statistics

	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
	Spill-overs from the Lucky Iron Leaf intervention					
	Lucky Iron Leaf households		Control households	Dialogic reading households		
	LIL lent to s.o. else	Use LIL $\geq$4 times per week & Neighbor's children eat regularly in HH	Heard about LIL		Heard about LIL	
No	0.88	0.92	0.93		0.92	
Yes	0.06	0.06	0.02		0.02	
N/A	0.06	0.02	0.05		0.05	
Observations	677	665	329		353	

Note: Columns 1 to 2 refer to spillovers *sent* from the respective treatment group households, columns 3 to 4 refer to spillovers *received* by pure control group households and columns 5 to 6 refer to spillovers *received* by the opposite treatment group 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
DR	0.006	0.105	−0.002	0.015	−0.001	0.042
	(0.020)	(0.131)	(0.023)	(0.039)	(0.032)	(0.039)
LIL	0.021	−0.073	−0.019	−0.059	0.007	0.059
	(0.020)	(0.117)	(0.023)	(0.040)	(0.036)	(0.042)
DR & LIL	0.028	0.001	−0.021	−0.015	−0.011	−0.001
	(0.022)	(0.125)	(0.027)	(0.043)	(0.034)	(0.037)
Controls	✓	✓	✓	✓	✓	✓
Observations	1096	1133	1137	1136	1116	1096
Adjusted R^2	0.11	0.01	0.05	0.00	0.01	0.03
Control mean	0.04	7.10	0.91	0.43	0.20	0.34

Note: Control variables include maternal education, reading ability and empowerment, the child's sex and subdistrict fixed effects. Standard errors are in parentheses. Conventional significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Multiple testing corrected significance levels: $_{+}$ $p<0.0084$, $_{++}$ $p<0.0042$, $_{+++}$ $p<0.0008$.