Enhancing Young Children's Language Acquisition through Parent-Child Book-Sharing: A Randomized Trial in Rural Kenya

Heather A. Knauer, Pamela Jakiela, Owen Ozier, Frances Aboud, and Lia C.H. Fernald

Abstract

Worldwide, 250 million children under five (43 percent) are not meeting their developmental potential because they lack adequate nutrition and cognitive stimulation in early childhood. Several parent support programs have shown significant benefits for children's development, but the programs are often expensive and resource intensive. The objective of this study was to test several variants of a potentially scalable, cost-effective intervention to increase cognitive stimulation by parents and improve emergent literacy skills in children. The intervention was a modified dialogic reading training program that used culturally and linguistically appropriate books adapted for a low-literacy population. We used a cluster randomized controlled trial with four intervention arms and one control arm in a sample of caregivers (n = 357) and their 24- to 83-month-old children (n = 510) in rural Kenya. The first treatment group received storybooks, while the other treatment arms received storybooks paired with varying quantities of modified dialogic reading training for parents. Main effects of each arm of the trial were examined, and tests of heterogeneity were conducted to examine differential effects among children of illiterate vs. literate caregivers. Parent training paired with the provision of culturally appropriate children's books increased reading frequency and improved the quality of caregiver-child reading interactions among preschool-aged children. Treatments involving training improved storybook-specific expressive vocabulary. The children of illiterate caregivers benefited at least as much as the children of literate caregivers. For some outcomes, effects were comparable; for other outcomes, there were differentially larger effects for children of illiterate caregivers.

Keywords: dialogic reading, word gap, early childhood, local-language storybooks, primary school readiness

Center for Global Development
www.cgdev.org

Working Paper 502
February 2019

Enhancing Young Children's Language Acquisition through Parent-Child Book-Sharing: A Randomized Trial in Rural Kenya

Heather A. Knauer
School of Public Health, University of California, Berkeley,
Present address: School of Social Work, University of Michigan

Pamela Jakiela
Center for Global Development

Owen Ozier
Development Research Group, The World Bank
oozier@worldbank.org

Frances Aboud
Department of Psychology, McGill University

Lia C.H. Fernald
School of Public Health, University of California, Berkeley

We thank Sheyda Esnaashari, Saahil Karpe, Rohit Chhabra, Emily Cook-Lundgren, Marilyn N. Ahun, and Jennifer Nakata for research assistance throughout the development of this project. We thank the staff of Innovations for Poverty Action (Kenya); specifically, Patricia Gitonga, Michael Meda, Jessica Jomo, and the field team they led during data collection for this project. We thank Moran Publishers, E. A., for their cooperation and assistance in producing adapted storybooks. We are especially grateful to Patricia Kariger, who led instrument adaptation, to the KLPS team, who generously shared their previously adapted instruments with us, and to Peter J. Cooper, who shared intervention slides from a dialogic reading training in South Africa. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent.

This work was supported by The World Bank, Washington, DC (via three facilities: the Strategic Impact Evaluation Fund, the Early Learning Partnership, and the Research Support Budget); and by Echidna Giving, San Francisco, CA.

Heather A. Knauer, Pamela Jakiela, Owen Ozier, Frances Aboud, Lia C.H. Fernald, 2019. "Enhancing Young Children's Language Acquisition through Parent-Child Book-Sharing: A Randomized Trial in Rural Kenya." CGD Working Paper 502. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/enhancing-young-childrens-language-acquisition-through-parent-child-book-sharing

Center for Global Development
2055 L Street NW
Washington, DC 20036

202.416.4000
(f) 202.416.4050

www.cgdev.org

Conflict of interest statement

The authors have no conflicts of interest to declare.

Contents

1. Introduction

Poverty has significant and lasting consequences for children's cognitive and language development (Bradley & Corwyn, 2002; Brooks-Gunn & Duncan, 1997). In low- and middle-income countries (LMICs), 43 percent of children under age five do not reach their developmental potential (Black et al., 2017). Even in Kenya, one of the best-educated countries in sub-Saharan Africa, only 30 percent of third graders can read at the second-grade level (Piper, 2010), and 34 percent of pre-school children are "on track" for language and numeracy development (Kenya National Bureau of Statistics, 2013). The lack of adequate literacy preparation is a key risk factor for poor performance in primary school worldwide (Behrman et al., 2006).

One reason children living in poverty do not achieve their potential in terms of language development and emergent literacy is that they do not have access to linguistically stimulating home environments (Can & Ginsburg-Block, 2016). The availability of storybooks in the home and the quantity and quality of parents' reading engagements with their children are consistently associated with children's cognitive and language development, school readiness, and achievement (Raikes, et al, 2006; Park, 2008; Rodriguez et al., 2009; Zauche, Thul, Mahoney, & Stapel-Wax, 2016). In particular, exposure to storybooks has a direct, positive, causal impact on children's vocabulary and language skills (Hargrave & Sénéchal, 2000).

In spite of the importance of books and parental engagement, 97 percent of households in sub-Saharan Africa have two or fewer children's books, and only half of parents report having engaged in any cognitively stimulating activities with their young children in the last three days (United Nations Children's Fund, 2017). In some African contexts, responsive conversations between parents and their young children are actively discouraged by cultural norms (Weber, Fernald, & Diop, 2017; Jukes et al., 2018). Yet, in order to learn vocabulary and build linguistic skills, children need opportunities to formulate and express ideas, receive feedback, and engage with caregivers in responsive and reciprocal interactions (Wasik & Hindman, 2015).

1.1 Book-Sharing as a "Vocabulary Acquisition Device"

Programs that distribute children's storybooks and encourage parent-child book-sharing may reduce disparities in parental stimulation and emergent literacy skills in LMICs (Weisleder et al., 2017); these "book-sharing programs" have been so successful that they have been called a "vocabulary acquisition device" (Ninio, 1983). In higher-income countries, programs such as Reach Out and Read—which combine book distribution and parent education—improve literacy outcomes among vulnerable children (Weitzman, Roy, Walls, & Tomlin, 2004). Though a number of similar programs exist in LMICs (cf. Literacy Boost, Tusome, and PRIMR), most programs target children during their primary school years, not during the pre-literacy period (Dowd, Friedlander, Guajardo, Mann, & Pisani, 2013; Piper, Destefano, Kinyanjui, & Ong'ele, 2018; Piper, Jepkemei, & Kibukho, 2015).

In many LMICs, young children learn one or more mother tongues at home but use other languages in school (Clegg & Simpson, 2016). Even when a government's official policy mandates initial literacy instruction in a mother tongue, as it does in Kenya, such mandates are often disregarded at the local level. In fact, many Kenyan parents oppose mother tongue instruction because they fear it will put their children behind at the higher-grade levels (Piper, Zuilkowski, Kwayumba, & Oyanga, 2018). Though existing evidence suggests that young children should be taught to read in their mother tongue (Ball, 2011), few children's storybooks are available in the indigenous languages of sub-Saharan Africa. A key question is whether parents—many of whom are illiterate—would make use of local language storybooks if they were available.

1.2 Dialogic Reading

Book-sharing has larger effects on children's language development when parents (or teachers) stimulate a dialogue related to the content of the story than if the parent reads aloud and the child passively listens (Mol, Bus, de Jong, & Smeets, 2008; Duursma, Augustyn, & Zuckerman, 2008). Dialogic reading is a set of book-sharing techniques designed to encourage children to formulate and articulate their ideas about book content, eventually acting as narrators who re-tell stories and actively engage instead of passively listening to them (Whitehurst et al., 1988; Zevenbergen & Whitehurst, 2003). The dialogic reading approach emphasizes the use of the "PEER" sequence and "CROWD" prompts (Zevenbergen and Whitehurst, 2003). In the PEER sequence, the adult prompts the child to say something about the book (P), evaluates the child's response (E), expands on the child's response by rephrasing and adding information to it (E), and repeats the prompt to make sure that the child has learned from the expansion (R). The CROWD prompts are completion (C), recall (R), open-ended (O), "wh-" questions (W), and distancing (D). The most important element of dialogic reading is the facilitation of conversation during storybook reading through contingent, and positive responsiveness (Blewitt & Langan, 2016), building on the understanding that responsive conversations between adults and children are needed to sustain receptive and promote expressive vocabulary (Tamis-LeMonda, Bornstein, & Baumwell, 2001; Werker & Hensch, 2015).

Dialogic reading programs have largely been implemented and evaluated in high-income countries, and have shown effects on emergent literacy skills and expressive vocabulary (Whitehurst et al., 1988; Zevenbergen & Whitehurst, 2003; Mol et al., 2008). Since the aim of dialogic reading is to stimulate a dialogue and not to read the text word-for-word (Duursma et al., 2008), these programs could be effective even in low-literacy populations.

A few studies in LMICs have shown benefits of dialogic reading, but most existing studies have targeted teachers, not parents. For example, a 5-day training in dialogic reading for pre-school teachers in Bangladesh improved expressive vocabulary scores in children (Opel, Ameer, & Aboud, 2009). Similarly, dialogic reading training for kindergarten teachers improved child phonological awareness in Egypt (Elmonayer, 2013) and receptive and expressive vocabulary in Turkey (Simsek & Erdogan, 2015). Indeed, most dialogic reading interventions in LMICs have been implemented in classroom settings, with teachers or

researchers engaging in dialogic reading with children. Few programs have targeted parents (Theriot et al., 2003), and of these, even fewer have been rigorously evaluated. Recently, a dialogic book-sharing program in South Africa that trained mothers on ways of engaging their 14- to 18-month-old children in book-sharing found greater sensitivity and content elaboration among intervention mothers, which, in turn, increased attention and vocabulary among children (Vally, Murray, Tomlinson, & Cooper, 2015; Murray et al., 2016).

1.3 Rationale for the Current Study

Integrated programs that combine parent education about nutrition and stimulation with home visits by trained specialists have been shown to improve child development outcomes in LMICs (Grantham-McGregor, Fernald, Kagawa, & Walker, 2014; Spier et al., 2016). However, the cost of "gold standard" home visiting programs can constrain the potential for scale-up; for example, the well-known Jamaica home visiting program was estimated to cost over $100 per child per year (Walker et al., 2015).

Integration of programs into existing health or educational infrastructure can allay programmatic costs and support scalability (Richter et al., 2017), however there are many challenges to successful multisector coordination. Deworming, an example of a simple and focused intervention built into existing infrastructure, has proven to have outsized effects on children's educational outcomes with a cost of about $3.50 per child per year (Kremer, 2003). However, evidence suggests that the World Health Organization (WHO) and Kenyan government policies for deworming have not been consistently followed in rural schools (Riesel, Ochieng', Wright, Vermund, & Davidson, 2010).

Center-based dialogic reading programs have positive impacts on child vocabulary in LMICs, however their scalability and sustainability are challenging in contexts with insufficient infrastructure to support them. Moreover, the youngest and most vulnerable children may not be reached by center-based programs (Martinez, Pereira, & Naudeau, 2012). There is an urgent need for evidence on focused interventions that are relatively low-cost and potentially scalable in LMICs, particularly parent education interventions that can successfully catalyze parents' energy to improve developmental outcomes for their young children.

To partially address these research gaps, we developed an intervention that combined locally appropriate children's storybooks with a modified dialogic reading training for primary caregivers of children aged two to six years. Our aim was to develop an intervention that would be valued by parents—so that they would use the storybooks and the lessons from the training, thus increasing the availability of early reading materials in the home, as well as the quantity and quality of book-sharing.

The first objective of the study was to test whether any variant of the intervention improved the quantity or quality of caregiver-child book-sharing. Varying the intensity of the training (one group session vs. two group sessions vs. group sessions plus a home visit) allowed us to test the hypothesis that the more involved interventions had larger effects, particularly on the quality of book-sharing. The second objective of the study was to assess the extent to which variants of our treatment increased child receptive and expressive vocabulary; we had

hypothesized that the more involved interventions would have a larger effect size. The final objective of the study was to test whether the treatments had differential effects on disadvantaged children, specifically with respect to caregiver literacy.

2. Methods and Materials

2.1 Storybook Selection and Modification

For the book component of our intervention, we began by conducting a survey of all bookstores, markets, and grocery stores in the greater Kisumu area to identify all readily-available children's storybooks. We were unable to find any storybooks intended for preschool-aged children that were printed in the dominant mother tongue, Luo. We pretested both existing, locally sourced English and Swahili storybooks and stories adapted from the African Storybook Project (http://www.africanstorybook.org), which we translated in English, Luo, and Swahili. We distributed these pilot books to households in peri-urban and rural communities surrounding Kisumu, and conducted follow-up interviews and focus groups to understand which books recipients had liked and why. Based on the feedback received, we partnered with Kenyan-owned Moran Publishers to adapt six of their (English-language) primary school level readers intended for beginning readers. In interviews and focus-groups, parents enjoyed the extremely detailed, contextually appropriate illustrations in the Moran books, which featured colorful pictures of African children living in rural areas and engaging in realistic activities—for example, a child taking a minibus to a local market with his mother. Members of the research team worked with Moran to elaborate the text of six storybooks, adding sections where parents could connect the story to children's experiences and including vocabulary related to the plots of the stories. The colorful illustrations and familiar content were expected to increase the likelihood that caregivers would respond positively to the books, use them, and repeat storybook vocabulary in their daily interactions (Hindman, Wasik, & Snell, 2016).

After adapting the English text of the storybooks, we translated them into Luo and Swahili. We used translators from the study areas to ensure that that vocabulary was appropriate for the dialect of the study population. We conducted forward and backward translations in each language to assure equivalency of the translations. Production of the final versions was overseen by a member of the research team who is a developmental psychologist with specific experience in creating content for young children in LMIC contexts. The storybooks were assembled into packets containing one of each title (six in total), with two titles in English, two titles in Swahili, and two titles in Luo. Which book was in which language varied across communities, though within each community, each caregiver received packets with the same book/language combination. The same books were distributed, regardless of child age within the 24- to 83-month range.

2.2 Adaptation of a Modified Dialogic Reading Training

For the second component of our intervention, we designed a modified dialogic reading training adapted to the cultural context and appropriate for a setting in which many caregivers had low levels of literacy and limited reading experience. We consider our intervention a modified dialogic reading program because we maintained the core message of dialogic reading—that children should be active and leading participants in book-sharing, with caregivers playing a supportive role as listeners and questioners that provide scaffolding for their child's continued exploration and learning (Whitehurst et al., 1988).

The foundational reading technique of dialogic reading is the PEER sequence, in which the adult prompts the child to say something about the book (P), evaluates the child's response (E), expands on the child's response by rephrasing and adding information to it (E), and repeats the prompt to make sure that the child has learned from the expansion (R) (Zevenbergen and Whitehurst, 2003). We adapted the PEER sequence to one that would be memorable and relevant to our context, "Read, Talk, Listen, and Expand." As most parents in our study area were accustomed to prompting children to identify objects in illustrations, we built on this foundation by introducing several of the CROWD dialogic reading prompts intended for older children, such as asking open-ended questions and "who, what, when, and why" prompts to parents, and for older children, prompts that relate the story and illustrations to the child's life (O, W, and D prompts). We did not introduce parents to completion or recall prompts (the C and R prompts).

The training we developed consisted of a 3-hour group session in which caregivers were taught skills about reading with their children (Whitehurst et al., 1994; Whitehurst et al., 1988), adapted from successful programs in South Africa (Vally et al., 2015) and Bangladesh (Opel et al., 2009). The core reading skills taught to parents were to identify objects in the illustrations, ask the child to identify objects in the illustrations, read the words that they knew how to read (or elicit help from others), and/or to support the attempts of their child to sound out words, ask the child simple (yes/no, who/what) questions, ask the child open-ended questions to elicit a dialogue, actively listen to the child and follow the child's lead in the conversation, and then to repeat what the child has said and expand on the child's responses by adding detail, and/or relating the story or illustrations to the child's life (Zevenbergen & Whitehurst, 2003). Parents received suggestions on how to select storybooks appropriate for their child, how to follow their child's lead when reading, and how to praise and encourage their child's engagement. We encouraged parents to read to their children as often as possible, with the goal of establishing a daily reading habit.

The group session included a question and answer period to address common barriers to reading with their children (e.g., too busy, child is uninterested), and to brainstorm solutions to challenges. We specifically addressed the challenge of caregiver literacy. For all caregivers, we emphasized the importance of engaging in a conversation with their child rather than reading the story word-for-word. For caregivers who could read most of the words, we suggested asking others about unknown words. For illiterate caregivers, we emphasized using the illustrations to create a story and enlisting the assistance of literate family members.

Content was delivered through an illustration and video-based PowerPoint presentation, with demonstrations and breakout sessions for parents to role play with partners and receive feedback in small groups from trainers. The training was delivered in Luo by a team (trained by the first author) who were from the study area; all trainers were native Luo speakers who were also fluent in English and Swahili. The training of the intervention team took course over a week, in which the team practiced reading the script of the presentation and using PowerPoint, as well as appropriate speaking pace, projection, eye contact, and tone. Trainers also learned how to facilitate role playing interactions of caregivers, and provide feedback and constructive guidance to parents, and practiced dialogic reading skills themselves with local children. The intervention team piloted the modified dialogic reading training in a pretesting community before implementation for the study. The full intervention was manualized and scripted, and supervisors were present at each training session to monitor fidelity and consistency of the implementation in every community. Supervisors did not report any deviations from script or protocol, but there was not a specific data metric used to measure fidelity.

2.3 Construction of the Baseline Sample

Our sample included 357 caregivers and 510 children aged 24 to 83 months. Study participants come from predominantly Luo-speaking areas within three hours of Kisumu, the third-largest city in Kenya. This area was selected because it has a low degree of ethnolinguistic heterogeneity. Sample communities were selected from rural parts of Homa Bay and Kisumu, Counties, where the overwhelming majority of the population speaks Luo as a mother tongue.

Our study area is a region of continuous settlement: communities are not separated by uninhabited farm or grazing land, and primary school catchment areas cut across village boundaries. This study is part of a longer-term research project evaluating the impact of early reading interventions on primary school preparedness and classroom-level outcomes. As such, for the purpose of our research, we define communities as primary school catchment areas. To select small, rural communities (where it was feasible to treat all the children entering primary school) for inclusion in our sample, nine primary schools were randomly chosen from the set of all public, coeducational day (i.e., not boarding) schools in Homa Bay and Kisumu Counties. We excluded larger schools (with average graduating cohorts of more than 20 students), and then further restricted the selection procedure so that no sampled school could be within 1.5 kilometers of another school in the study or within three kilometers of any school included in other ongoing research projects.

Within in each community (i.e., primary school catchment area) chosen for inclusion in our sample, we conducted a census to generate a listing of all households living within a radius of 750 meters of the school. Compounds including at least one child aged 24 to 83 months as well as the primary caregiver of that child met the inclusion criteria to participate in the study (irrespective of whether the household's children were enrolled in school). Through our census activities, we identified 592 households with children in the eligible age range. These

households resided in 524 unique compounds. A further 451 compounds did not include any children in the eligible age range.

In order to avoid cross-contamination, we randomly selected a subset of eligible households from neighborhoods where dwellings were less than 100 meters apart. When a community contained a large number of households grouped together, a subset was randomly chosen for inclusion in the baseline sample. We then grouped selected households that were less than 100 meters apart into clusters of no more than two households. Otherwise-eligible households that were excluded from the study for proximity reasons were invited to attend a book-sharing training that occurred after the follow-up data collection.

After randomly selecting eligible households from areas where dwellings were densely spaced, we generated a list of 410 eligible households that we invited to participate in the study. After obtaining written consent from the primary caregiver within a household, field workers enrolled participants between March and October of 2016. We enrolled 357 caregivers (87 percent of those invited) with 510 children between the ages of 24 and 83 months.

2.4 Study Design

Caregivers who completed the baseline survey were stratified by community and grouped into geographic clusters containing either one or two eligible caregivers each. Within each community, caregivers who completed the baseline were blocked by the number of caregivers within the geographic cluster; clusters were then randomly assigned to either the control group or one of four treatment arms (Figure 1). The study design was preregistered through the ISRCTN registry (id: ISRCTN68855267) accessible at: https://doi.org/10.1186/ISRCTN68855267. Study procedures were approved by local and international human subjects review boards.

The treatment arms were:

T1: Storybooks: Households received a packet of 6 children's storybooks—two books in English, two in Luo, and two in Swahili.

T2 = T1 + Training: In addition to T1 provisions, caregivers were invited to attend a modified dialogic reading training; they also received several text message (SMS) reminders encouraging them to read with their children in the week following the training session.

T3 = T2 + Booster: In addition to T2 provisions, households received a follow-up "booster" training session two weeks after the initial training.

T4 = T3 + Home Visit: In addition to T3 provisions, households received a home visit from one of the dialogic reading trainers approximately one week after the initial training.

Control: Households randomly assigned to the control group did not receive storybooks, training, or home visits between the baseline and follow-up assessments. Households were

informed that a second training session would be held approximately two months later. These sessions were held in each community after follow-up data had been collected. Both caregivers assigned to the control group and those who were not included in the baseline sample were invited to attend.

Thus, we varied the intensity of the treatment arms with the intention of identifying the (conceptual) dose required to yield an effect on parent reading behaviors and child outcomes. Similar interventions evaluated in LMICs have varied in duration from four to eight weeks (Hargrave & Sénéchal, 2000; Opel et al., 2009; Vally et al., 2015). All study participants lived in households with access to a mobile phone.

2.5 Implementation of the Interventions

We delivered packets of storybooks to the homes of caregivers who were randomly assigned to T1. Caregivers randomized to receive training (T2, T3, and T4) were invited to participate in a group parenting program that was held at a central meeting space within the community. At the end of the training session, packets of storybooks were distributed to all caregivers in attendance. Storybooks were also distributed to the homes of caregivers who did not attend the training.

Caregivers assigned to T2, T3, and T4 also received text messages that were intended as small encouragements to keep parents engaged in reading and reinforce the lessons of the training. Participants received SMS reminders two and three weeks after the training session. The first message asked a question about a main character in one of the storybooks, and the second message asked when caregivers read to their children. Caregivers who responded to each text message received a small amount of mobile airtime (delivered to their phones electronically).

Caregivers assigned to T3 and T4 were invited to attend a manualized and scripted booster training session two weeks after the initial training. The booster training was approximately 1.5 hours long, and was supervised to monitor fidelity. During the booster session, caregivers were given a presentation reviewing the key takeaways from the initial training, and were provided additional practice time during which they received feedback from trainers. The booster training included a question and answer session to address specific issues that caregivers had encountered while reading to their children.

Caregivers assigned to T4 received a home visit from a trainer one week after the initial training. Home visits lasted approximately one hour. During the home visit, the trainer followed a manual with scripted prompts to answer caregiver questions, discuss concerns and difficulties, and offer encouragement and support. Trainers also observed the caregiver reading to the child and offered feedback, reminding caregivers of the benefits of reading with their child. Supervisors conducted drop-in observations to ensure fidelity.

All trainers participated in the initial training and booster sessions, rotating through speaking parts of the script and small group leadership roles. Parents in the T2 to T4 arms interacted with the same set of trainers throughout the intervention. It was not feasible to assign

trainers to conduct the home visit (for T4) with the same caregivers that participated in their small group given the geographic spread and terrain of many of the communities, but all trainers who conducted the home visit also conducted the dialogic reading training and vice versa.

2.6 Data Collection Procedures

Baseline and follow-up data were collected through in-home interviews of primary caregivers and direct child assessments. Direct child assessments were administered in either Luo or Swahili. Enumerators asked children which language they preferred (English, Luo, or Swahili), and conducted the assessments in the child's preferred language. The only exceptions were the stimuli used in the English, Luo, and Swahili receptive vocabulary assessments, which were necessarily administered in the language being assessed—though explanations and prompts were always given in the child's preferred language. The overwhelming majority (94 percent) of children preferred to be interviewed in Luo, and the remainder preferred Swahili.

Baseline and follow-up data were collected by trained enumerators, all of whom had university degrees, were from the study areas, were native speakers of the local Luo dialect, and were also fluent in English and Swahili. Enumerators received 15 days of training from members of the research team on child anthropometry and developmental assessments. Enumerators also received extensive training and monitoring from senior research staff at the implementing organization, Innovations for Poverty Action – Kenya. Although not informed of treatment group assignment, enumerators were not blinded to treatment, as it would be immediately obvious when observing the children's books that were present in the home at endline. It was cost prohibitive to train two separate intervention and enumeration teams, but we conducted high-frequency checks during data collection to look for data falsification, and all field teams were led by supervisors who conducted daily drop-in observations of data collection. Baseline data was collected one week before the intervention. Follow-up was collected four to five weeks after the first modified dialogic reading training session.

2.7 Measures

We use data from an adapted version of the Family Care Indicators (FCI) questionnaire to address the first study objective (Bradley & Corwyn, 2002; Hamadani et al., 2010; Kariger et al., 2012). The FCI includes questions about the variety of play materials and children's books in the home, as well as stimulation activities in the three-day period preceding the interview. From the FCI data, we constructed a child-level indicator for having been read to in the past three days.

We collected two additional quantitative measures of book-sharing. First, we asked primary caregivers how frequently they had read to their young children over the seven-day period prior to the follow-up survey. This measure was distinct from the FCI questions about book-sharing because it was specific to the primary caregiver who was invited to attend the

modified dialogic reading training (in T2, T3, and T4). Second, to provide an objective measure of the extent to which children had used the storybooks, we also assessed their familiarity with book content using a set of 13 pictorial comprehension questions. Children were shown cropped or otherwise partially obscured illustrations from the books and were asked simple questions about them. Children could respond to the storybook comprehension questions in their preferred language. Questions were open-ended, but extensive piloting allowed us to build a comprehensive menu of likely responses, so coding the answers as correct or incorrect did not rely on enumerator discretion. From the binary responses to these questions, we constructed an age-normalized index (z-score) of storybook comprehension. The strength of this assessment is that it provides a measure of storybook use that does not rely on caregiver self-reports—with the caveat that it may also respond to children's independent play with the book.

To further address the first objective, the quality of book-sharing interactions between primary caregivers and their children was measured using an adapted version of the Mother-Child Picture Observation ("MCPO") assessment (Aboud, 2007; Aboud & Akhter, 2011; Rasheed & Yousafzai, 2015). The primary caregiver was given a locally sourced book not used in the intervention and was asked to demonstrate how they would normally interact with their child. Trained assessors live-coded caregivers' actions during 20 10-second intervals, with a 5-second coding window between each interval. Assessors indicated whether caregivers engaged in (i) directive reading (e.g., reading words or sentences, pointing out and naming objects); (ii) more advanced dialogic reading (e.g., asking children questions, expanding on their statements, asking them to elaborate or expand their statements, or answering their questions); or (iii) were distracted or off-task. The number of 10-second intervals during which the caregiver engaged in each type of behavior was summed, creating ordinal indices of basic reading activities, interactive reading activities, and disengagement. Assessor training for the observations consisted of a discussion of the definitions of each code and refinement until group consensus was achieved. Assessors practiced coding videotaped observations of local parent-child dyads from non-study locations until inter-rater reliability was consistently greater than 90 percent. Live coding in tight increments could have potentially affected reliability in the field, but it was not feasible to videotape caregivers and children during data collection.

To address the second study objective, we assessed receptive vocabulary in Luo, English, and Swahili using adapted versions of the British Picture Vocabulary Scale (BPVS) - III (Dunn & Dunn, 2009), which includes 168 items for use with children aged 3 to 17. In the standard BPVS, knowledge of receptive vocabulary is measured by asking the respondent to point to one of four pictures that corresponds to a word (object, person, or action) spoken by the assessor. The BPVS has previously been adapted to create receptive vocabulary tests appropriate for coastal Kenya (Holding et al., 2004) and Indonesia (Prado, Alcock, Muadz, Ullman, & Shankar, 2012). As discussed above, one of the key considerations in our setting was the need to develop assessments that captured the development of vocabulary skills in a multilingual environment. There are trade-offs for assessing vocabulary in multilingual contexts, as children may have a larger vocabulary in their first language, or in the language of instruction. For our study we were interested in capturing children's development in both,

so we developed Luo, English, and Swahili subscales and adapted the BPVS to the local context and ages of the children in our study (Fernald, Prado, Kariger, & Raikes, 2017). To minimize the testing burden, we extensively pretested all items and ordered them by difficulty using item response theory, with a final assessment of 27 Luo items (Cronbach's alpha: $\alpha = 0.78$, Cohen's Kappa for IRR: $\kappa = 1$), 32 Swahili items ($\alpha = 0.76$, $\kappa = 0.89$), and 34 English items ($\alpha = 0.57$, $\kappa = 0.95$) that demonstrated variability among children in our study area and age range (author manuscript, under review). While this adaptation decreased comparability of our vocabulary scores to other studies or populations, we were primarily concerned with developing measures that would be appropriate for the multilingual study environment.

In collaboration with local artists, we also developed a locally appropriate measure of expressive vocabulary after pre-testing suggested that the stimulus words and pictures in existing measures of expressive language were not appropriate to our context (author manuscript, under review). The measure was based on words from the BPVS, ordered by difficulty. This 30-item measure of expressive vocabulary was administered by having the assessor show a picture and ask the respondent to name the object or concept the illustration depicted. Each item was scored as 0 or 1, with credit received for responding in English, Luo, or Swahili (α=0.66, κ=0.95). To assess children's knowledge of the vocabulary words embedded in the storybooks, we added seven expressive vocabulary items from the project storybooks (crayon, goose, kite, nest, rabbit, toothbrush, and umbrella). For all vocabulary outcomes, age-normalized z-scores were constructed by estimating the relationship between age and the outcome variable non-parametrically (Fan, 1993).

Child characteristics measured at baseline included age (in months), sex, and height (Onis, 2006). Primary caregiver characteristics measured included the caregiver's relationship to the child, educational attainment (in years of completed schooling), and depressive symptoms in the past week (CESD-10; Kilburn et al., 2018). We constructed a measure of household wealth based on housing quality (crowding, roofing and floor materials, access to electricity, sanitation facilities) and ownership of durable assets such as vehicles and livestock (Filmer & Pritchett, 2001).

To address the third study objective, we assessed caregiver literacy by asking caregivers to read one simple sentence (second grade level) in each of the three study languages—Luo, English, and Swahili. Caregivers who either stated that they could not read or who were unable to correctly read more than one word in any of the three languages were categorized as illiterate.

2.8 Balance and Attrition

Most baseline child characteristics (e.g., age, sex, height-for-age z-score, school enrollment, and receptive vocabulary) did not vary across study arms, though we did observe some variation in caregiver and household characteristics (Supplementary Tables A.1 and A.2). Given the large number of treatments, some statistically significant imbalances are to be

expected. We included controls for variables that showed statistically significant levels of baseline imbalance in our analysis.

Households were followed-up five to six weeks after baseline, with 48 caregivers and 68 children lost to follow-up. Attrition was not correlated with either randomly assigned treatments or baseline characteristics (Supplementary Tables A.3 and A.4), and thus is unlikely to bias the estimates of treatment effects.

2.9 Analytic Approach

To address the first and second study objectives, we conducted intent-to-treat analyses for all study groups and outcomes of interest. In our main analysis, we estimated OLS regression models including mutually exclusive indicators for the treatment arms (T1 through T4) as independent variables in the same regression, estimating a separate model for each of the outcome variables of interest. We included control variables to improve power and account for baseline differences between treatment arms; the variables were child age and sex, an indicator for whether the mother was the primary caregiver, household size, a household wealth index, primary caregiver depressive symptoms, and community fixed effects. When baseline values of the outcome variable were available, we included them in the regression to maximize statistical power. Cluster-robust standard errors were used to account for the geographic clustering in the random assignment of treatments.

To address the third objective, we estimated the interaction of treatment group assignment and caregiver literacy. We also included an uninteracted indicator for being illiterate, to capture the outcome difference between literate and illiterate caregivers in the control group. All analyses were conducted using Stata 14.2 (StataCorp, L. P., 2016).

3. Results

3.1 Descriptive Statistics

The mean baseline age of children in our follow-up sample was 54.77 months; 52 percent of children were male; 12 percent were stunted; and 84 percent of primary caregivers in the sample are the mothers of the sample children (Table 1). Most children in our sample did not have access to age-appropriate reading materials prior to our study: Only 13 percent of homes had any children's books (Table 1). At baseline, no household in our sample owned a children's storybook in Luo, the first language of all of the sample children. Despite this, about half of the caregivers reported that children had been read to in the three days prior to the baseline survey.

3.2 Study Objective 1: Treatment Effects on Book-Sharing

Randomization to any of the four treatment groups increased the likelihood that children had been read to in the three days prior to the follow-up survey, and we cannot reject the hypothesis that storybooks alone generated impacts that were as large as those of the more intensive treatments (Table 2). Children in all treatment arms also demonstrated significantly higher storybook comprehension compared to children in the control group, suggesting that the observed treatment effects on reading frequency were genuine (and not explained by inaccurate self-reports among caregivers assigned to treatment). However, assignment to T1 did not have a statistically significant impact on the frequency of caregiver-child reading interactions over the week prior to the follow-up (Table 2).

The treatments involving modified dialogic reading training all increased the frequency of caregiver-child reading interactions over the week prior to the follow-up (Table 2), and they specifically increased the likelihood that caregiver-child book-sharing occurred at least four times in the week prior to the follow-up survey (results not shown). In the Mother Child Picture Observation assessment, all three treatments involving modified dialogic reading training increased the number of dialogic reading behaviors observed and decreased the amount of time spent distracted or off-task (Table 2).

The addition of a booster training session (T3) or home visit (T4) did not further increase the quantity or quality of caregiver-child book-sharing. Relative to T2, the more intensive treatments did not have significantly larger impacts on the likelihood that anyone had read to a child, the frequency of caregiver-child book-sharing, storybook comprehension, the number of dialogic reading interactions observed, or the amount of time spent off-task during the MCPO assessment (Table 2).

3.3 Study Objective 2: Treatment Effects on Child Vocabulary

T2 and T3 increased book-specific expressive vocabulary skills, and we cannot reject the hypothesis that T4 had an equally large effect (though the main effect of T4 is not statistically significant, with a *p*-value of 0.107). There were no significant effects of any of the treatments on children's knowledge of expressive vocabulary that was not specific to the storybooks, or on receptive vocabulary in English, Luo, or Swahili (Supplementary Table A.5).

3.4 Study Objective 3: Interaction Effects by Caregiver Literacy

The impacts of treatment were generally larger for illiterate caregivers than for literate caregivers (Supplementary Table A.6). At baseline, the children of illiterate caregivers were 18 percentage points less likely to have had anyone read to them in the three days prior to the survey, and our intervention appears to partially address these baseline gaps (Figure 2). Being randomized to any of the treatment arms increased the likelihood that anyone had read to a child of an illiterate caregiver in the last three days; impacts were significantly larger for the children of illiterate caregivers than for the children of literate caregivers in T4 (Supplementary Table A.6). In addition, T3 had a differentially larger impact on reading

frequency for the children of illiterate caregivers, while T4 had a differentially larger impact on storybook-specific expressive vocabulary for the same set of children (Supplementary Table A.6).

Being randomized to any of the four treatment arms significantly increased comprehension for the children of both literate and illiterate caregivers (all eight p-values < 0.001), but caregiver literacy did not modify the effects of treatment on storybook comprehension (Supplementary Table A.6). Caregiver literacy also did not modify the effect of treatment on the quality of caregiver-child reading interactions (as measured through the Mother-Child Picture Observation). For the more intensive treatments (T3 and T4), there were larger coefficient estimates for illiterate caregivers than for literate caregivers, but the differences were not statistically significant.

4. Discussion

In this cluster randomized controlled trial, we found that distributing storybooks encouraged caregivers to read with their children and increased children's knowledge of book content. Adding modified dialogic reading training for the parents to the distribution of storybooks was necessary to improve the quality of their reading interactions with their children, and to increase children's knowledge of vocabulary words from the storybooks. We found that one training session was sufficient to achieve these results, and higher doses of contact with caregivers did not further increase impacts on reading frequency, storybook comprehension, storybook-specific expressive vocabulary, or the quality of reading interactions. While, for most outcomes, there were no differences in intervention effectiveness by literacy of the caregiver, improvements in storybook-relevant expressive vocabulary proved to be largest for children of illiterate caregivers.

4.1 Comparison of Study Effects to Other Interactive Reading Training Programs

Previous dialogic reading interventions targeting parents have had a moderate positive effect (d = 0.59) on children's expressive vocabulary and a small effect (d = 0.22) on receptive vocabulary, but effects were smaller among children "at risk of literacy delays" (Mol et al., 2008). We document small but statistically significant impacts on storybook-specific expressive vocabulary (between 0.152 and 0.292 SD, as shown in Table 2) in an at-risk population with low baseline levels of caregiver literacy and limited access to age-appropriate reading materials. It is therefore unsurprising that we do not detect impacts on receptive vocabulary, as we might expect them to be smaller than impacts on expressive vocabulary, and thus perhaps statistically undetectable in this sample. One previous study comparing dialogic reading to book distribution without parental education did not detect any impacts of book distribution alone, but the study population had substantially higher quality home literacy environments at baseline compared with our population (Chow, McBride-Chang, Cheung, & Chow, 2008).

While dialogic reading has improved expressive vocabulary among children in at-risk populations (Valdez-Menchaca & Whitehurst, 1992; Whitehurst et al., 1994; Hargrave & Sénéchal, 2000), previous work has shown that children from higher-income families are more likely to benefit from vocabulary-related interventions than are children from lower-income backgrounds, potentially exacerbating the word gap (Marulis & Neuman, 2010). Our findings provide two sharp contrasts to this pattern in previous work. First, the area in which we implemented the intervention is very poor and very rural; the vast majority (86 percent) of the households in our study did not have any children's storybooks in the home prior to the intervention. Thus, the presence of positive effects in this population is already a promising counterpoint to findings in the existing literature. Second, even within our study area, there remains marked variation in the levels of stimulation that children experience: in the absence of treatment, children of illiterate caregivers were exposed to less parental engagement and had significantly lower (age-adjusted) levels of expressive and English-language receptive vocabulary than children of literate caregivers. Yet even for this relatively more disadvantaged group, we found that a modified dialogic reading training empowered illiterate caregivers to engage their children in storybook-centered conversations. The intervention did not exacerbate disparities between children of literate and illiterate caregivers: it actually had greater effects on the likelihood of adults engaging in shared reading and expressive vocabulary for children in low-literacy households. Thus, along some dimensions, our intervention closed the gap that existed between children of literate and illiterate caregivers (Figure 2).

4.2 Relationship to Existing Parenting and School-Based Literacy Interventions in LMICs

Parenting interventions typically aim to improve the quantity and quality of parents' literacy engagements with their children, but in LMICs they must also address the lack of available learning resources for children. There has been growing interest in increasing the availability of books for children in African languages, but the linguistic complexity and variety of languages in Africa present challenges to storybook authors (Edwards & Ngwaru, 2012). Our study demonstrates that it is possible to partner with a local publisher to produce storybooks in local languages, with a result that parents value: parents kept the books that were distributed. Hence, distributing low-cost books ($1/book) was effective at improving children's home literacy environment. Finally, providing books in several languages gave caregivers the option of choosing the language to read to their child in. Most caregivers reported that they read with their child in Luo (70 percent) followed by English (16 percent).

There are a number of school based literacy programs that have been demonstrated to be effective in sub-Saharan Africa and in other LMICs—for example, Literacy Boost, Tusome, PRIMR, and Mango Tree—but these programs are center-based, emphasizing teacher training, and targeting school-aged children (Dowd et al., 2013; Piper, Destefano, Kinyanjui, & Ong'ele, 2018; Piper et al., 2015; Kerwin & Thornton, 2018). Modifications to center-based programs may have logistical advantages in some settings, but they may be less likely to reach younger children. In the present study, only 42 percent of 3-year-olds were enrolled in pre-schools. In a study in Mozambique, even conditional on the availability of high-quality

pre-schools in the community, only 55.6 percent of 3- and 4-year-olds were enrolled in pre-schools (Martinez et al., 2012). Scaling up access to high-quality preschool programs in LMICs may be costlier and require more political will than implementing parenting education programs; parenting programs may have broader reach in rural areas, especially for younger children (Daelmans et al., 2017). A comparative study of preschool and parenting education programs in Ethiopia by Save the Children found the parenting program to be as effective at improving children's early literacy skills and language development as the preschool program (Borisova, Pisani, Dowd, & Lin, 2017). Interventions targeting parents and center-based programs can easily be complements: there is no reason a policymaker could not aim for improvements on both margins. For example, a study of a childcare center in Brazil found that adding parent reading training to their program yielded significant improvements in parent reading behaviors, and higher child receptive vocabulary and working memory scores after 9 months (Weisleder et al., 2017). In reflecting on the policy options, we think it is helpful to be clear on the relative advantages of each approach.

4.3 Limitations

There were several limitations of this study. First, the follow up was conducted five to six weeks after the intervention, giving us the opportunity to see only short-run effects of the intervention. This timing is in line with existing work documenting the effects of dialogic reading interventions four to six weeks after treatment (cf. Whitehurst et al., 1988; Hargrave & Sénéchal, 2000). However, our design does not allow us to assess the longevity of these effects. A longer term evaluation is underway to examine if parents retain what they learned from the training and continue to practice dialogic reading, and if children demonstrate lasting gains in their vocabulary.

Though we explored conceptual "dosage" by varying the intensity of the training and frequency of interactions with trainers across treatment arms, even the highest-intensity variant remained a very light intervention in comparison to regular home-visit programs. Caregivers interacted with trainers no more than four times over the course of a month. This clearly poses a limitation in that we did not trace the full arc of possible effects by ever-higher program intensities; our goal here, however, was to test whether a modest intervention could have impacts in a population such as this one.

For intervention implementation, we were able to monitor fidelity of book distribution, but did not collect data on deviations from the protocol or script of the dialogic reading training. We relied on supervisors to be present at each group session, and conduct spot checks of home visits and data collection procedures at participants' homes, and no substantial deviations from the script or the protocol were reported. For the MCPO, we trained interviewers until we achieved inter-rater reliability above 90 percent before implementation in the field. Unfortunately, it was not feasible to video-tape caregiver-child interactions to assess inter-rater reliability in the field. However, there were no observer fixed effects in our data, meaning that there were no systematic effects of the interviewers on the measured outcome.

The age range of the children in the study was wide, and we observed substantial differences in program effect on storybook comprehension according to child age—though we found positive treatment effects on comprehension for children at all age levels (Supplementary Materials Figure 1). In our analysis, we controlled for the potential confounding effect of child age on all outcomes, and examined differential treatment effects by age. We chose to use the same set of storybooks and comprehension questions for all children so that treatment effects would be comparable and well-defined. This reflects the tension in the age-targeting of the intervention: intervening early may, for some interventions, produce the largest impacts, yet at later ages, book comprehension is much easier to assess. In an LMIC context, we expect high levels of heterogeneity in emergent literacy skills at older ages, so age-specific targeting may be more challenging than in HICs. Nevertheless, the broad age range might plausibly have diluted the hypothetical effects of a more age-specific variant of this intervention.

Finally, our relatively simple intervention was not intense enough to overcome all deficits these households face. However, we were able to deliver the full intervention for $28.27 per household. The books were about $1 each, and the staff time for the intervention cost about $12.96 per household. An additional $9.31 per household was spent on adaptation and training, transportation, equipment, text messages and airtime, local guides, and refreshments for participants. Thus, our light touch (single session) intervention can be effectively delivered at substantially lower cost than "gold standard" home visitation programs (Walker et al., 2015). The design of our training (using videos, demonstrations, small group practice and feedback, etc.) made the intervention accessible to caregivers with limited or no literacy, and only required a single group training session. This low-cost approach has the potential to address the word gap between the poor, rural children in our study population, and wealthier children: we found evidence of greater treatment effects on storybook-relevant vocabulary for the most vulnerable children, those with illiterate caregivers.

4.4 Future Directions

Our storybook and parent education intervention is immediately applicable to about 4 million Luo-speakers in Kenya, and if translated to other mother tongues, could benefit up to 6 million children ages 2 to 6 years old in Kenya (Kenya National Bureau of Statistics, 2010). This intervention has the potential to reach approximately 40 million children in sub-Saharan Africa (UNICEF, 2017), as a lack of children's storybooks and high rates of caregiver illiteracy are not specific to Kenya, and our publishing partner is active in six countries in the region. The scripted and manualized modified dialogic reading training could be successfully delivered by trained laypersons, and being PowerPoint based, is easily portable. A larger scale study of the intervention with a longer follow up could yield a definitive picture of the scalability and cost-effectiveness of this intervention.

5. Conclusions

Investing in children early in their development can have high returns over the life cycle because early childhood is a period of rapid growth and development (Campbell et al., 2014). Our study suggests two important conclusions. First, a well-designed parent-child book-sharing training program can be implemented in a high-poverty setting with low literacy among caregivers, with positive effects on caregivers' reading interactions with their children. Second, children as young as two years old are able to engage with and benefit from local-language, contextually appropriate children's storybooks. Simple storybooks are also engaging for older children, who appear to become very familiar with their content—suggesting regular and sustained use.

Ours is among the first studies to test reading training programs in a context with a substantial fraction of illiterate caregivers; not only is this one of the most important dimensions of socioeconomic status for interventions of this type, we find that it is children of the illiterate caregivers who appear to benefit the most from the intervention. The potential to reduce educational inequality is one of the many potential advantages of a parent-focused program such as this one, in relation to more common center-based programs. Longer term follow-up study is needed to build on these promising findings, to determine the effectiveness and cost-effectiveness of the intervention, and to establish whether effects on children's vocabulary development last beyond the short term.

References

Aboud, F. E. (2007). Evaluation of an early childhood parenting programme in rural Bangladesh. *Journal of Health, Population and Nutrition, 25*(1), 3-13.

Aboud, F. E., & Akhter, S. (2011). A cluster-randomized evaluation of a responsive stimulation and feeding intervention in Bangladesh. *Pediatrics, 127*(5), e1191-e1197.

Ball, J. (2011). *Enhancing learning of children from diverse language backgrounds: mother tongue-based bilingual or multilingual education in the early years.* Paris, France: UNESCO.

Behrman, J. R., Hoddinott, J., Maluccio, J. A., Soler-Hampejsek, E., Behrman, E. L., Martorell, R., . . . Stein, A. D. (2006). What determines adult skills? Impacts of pre-school, school-years, and post-school experiences in Guatemala. *Latin American Econmic Review*, *23*(4), 1-32.

Black, M. M., Walker, S. P., Fernald, L. C., Andersen, C. T., DiGirolamo, A. M., Lu, C., ... Devercelli, A. E. (2017). Early childhood development coming of age: science through the life course. *The Lancet*, *389*(10064), 77-90.

Blewitt, P., & Langan, R. (2016). Learning words during shared book reading: the role of extratextual talk designed to increase child engagement. *Journal of Experimental Child Psychology, 150*, 404-410.

Borisova, I., Pisani, L., Dowd, A. J., & Lin, H.-C. (2017). Effective interventions to strengthen early language and literacy skills in low-income countries: comparison of a family-focused approach and a pre-primary programme in Ethiopia. *Early Child Development and Care, 187*(3-4), 655-671.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*(1), 371-399.

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *Future Child, 7*(2), 55-71.

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science, 343*(6178), 1478-1485.

Can, D. D., & Ginsburg-Block, M. (2016). Parenting stress and home-based literacy interactions in low-income preschool families. *Journal of Applied Developmental Psychology, 46*, 51-62.

Chow, B. W. Y., McBride-Chang, C., Cheung, H., & Chow, C. S. L. (2008). Dialogic reading and morphology training in Chinese children: effects on language and literacy. *Developmental Psychology, 44*(1), 233.

Clegg, J., & Simpson, J. (2016). Improving the effectiveness of English as a medium of instruction in sub-Saharan Africa. *Comparative Education, 52*(3), 359-371.

Daelmans, B., Darmstadt, G. L., Lombardi, J., Black, M. M., Britto, P. R., Lye, S., . . . Richter, L. M. (2017). Early childhood development: the foundation of sustainable development. *The Lancet, 389*(10064), 9-11.

Dowd, A. J., Friedlander, E., Guajardo, J., Mann, N., & Pisani, L. (2013). *Literacy Boost cross country analysis results*. Washington, D.C. : Save the Children.

Dunn, L., & Dunn, D. (2009). The British Picture Vocabulary Scale Third Edition BPVS III. *National Foundation for Educational Research.*

Duursma, E., Augustyn, M., & Zuckerman, B. (2008). Reading aloud to children: the evidence. *Archives of Disease in Childhood*, 93(7), 554-557.

Edwards, V., & Ngwaru, J. M. (2012). African language books for children: issues for authors. *Language Culture and Curriculum, 25*(2), 123-137.

Elmonayer, R. A. (2013). Promoting phonological awareness skills of Egyptian kindergarteners through dialogic reading. *Early Child Development and Care, 183*(9), 1229-1241.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, 196-216.

Fernald, L. C., Prado, E., Kariger, P., & Raikes, A. (2017). *A toolkit for measuring early childhood development in low- and middle-income countries*. Washington, D.C. : World Bank Group.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography, 38*(1), 115-132.

Grantham-McGregor, S. M., Fernald, L. C., Kagawa, R., & Walker, S. (2014). Effects of integrated child development and nutrition interventions on child development and nutritional status. *Annals of the New York Academy of Sciences*, *1308*(1), 11-32.

Hamadani, J. D., Tofail, F., Hilaly, A., Huda, S. N., Engle, P., & Grantham-McGregor, S. M. (2010). Use of family care indicators and their relationship with child development in Bangladesh. *Journal of Health, Population and Nutrition, 28*(1), 23-33.

Hargrave, A. C., & Sénéchal, M. (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.

Hindman, A. H., Wasik, B. A., & Snell, E. K. (2016). Closing the 30 million word gap: next steps in designing research to inform practice. *Child Development Perspectives, 10*(2), 134-139.

Holding, P. A., Taylor, H. G., Kazungu, S. D., Mkala, T., Gona, J., Mwamuye, B., . . . Stevenson, J. I. M. (2004). Assessing cognitive outcomes in a rural African population: development of a neuropsychological battery in Kilifi District, Kenya. *Journal of the International Neuropsychological Society, 10*(02), 246-260.

Jukes, M. C. H., Gabrieli, P., Mgonda, N. L., Msolezi, F. S., Jeremiah, G., Tibenda, J. J., Bub, K. L. (2018). "Respect is an investment": community perceptions of social and emotional competencies in early childhood from Mtwara, Tanzania. *Global Education Review*, *5*(2), 160-188.

Kariger, P., Frongillo, E. A., Engle, P., Britto, P. M., Sywulka, S. M., & Menon, P. (2012). Indicators of family care for development for use in multicountry surveys. *Journal of Health, Population and Nutrition, 30*(4), 472-486.

Kenya National Bureau of Statistics (2010). *The 2009 Kenya population and housing census. Volume IC, population distribution by age, sex and administrative units*. Nairobi, Kenya: Author.

Kenya National Bureau of Statistics (2013). *Nyanza Province Multiple Indicator Cluster Survey 2011, Final Report*. Nairobi, Kenya: Author.

Kerwin, J. T., & R. Thornton. (2018). Making the grade: the trade-off between efficiency and effectiveness in improving student learning. *Mimeo*.

Kilburn, K., Prencipe, L., Hjelm, L., Peterman, A., Handa, S., & Palermo, T. (2018). Examination of performance of the Center for Epidemiologic Studies Depression Scale

Short Form- 10 among African youth in poor, rural households. *BMC Psychiatry*, *18*(1), 201.

Kremer, M. (2003). Randomized evaluations of educational programs in developing countries: some lessons. *American Economic Review*, *93*(2), 102-106.

Martinez, S., Naudeau, S., & Pereira, V. (2012). *The promise of preschool in Africa: a randomized impact evaluation of early childhood development in rural Mozambique*. Washington D.C. : The World Bank.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: a meta-analysis. *Review of Educational Research, 80*(3), 300-335.

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent-child book readings: a meta-analysis. *Early Education and Development, 19*(1), 7-26.

Murray, L., De Pascalis, L., Tomlinson, M., Vally, Z., Dadomo, H., MacLachlan, B.,... Cooper, P. J. (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,* 1370–1379.

Ninio, A. (1983). Joint book reading as a multiple vocabulary acquisition device. *Developmental Psychology*, *19*(3), 445-451.

Onis, M. (2006). WHO child growth standards based on length/height, weight and age. *Acta Paediatrica, 95*(S450), 76-85.

Opel, A., Ameer, S. S., & Aboud, F. E. (2009). The effect of preschool dialogic reading on vocabulary among rural Bangladeshi children. *International Journal of Educational Research, 48*(1), 12-20.

Park, H. (2008). Home literacy environments and children's reading performance: a comparative study of 25 countries. *Educational Research and Evaluation, 14*(6), 489-505.

Piper, B. (2010). Kenya Early Grade Reading Assessment Findings Snapshot. Research Triangle Park, North Carolina, USA: RTI International.

Piper, B., Destefano, J., Kinyanjui, E. M., & Ong'ele, S. (2018). Scaling up successfully: lessons from Kenya's Tusome national literacy program. *Journal of Educational Change*, *19*(3), 293-321.

Piper, B., Jepkemei, E., & Kibukho, K. (2015). Pro-poor PRIMR: improving early literacy skills for children from low-income families in Kenya. *Africa Education Review*, *12*(1), 67-87.

Piper, B., Zuilkowski, S. S., Kwayumba, D., & Oyanga, A. (2018). Examining the secondary effects of mother-tongue literacy instruction in Kenya: impacts on student learning in English, Kiswahili, and mathematics. *International Journal of Educational Development*, *59*, 110-127.

Prado, E. L., Alcock, K. J., Muadz, H., Ullman, M. T., & Shankar, A. H. (2012). Maternal multiple micronutrient supplements and child cognition: a randomized trial in Indonesia. *Pediatrics*, e536-e546.

Raikes, H., Alexander Pan, B., Luze, G., Tamis-LeMonda, C. S., Brooks-Gunn, J., Constantine, J., . . . Rodriguez, E. T. (2006). Mother–child bookreading in low-income families: correlates and outcomes during the first three years of life. *Child Development, 77*(4), 924-953.

Rasheed, M.A., & Yousafzai, A. K. (2015). The development and reliability of an observational tool for assessing mother-child interactions in field studies- experience from Pakistan. *Child Care Health and Development*, *41*, 1161–1171.

Richter, L. M., Daelmans, B., Lombardi, J., Heymann, J., Boo, F. L., Behrman, J. R., . . . Darmstadt, G. L. (2017). Investing in the foundation of sustainable development: pathways to scale up for early childhood development. *Lancet*, *389*(10064), 103–118.

Riesel, J. N., Ochieng', F. O., Wright, P., Vermund, S. H., & Davidson, M. (2009). High prevalence of soil-transmitted helminths in Western Kenya: failure to implement deworming guidelines in rural Nyanza Province. *Journal of Tropical Pediatrics*, *56*(1), 60-62.

Rodriguez, E. T., Tamis-LeMonda, C. S., Spellmann, M. E., Pan, B. A., Raikes, H., Lugo-Gil, J., & Luze, G. (2009). The formative role of home literacy experiences across the first three years of life in children from low-income families. *Journal of Applied Developmental Psychology, 30*(6), 677-694.

Simsek, Z. C., & Erdogan, N. I. (2015). Effects of the dialogic and traditional reading techniques on children's language development. *Procedia-Social and Behavioral Sciences, 197*, 754-758.

Spier, E., Britto, P., Pigott, T., Kidron, Y., Lane, J., Roehlkepartain, E., . . . Song, M. (2016). Parental, familial, and community support interventions to improve children's literacy in developing countries: a systematic review. *Campbell Systematic Reviews, 12*(4).

StataCorp, L. P. (2016). Stata SE 14.2 for Windows 64-bit. College Station, TX: StataCorp

Tamis-LeMonda C.S., Bornstein, M.H., & Baumwell, L.(2001). Maternal responsiveness and children's achievement of language milestones. *Child Development,* 72, 748–67.

Theriot J.A., Franco S.M., Sisson B.A., Metcalf S.C., Kennedy M.A., & Bada H.S. (2003). The impact of early literacy guidance on language skills of 3-year-olds. Clinical Pediatrics, *42*, 165–172.

United Nations Children's Fund (UNICEF). (2017). *State of the worlds children 2017: children in a digital world*. New York: UNICEF.

Valdez-Menchaca, M. C., & Whitehurst, G. J. (1992). Accelerating language development through picture book reading: A systematic extension to Mexican day care. *Developmental Psychology, 28*(6), 1106-1114.

Vally, Z., Murray, L., Tomlinson, M., & Cooper, P. J. (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*, 865–873.

Walker, S. P., Powell, C., Chang, S. M., Baker-Henningham, H., Grantham-McGregor, S., Vera-Hernandez, M., López-Boo, F. (2015). *Delivering parenting interventions through health services in the Caribbean: impact, acceptability and costs. IDB Working Paper Series, No. IDB-WP-642*. Washington, D.C. : Inter-American Development Bank (IDB).

Wasik, B. A., & Hindman, A. H. (2015). Talk alone won't close the 30-million word gap. *Phi Delta Kappan, 96*(6), 50-54.

Weber, A., Fernald, A., & Diop, Y. (2017). When cultural norms discourage talking to babies: effectiveness of a parenting program in rural Senegal. *Child Development*, *88*(5), 1513-1526.

Weisleder, A., Mazzuchelli, D. S., Lopez, A. S., Neto, W. D., Cates, C. B., Gonçalves, H. A., . . . Mendelsohn, A. L. (2017). Reading aloud and child development: a cluster-randomized trial in Brazil. *Pediatrics*, *41*(1), 1-12.

Weitzman, C. C., Roy, L., Walls, T., & Tomlin, R. (2004). More evidence for reach out and read: a home-based study. *Pediatrics*, *113*(5), 1248-1253.

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: new directions. *Annual Review of Psychology,* 66, 173–96.

Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology, 30*(5), 679-689.

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

Zauche, L. H., Thul, T. A., Mahoney, A. E. D., & Stapel-Wax, J. L. (2016). Influence of language nutrition on children's language and cognitive development: an integrated review. *Early Childhood Research Quarterly, 36*, 318-333.

Zevenbergen, A. A., & Whitehurst, G. J. (2003). Dialogic reading: a shared picture book reading intervention for preschoolers. In A. van Kleeck, S.A. Stahl, & E.B. Bauer (Eds.), *On Reading Books to Children: Parents and Teachers* (pp. 177-200). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Figure 1. CONSORT flow diagram

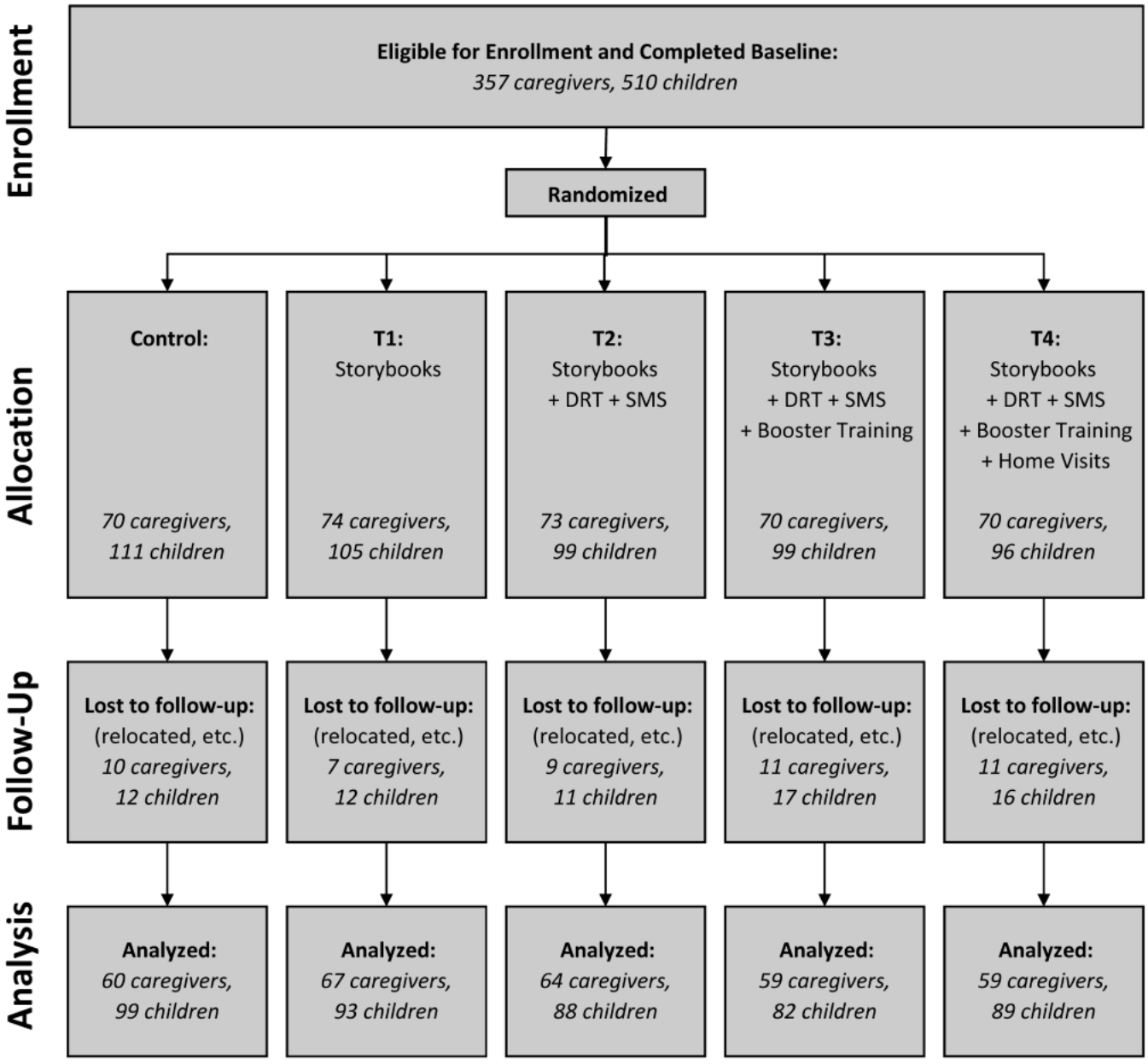
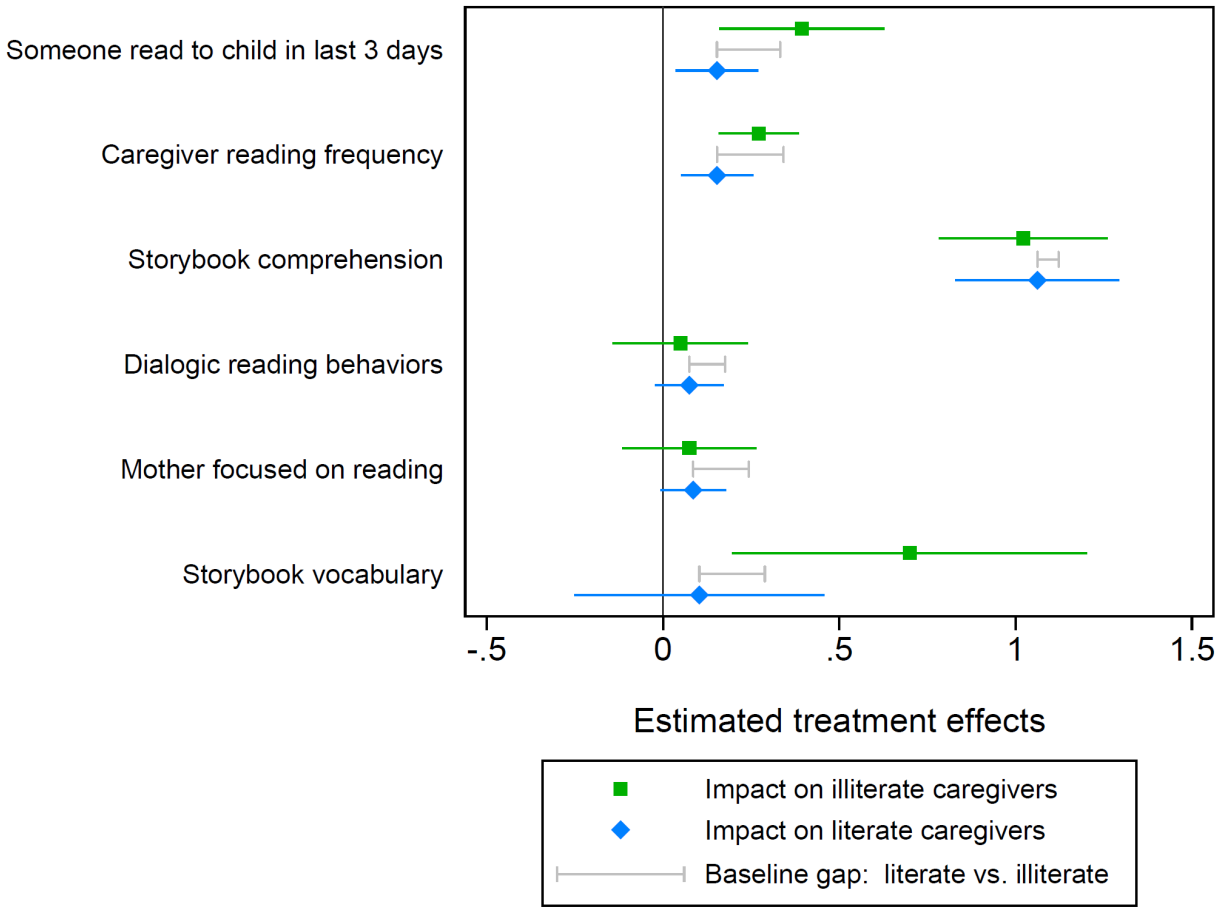


Figure 2. Aggregate program impacts by caregiver literacy in relation to baseline gaps



Notes. The "pooled" treatment effects are shown for treatment arms T1 through T4 (any treatment vs control), interacted with caregiver literacy. Caregiver literacy was a defined as caregiver's ability to read three simple sentences with no more than one error in English, Swahili, and Luo. The point estimate is indicated by a square (illiterate caregiver) or diamond (literate caregiver), with the 95-percent confidence interval represented by the line going through the square or diamond. The gray bar represents the gap between outcomes, in the absence of the intervention, for literate and illiterate caregivers (or their children). Outcomes were scaled for graphical representation, as follows: "Someone read to child in last 3 days," is binary: 1=yes, 0=no. "Reading frequency" ranges from zero to one, where one represents reading every day of the past week, while zero represents no reading in the past week. "Storybook comprehension" and "Storybook vocabulary," are age-normalized to have a conditional mean of zero and standard deviation of one. "Dialogic reading behaviors" and "Mother focused on reading" are observations of caregiver reading behaviors scaled to range from zero to one, representing the fraction of ten-second observations (out of 20 such observations) during which a behavior was recorded. Dialogic reading is the number of intervals in which the caregiver engaged in dialogic reading behaviors. Mother focused on reading is the fraction of intervals in which the caregiver was not distracted. Baseline data were available for all measures except "Reading frequency." For other outcomes, the gap is calculated as the difference between baseline outcomes for children of literate and illiterate caregivers. For "Reading frequency," the gap is calculated as the difference between endline outcomes for children of literate and illiterate caregivers within the control group only.

Table 1. Summary Statistics on Caregivers and Children Located at Follow-Up

	Obs.	Mean	S.D.	Min.	Max.
Child Characteristics:					
Child age in months	442	54.77	17.45	24	83
Child is male	442	0.52	0.50	0	1
Height-for-age z-score	429	-0.41	1.42	-4.45	4.07
Child is stunted	429	0.12	0.32	0	1
Child is enrolled in school	363	0.74	0.44	0	1
Expressive vocabulary (out of 20)	442	5.82	5.35	0	20
Receptive vocabulary in English (out of 34)	442	6.68	4.32	0	18
Receptive vocabulary in Luo (out of 27)	442	8.91	6.11	0	24
Receptive vocabulary in Swahili (out of 31)	442	8.97	5.22	0	28
Primary Caregiver Characteristics:					
Caregiver is child's mother	309	0.84	0.36	0	1
Caregiver is child's grandmother	309	0.12	0.32	0	1
Caregiver is not mother or grandmother	309	0.14	0.35	0	1
Caregiver education: no formal schooling	309	0.02	0.15	0	1
Caregiver education: incomplete primary school	309	0.49	0.50	0	1
Caregiver education: completed primary, not secondary	309	0.40	0.49	0	1
Caregiver education: completed secondary school	309	0.09	0.28	0	1
Caregiver illiterate	309	0.28	0.45	0	1
Caregiver depressive symptoms (out of 26)	302	10.36	5.27	0	26
Household Characteristics:					
Any children's books in the home	309	0.14	0.34	0	1
Someone has read to the child in past 3 days	309	0.50	0.50	0	1
Family Care Indicators score (out of 18)	309	8.13	3.54	0	16
Household size	309	3.04	0.85	2	6
Rooms per person	309	0.88	0.43	0.20	3
Household has cement floor	309	0.23	0.42	0	1
Household has iron roof	309	0.96	0.19	0	1
Household has electricity	309	0.31	0.46	0	1
Household has latrine	309	0.70	0.46	0	1

Notes. Summary statistics on 309 caregivers and 442 children for whom follow-up data is available. Summary statistics on the entire sample of 357 caregivers and 510 children are similar, and are available upon request. Baseline height data is missing for 13 children. Baseline data on school enrollment is missing for 93 children. Seven caregivers declined to answer the questions on depressive symptoms.

Table 2. Regression Analysis of the Impacts of Treatment on Primary Outcomes

	Someone in Household Read to Child Past 3 Days		**Caregiver Reading Frequency (Times/Week)**		**Storybook Comprehension Index (z-score)**		**Storybook Expressive Vocabulary (z-score)**		**Dialogic Reading Behaviors Observed**		**Times Caregiver Was Distracted while Reading**	
Treatment	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
T1: storybooks	0.201	<0.001	0.464	0.233	0.978	<0.001	0.152	0.228	-0.653	0.408	-0.346	0.628
	(0.059)		(0.388)		(0.137)		(0.126)		(0.787)		(0.713)	
T2: training	0.223	<0.001	1.426	<0.001	0.924	<0.001	0.223	0.089	1.923	0.008	-2.021	0.005
	(0.061)		(0.423)		(0.125)		(0.131)		(0.722)		(0.715)	
T3: booster	0.243	<0.001	1.497	<0.001	1.119	<0.001	0.292	0.027	1.696	0.031	-1.463	0.047
	(0.055)		(0.361)		(0.134)		(0.132)		(0.782)		(0.733)	
T4: home visit	0.262	<0.001	2.069	<0.001	1.194	<0.001	0.239	0.107	2.437	0.002	-2.268	0.003
	(0.058)		(0.398)		(0.162)		(0.148)		(0.782)		(0.757)	
T1 = T2		0.669		0.019		0.716		0.571		<0.001		0.007
T2 = T3		0.646		0.863		0.204		0.604		0.749		0.399
T3 = T4		0.625		0.145		0.696		0.725		0.328		0.227
Control mean	0.697		1.889		-0.680		-0.156		14.897		4.103	

Notes. OLS regression results reported. All specifications control for community effects, child age and gender, an indicator for having the mother as the primary caregiver, household size, household wealth (a durable asset index), and caregiver depressive symptoms. Columns 1, 4, 5, and 6 also include baseline values of the outcome variable (baseline data on reading frequency and storybook comprehension was not collected). Regressions estimated using reg command in Stata 14.2.

Table A1. Summary Statistics by Experimental Treatment

	Control		Storybooks		+ Training		+ Booster		+ Home Visit	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Child Characteristics:										
Child age in months	50.89	17.08	55.77	17.71	54.92	16.65	57.01	17.77	55.95	17.80
Child is male	0.52	0.50	0.52	0.50	0.53	0.50	0.51	0.50	0.53	0.50
Height-for-age z-score	-0.28	1.56	-0.49	1.52	-0.35	1.41	-0.32	1.21	-0.63	1.36
Child is enrolled in school	0.68	0.47	0.79	0.41	0.72	0.45	0.80	0.41	0.74	0.44
Expressive vocabulary (out of 20)	4.97	4.84	6.72	5.76	5.25	5.10	6.56	5.76	5.69	5.13
Receptive vocabulary in English (out of 34)	6.34	4.45	6.74	4.06	6.78	4.25	6.80	4.40	6.78	4.52
Receptive vocabulary in Luo (out of 27)	8.03	6.01	9.57	6.43	8.40	5.80	9.72	6.38	8.97	5.84
Receptive vocabulary in Swahili (out of 31)	8.41	5.00	9.40	5.94	8.43	5.20	9.38	4.87	9.32	4.95
Primary Caregiver Characteristics:										
Caregiver is child's mother	0.81	0.40	0.77	0.42	0.93	0.25	0.85	0.36	0.85	0.36
Caregiver is child's grandmother	0.12	0.33	0.16	0.37	0.06	0.23	0.11	0.31	0.14	0.35
Caregiver is not mother or grandmother	0.18	0.39	0.16	0.37	0.07	0.25	0.15	0.36	0.15	0.36
Caregiver years of education	7.52	2.33	7.09	2.89	7.02	2.84	6.84	2.45	7.19	2.09
Caregiver illiterate	0.29	0.46	0.30	0.46	0.26	0.44	0.29	0.46	0.26	0.44
Caregiver depressive symptoms (out of 26)	11.10	5.59	9.29	4.45	9.69	5.49	10.63	4.80	9.55	5.14
Household Characteristics:										
Household size	3.55	1.02	3.17	0.83	3.20	0.92	3.20	0.91	3.26	0.78
Number of children's books in the home	0.07	0.29	0.35	0.90	0.30	0.90	0.28	0.76	0.25	0.80
Someone has read to the child in past 3 days	0.55	0.50	0.49	0.50	0.51	0.50	0.56	0.50	0.47	0.50
Family Care Indicators score (out of 17)	8.14	3.54	7.77	3.67	8.03	3.44	8.78	3.68	8.41	3.69
Household wealth index	-0.08	2.51	0.17	2.24	-0.30	1.69	0.56	2.47	-0.50	1.82

Summary statistics on 309 caregivers and 442 children for whom follow-up data is available. Summary statistics on the entire sample of 357 caregivers and 510 children are similar, and are available upon request. Baseline height data is missing for 13 children; baseline school enrollment data is missing for 93 children. Seven caregivers declined to answer the questions on depressive symptoms.

Table A2. Tests of Equality of Baseline Characteristics Across Treatment Arms

	F-test	**Pairwise Tests of Equality**						
	(p-value)	C vs. T1	C vs. T2	C vs. T3	C vs. T4	T1 vs. T2	T2 vs. T3	T3 vs. T4
Child age in months	0.14							
Child is male	1.00							
Height-for-age z-score	0.49							
Child is enrolled in school	0.46							
Expressive vocabulary	0.10	0.04	0.69	0.05	0.34	0.08	0.10	0.30
Receptive vocabulary in English	0.94							
Receptive vocabulary in Luo	0.27							
Receptive vocabulary in Swahili	0.47							
Caregiver is child's mother	0.05	0.67	0.04	0.49	0.56	0.02	0.13	0.95
Caregiver is child's grandmother	0.27							
Caregiver is not mother, grandmother	0.24							
Caregiver years of education	0.47							
Caregiver illiterate	0.96							
Caregiver depressive symptoms	0.08	0.05	0.19	0.64	0.14	0.64	0.32	0.25
Household size	0.02	0.05	0.10	0.09	0.15	0.85	0.96	0.71
Number of children's books in home	0.10							
Someone read to child in past 3 days	0.79							
Family Care Indicators score	0.42							
Household wealth index	0.02	0.65	0.66	0.25	0.41	0.21	0.03	0.01

Summary statistics on 309 caregivers and 442 children for whom follow-up data is available. Summary statistics on the entire sample of 357 caregivers and 510 children are similar, and are available upon request. Baseline height data is missing for 13 children; baseline school enrollment data is missing for 93 children. Seven caregivers declined to answer the questions on depressive symptoms. P-value from an F-test of the joint significance of treatments reported in the first column, calculated using anova command in Stata 14.2. Remaining columns report Wald tests calculated using the reg and test commands in Stata 14.2.

Table A3. OLS Regressions of the Likelihood of Attrition – by Treatment Arm

	OLS Regression Results		
	Coef.	S.E.	p-value
Storybooks only treatment	0.006	0.043	0.886
Dialogic reading treatment	0.003	0.043	0.945
Booster training treatment	0.064	0.048	0.188
Home visit treatment	0.059	0.048	0.226
Constant	0.108	0.030	0.000

OLS regressions estimated using Stata 14.2. Dependent variable is an indicator for attriting from the sample (i.e. completing the baseline but not the endline survey).

Table A4. OLS Regressions of the Likelihood of Attrition – by Baseline Characteristics

	OLS Regression Results		
	Coef.	S.E.	p-value
Child age in months	0.001	0.001	0.700
Child is male	0.016	0.031	0.592
Height-for-age z-score	0.007	0.013	0.568
Child is enrolled in school	-0.006	0.044	0.894
Expressive vocabulary (out of 20)	0.005	0.005	0.277
Receptive vocabulary in English (out of 34)	-0.009	0.004	0.039
Receptive vocabulary in Luo (out of 27)	-0.004	0.004	0.351
Receptive vocabulary in Swahili (out of 31)	-0.002	0.004	0.553
Caregiver is child's mother	-0.078	0.142	0.585
Caregiver is child's grandmother	-0.145	0.114	0.205
Caregiver is not mother or grandmother	-0.024	0.180	0.894
Caregiver years of education	-0.002	0.007	0.804
Caregiver illiterate	-0.032	0.038	0.404
Caregiver depressive symptoms (out of 26)	-0.000	0.003	0.973
Household size	-0.015	0.017	0.353
Any children's books in the home	-0.024	0.047	0.611
Someone has read to the child in past 3 days	-0.026	0.036	0.457
Family Care Indicators score (out of 18)	-0.008	0.006	0.174
Household wealth index	0.008	0.008	0.302
Some baseline data missing	0.014	0.049	0.778
Constant	0.416	0.179	0.020

OLS regressions estimated using Stata 14.2. Dependent variable is an indicator for attriting from the sample (i.e. completing the baseline but not the endline survey).

Table A5. Treatment Effects on Vocabulary

	Expressive Vocabulary		**Receptive Vocabulary in…**					
			English		**Luo**		**Swahili**	
Treatment	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
T1: storybooks	0.034	0.711	-0.070	0.659	-0.027	0.827	-0.070	0.610
	(0.093)		(0.158)		(0.125)		(0.137)	
T2: training	0.096	0.288	0.047	0.778	-0.073	0.569	-0.063	0.612
	(0.090)		(0.168)		(0.129)		(0.124)	
T3: booster	0.019	0.844	-0.009	0.962	0.083	0.523	0.189	0.186
	(0.098)		(0.180)		(0.130)		(0.142)	
T4: home visit	-0.019	0.833	0.124	0.472	-0.017	0.907	0.048	0.709
	(0.088)		(0.173)		(0.142)		(0.129)	
T1 = T2		0.496		0.372		0.702		0.953
T2 = T3		0.438		0.718		0.202		0.051
T3 = T4		0.696		0.406		0.462		0.316
Control mean	-0.030		0.005		0.041		0.016	

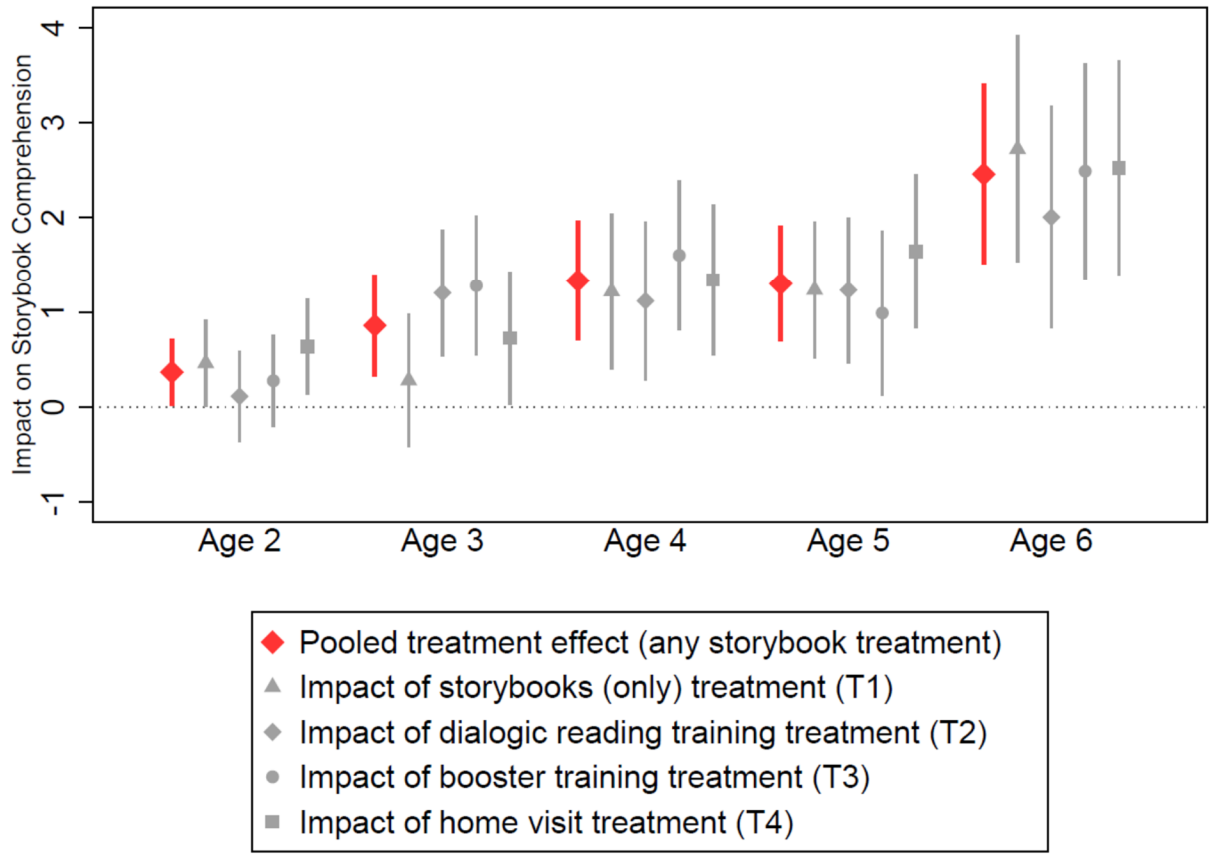
OLS regression results reported. All specifications control for community effects, child age and gender, an indicator for having the mother as the primary caregiver, household size, household wealth (a durable asset index), caregiver depressive symptoms, and baseline values of the outcome variable. Regressions estimated using reg command in Stata 14.2.

Table A6. Treatment Effect Heterogeneity by Caregiver Literacy

Dependent Variable:	**Read to Child Past 3 Days**		**Reading Frequency**		**Storybook Comprehension**		**Storybook Expressive**		**Dialogic Reading Behaviors**		**Distracted while Reading**	
Treatment	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
T1: storybooks												
T1 x Illiterate	0.329	0.020	0.738	0.161	0.737	0.000	0.239	0.165	1.303	0.402	-2.499	0.080
T1 x Literate	0.148	0.024	0.399	0.372	1.091	0.000	0.117	0.463	-1.345	0.104	0.489	0.521
Test of equality	0.272		0.611		0.130		0.583		0.113		0.052	
T2: training												
T2 x Illiterate	0.366	0.008	1.528	0.009	0.976	0.000	0.340	0.157	0.558	0.761	-0.541	0.781
T2 x Literate	0.159	0.018	1.345	0.007	0.900	0.000	0.173	0.264	2.165	0.001	-2.275	0.000
Test of equality	0.196		0.809		0.766		0.549		0.386		0.375	
T3: booster												
T3 x Illiterate	0.405	0.002	3.005	0.000	1.287	0.000	0.457	0.018	3.462	0.040	-2.672	0.131
T3 x Literate	0.174	0.009	0.866	0.037	1.052	0.000	0.224	0.177	1.019	0.193	-0.934	0.190
Test of equality	0.142		0.010		0.352		0.343		0.166		0.348	
T4: home visit												
T4 x lliterate	0.513	0.000	2.698	0.000	1.154	0.000	0.609	0.001	3.805	0.015	-3.095	0.055
T4 x Literate	0.160	0.018	1.772	0.000	1.198	0.000	0.098	0.600	1.898	0.016	-1.892	0.012
Test of equality	0.015		0.266		0.879		0.046		0.239		0.476	
CG illiterate	-0.250	0.073	-1.515	0.000	-0.200	0.063	-0.194	0.177	-2.532	0.016	2.427	0.049
Control group mean	0.697		1.889		-0.680		-0.156		14.897		4.103	
N	431		431		431		431		292		292	

OLS regression results reported. All specifications control for community effects, child age and gender, an indicator for having the mother as the primary caregiver, household size, household wealth (a durable asset index), and caregiver depressive symptoms. Columns 1 and 3 through 6 also include baseline values of the outcome variable. Regressions estimated using reg command in Stata 14.2.

Figure A1. Treatment Effects on Storybook Comprehension by Age



Notes. In Figure A1, each marker (circle, square, triangle, or diamond) indicates the point estimate for a treatment effect, while the associated 95-percent confidence interval is represented by the line going through the marker. Treatment effects are separated by age. Effects are shown first, in red, with all four treatment arms (T1 through T4) aggregated and analyzed together in a regression specification comparing outcomes in any treatment group to outcomes in the comparison group (the "Pooled treatment effect"), then are shown separately by treatment arm in gray. The outcome, "Storybook Comprehension," is not age-normalized in this figure; instead, it is a measure ranging from zero to six: children are asked between one and three questions about the content of each of the six storybooks; the fraction of questions they answer correctly about each book is summed, so that comprehension of each book contributes equally (up to one point) on this outcome scale.