Informing Mothers about the Benefits of Conversing with Infants: Experimental Evidence from Ghana*

Pascaline Dupas, Camille Falezan, Seema Jayachandran, and Mark Walsh

January 31, 2024

Abstract

Despite the well-established importance of verbal engagement for infant language and cognitive development, many parents in low-income contexts do not converse with their infants regularly. This paper reports on a randomized field experiment evaluating a low-cost intervention designed to boost verbal engagement with infants. The intervention entails showing recent or expectant mothers a 3-minute informational video and providing them with a themed wall calendar. Six to eight months later, mothers who participated reported a stronger belief in the benefits of verbally engaging with infants, more frequent parent-infant conversations, and more advanced language and communication skills of their infants. Treatment effects on objective measures of parent-child conversation (from a recording device) and infant language and cognitive skills (from surveyors' observations) were statistically insignificant but consistently positive. We find larger effects on objectively measured parent-child conversation immediately after the intervention, suggesting scope for a larger long-term effect had the behavior change stuck more. The intervention's potential for low-cost implementation via health clinics makes it a promising strategy for early childhood development in low-income contexts, particularly if complemented by efforts to support habit formation.

JEL: C93, D19, I21

Keywords: early childhood development; infant-directed speech; human capital; information intervention

*The authors are from Princeton University, NBER, and CEPR (pdupas@princeton.edu), MIT (cfalezan@mit.edu), Princeton University (jayachandran@princeton.edu), and Stanford University (mwalsh24@stanford.edu). Acknowledgments: This research was supported by a grant from the World Bank's Strategic Impact Evaluation Fund. It builds on pilot work made possible by funding from the Weiss Fund. We gratefully acknowledge their support. These funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank IPA Ghana for their partnership. The research protocol was approved by the Ghana Health Service Ethical Review Committee and by the Stanford University IRB. We registered the RCT in the AEA RCT Registry (ID AEARCTR-0007161) and at ClinicalTrials.gov (ID NCT04807907) (Dupas, Jayachandran, and Walsh, 2021).

1 Introduction

While parents universally use "baby talk" to soothe an infant or get their attention, engaging in a richer form of infant-directed speech (IDS)—that includes responding to their infant's gestures and babbles and talking to them in complete, if simplified, sentences, using a wide variety of words—varies by socioeconomic status (SES) within societies (Hart and Risley, 1995; Hoff, 2003) and across societies (Farran et al., 2016). Given the benefits of parent-infant conversations for cognitive development (Monnot, 1999; Weisleder and Fernald, 2013), these SES gaps are likely to compound the disadvantages that children in poorer families face. The problem may be especially acute in lower-income countries where 43% of children under 5 years (over 250 million children) are at risk of not reaching their developmental potential (Black et al., 2017).

One explanation for parental under-investment in conversing with infants is inaccurately low expectations about the benefits. A large body of literature in the US has shown that lower-SES parents are less aware of the returns to verbal engagement with infants (List et al., 2021). In the Northern region of Ghana, our study setting, only 11% of mothers reported that parents should start talking to their baby at birth,[1] and 61% reported that they should begin talking to the child once he or she is older than six months. While one might have hoped that rising educational attainment in low- and middle-income countries (LMICs) would narrow this gap, Duflo et al. (2024) find that a subsidy program that increased secondary school completion in Ghana did not change maternal beliefs about the importance of conversing with infants. The persistence of misperceptions is perhaps unsurprising, as it is not intuitive that speaking to a 1-week or even 3-month-old baby boosts language skills and cognitive development. Young infants are not noticeably responsive to language, and the benefits materialize later, so talking to babies might not be a practice that arises organically, but only by parents explicitly being taught its value. If this explanation is correct, a cheap

[1]This is lower than in other contexts. For example, in urban areas of Turkey, 50% of mothers reported that one should begin talking to their child at birth (Ertem et al., 2007). Other studies have found low levels of caregiver knowledge in other low-and-middle-income countries such as Morocco (Zellman et al., 2014) and Nepal (Shrestha et al., 2019).

information intervention might be enough to correct parental beliefs, cause behavior change, and cost-effectively enhance infant development outcomes.

In this paper, we report on the effects of a cheap, scalable intervention designed to inform mothers about the benefits of conversing with infants. The intervention consists of showing the recent or expectant mother a 3-minute video about parent-infant conversations and giving her a wall calendar with visual reminders of the video's message. The video is a simple animation with a voice-over describing the value of parent-infant conversations and encouraging the viewer to speak to her baby and to tell family members to do so too. The purpose of the calendar (see Figure A.1) is to (1) act as a reminder of the message, keeping it salient, (2) facilitate common knowledge among household members about these lessons, and (3) provide a method of forming a parent-infant conversation habit (the treatment respondents were encouraged to fill in the stars next to each week on the calendar if they conversed with their infant each day that week). The video was shown and a calendar handed out to women visiting local government health clinics for prenatal or postnatal checkups.

To evaluate this intervention, we conducted a randomized controlled trial (RCT) in which we delivered the intervention to 705 randomly selected Northern Ghanaian women from a sample of 1,408 who were pregnant or had a young infant. We use data from a follow-up survey conducted 6 to 8 months later to estimate the impacts of the intervention on maternal beliefs about the benefits of parent-infant conversations, self-reported parental verbal inputs, and mother-reported early childhood development outcomes. Mothers who received the intervention report greater belief in the benefits of conversing with infants, more verbal engagement behaviors, and that their infants have larger vocabularies and more advanced gestural skills. The magnitudes of these reported effects are consistently about 0.1 standard deviation.

To address concerns about experimenter demand effects, we also collected observed measures of child cognitive development and parent-child conversation. For child cognitive development, we asked surveyors to record children's behavioral responses to actions or verbal prompts, using two problem-solving questions adapted from the Ages & Stages Questionnaire (ASQ) and seven questions adapted from the Oxford Neurodevelopment Assessment (Ox-NDA). While the OX-NDA was originally designed for 10-14 month olds, we modified it for our sample of 2 to 18 month olds (Fernandes, 2021).[2] In addition, surveyors observed and recorded whether the infant babbled at some point during the survey. To measure parental verbal inputs objectively and to provide another objective measure of child vocalizations, we used Language ENvironment Analysis (LENA) recording devices to collect day-long recordings of the child's auditory environment (for only half of our sample, due to budget constraints). We find positive point estimates across all of these outcomes, but they are mostly not statistically significant.

Given the 'light-touch' nature of the intervention, we planned for the possibility that mothers might initially change their behavior but revert to their pre-intervention behavior by the time of the follow-up survey (6 to 8 months later). To distinguish between participants never embracing the recommended practices versus adopting them initially but not persisting, we also administered our informational video treatment to a subset of control-group respondents the morning after they completed their day-long endline LENA recording. We refer to this as the 'endline intervention' to distinguish it from the main intervention described above that was delivered at baseline. For these respondents, we recorded the child's auditory environment again over the next day. To estimate the effect of the endline intervention, we compare the LENA measures for these respondents from *the day before* receiving the intervention to the same measures *the day after*. The newly-informed mothers speak 1.4 more words per minute (≈8.4% of the mean, p=0.058) post-intervention relative to their pre-intervention levels. This effect is eight times larger than the main intervention effect estimated after 6-8 months. The large immediate impact of the endline intervention on parental behavior shows that mothers are willing and able to verbally engage with their children when (1) they are told that they should, and (2) they know their mother-child

[2] We could not identify a cognitive development tool designed for our younger age range that was not based on parental reports and could be implemented by survey staff at the homes of our Northern Ghanaian respondents.

interactions are being recorded. This suggests that there is no 'technological barrier' to verbal engagement with infants: once they know they should do it, mothers know *how* to do it. But the fact that treatment effects of the main intervention are much lower after 6-8 months suggests that sustaining the behavior over time is difficult.

When we ask mothers who received either the main intervention or endline intervention about likely barriers to parent-infant conversations, the most common answers are fear of social sanctions (scorn) and difficulty in habit formation. The relative importance of these two perceived barriers vary by time since the intervention delivery. Mothers who received the main intervention at baseline (i.e., 6-8 months prior to being asked about barriers) are 39% less likely to report social scorn or mockery as the main barrier (p=0.002), compared to mothers who received the intervention at endline (i.e., less than 24 hours prior to being asked about barriers). In contrast, habit formation is equally likely to be cited as the main barrier across the two groups. One interpretation of this pattern is that mothers quickly get over the social awkwardness of verbally engaging their infants, while transforming a new behavior into a sticky habit is fundamentally difficult (Rothman et al., 2015; Webb and Sheeran, 2006; Lally and Gardner, 2013).

Recent meta-analyses have already shown that there is strong evidence that interventions encouraging 'responsive caregiving' (which includes parent-infant conversations) promote maternal knowledge and mother-infant interactions, but our intervention is cheaper and lighter-touch than any of the studies included in recent meta-analyses (Jeong et al., 2018, 2021; Verguet et al., 2022). In the most thorough recent meta-analysis (Jeong et al., 2021), almost all (67) of the 70 responsive-caregiving interventions required multiple visits or sessions with a skilled trainer. The closest studies to ours are Suskind et al. (2018) and List et al. (2021)'s experiments in the Chicago metropolitan area in the United States. Suskind et al. (2018) finds significant effects on parental beliefs from mothers watching a 10-minute video but does not measure parental behavior or child development. List et al. (2021) evaluates the effect of mothers watching 10-minute videos when their child is 1, 2, 4, and 6 months old and measures persistent effects on beliefs, short-run effects on parental verbal inputs, but noisy null effects on mother-reported vocabulary. Our study tests an intervention that is significantly shorter and finds positive effects on infant language and cognitive development, which were unmeasured outcomes in Suskind et al. (2018) and may have been undetected in List et al. (2021) due to a lack of statistical power.[3]

There is less evidence on parenting interventions in low- and middle-income countries (LMICs), but the existing evidence is promising. Jeong et al. (2021) estimates that parenting interventions have 3-4 times larger effects in LMICs compared to high-income countries. Almost all of the rigorously-evaluated programs in LMICs are home-visiting programs or comprehensive village-level initiatives with regular group meetings. These types of resource-intensive interventions may not be scalable for budget-constrained LMICs.

We estimate that the cost per child beneficiary in our research trial was $3.01. At scale, one could use existing health center staff rather than surveyors to hand out the calendar and show the video, lowering the cost to $0.45 per child. The low cost, combined with the treatment effects we estimate, implies that the intervention could be among the most cost-effective known ways for LMICs to increase infant language development. We calculate that, at scale, it would deliver a 1 standard deviation (SD) improvement in a child's cognitive or language development for $4.52 to $10.01, depending on whether we use our reported or observed measures of child development (or $30 to $67 under our research trial conditions). In Verguet et al. (2022)'s meta-analysis of 12 early childhood interventions, the median intervention costs $329 per SD improvement in child cognitive or language development.[4] Even if the true effect of our intervention, or the effect achievable at scale, were a quarter of the magnitude we estimated (so the cost becomes $18.09-$40.02 per SD improvement), the intervention would still be more cost-effective than 11 of the 12 interventions assessed by Verguet et al. (2022).

One reason that our results could overstate the effect of the scaled-up intervention is if

[3] List et al. (2021)'s sample size is 475 compared to our sample size of 1,408.

[4] We convert Verguet et al. (2022)'s estimates to 2021 USD to facilitate comparisons. Among the 12 interventions they assess, 10 cost over $50 per child and the cheapest costs $22.32 per child.

compliance fades out over time, or alternatively, the mother-reported results might be subject to experimenter demand effects. Future research could use larger sample sizes or more precise objective measurements to better understand the effects of light-touch IDS interventions. Additionally, our analysis of mechanisms suggests that supplementing the intervention with habit formation support could increase the effectiveness.

2 Study Design

2.1 Sampling and intervention

We received approval from the Ghana Health Service, which is the government agency overseeing health clinics, to survey prenatal and postnatal patients in 10 of the public health clinics around the city of Tamale in early 2021 (see Table A.1 for the list of facilities). Tamale is the third-largest city in Ghana and the largest city in the Northern region of Ghana, which is poorer than the rest of Ghana.[5]

In March 2021, we employed a team of surveyors from Innovations for Poverty Action (IPA) Ghana to enroll a sample of prenatal and postnatal patients from the health clinics. Surveyors approached patients before/after their prenatal or postnatal clinic visits and, if the patients were interested, screened them for eligibility. In order to participate, women had to (1) be aged 18-40 years old, (2) have an infant or be pregnant with a child who would be 2-18 months six months later at the time of the follow-up survey, and (3) speak English or Dagbani (the main language in Tamale).[6] We aimed to survey 1,400 women and ended up surveying slightly more, 1,408.

Half of respondents were randomly allocated to the treatment group and selected to watch a 3-minute intervention video (see `https://www.facebook.com/ghanababytalk`) and receive

[5] The average monthly household income in the Northern region is $\sim$ \$38, while the national average is $\sim$ \$156.

[6] Of the 1,765 women approached, 1,462 were eligible. 17 were ineligible because of their age, 283 were ineligible because of their child's age or due date, and 3 were ineligible because they did not speak English or Dagbani. Of the 1,462 eligible women, 1,408 completed the survey and were administered the intervention. One did not pass the COVID symptom screening, 15 refused to participate, and 38 chose not to participate partway through the baseline survey.

the intervention calendar at the end of the baseline survey (see Figure A.1). The narrator of the video (which was available in English or Dagbani) conveys information about the benefits of verbal engagement with infants. Examples of the information in the video include that conversing with infants helps them learn even if they are “too young to talk themselves”, that infants learn more from “hearing words and sentences directed at them”, and that “back-and-forth moments” are particularly important for child development. The video then provides a few ideas for how to converse with your infant such as: describing what you see “when you are walking across the village or town with her”, telling your baby what you are preparing “when you are cooking”, “telling stories”, “singing songs”, or describing pictures in/“reading books”. This narration is paired with images of family members talking to an infant while doing the suggested activities. In short, the video informs mothers about the benefits of verbal engagement with infants and about how to verbally engage an infant. The intervention calendar highlights a few key points from the video, displays images from the intervention video, and has hollow stars next to each week that respondents were encouraged to color in if they talked to their infant at least once a day during that week.[7]

The remaining 50% of respondents form the control group. They did not watch the video, and they received a calendar with a picture of Stanford University (see Figure A.2).

We implemented the intervention at the public health clinics after the patients' prenatal or postnatal visit, which mirrors how we expect this intervention would be implemented at scale.[8] To enable within-clinic randomization, we had surveyors show the video on a tablet to individual mothers, but the intervention could be even cheaper at scale if existing clinic staff show individual patients the video or the video is shown to a group of patients, perhaps on a television monitor in the waiting room.[9]

[7]The calendar also included a link to the webpage with the video. There were 26 >3-second viewers of the Dagbani version of the video during the study period and 10 >60-second viewers. The low usage of the webpage could be due to the internet data charges to stream a video being expensive for this population.

[8]We partnered with officials in the Ghana Health Service who agreed that this was a reasonable expectation.

[9]In the latter scenario, one would need to ensure that the one-on-one engagement of the surveyor and the mother was not a key mechanism for the treatment effects. Unfortunately, our experiment cannot speak to this.

2.2 Sample characteristics and baseline behavior

Table 1 presents baseline characteristics for our sample. Reassuringly, only 1 of the variables in the table is significantly different at the 5% level between the treatment and control group, and the joint test does not reject the null of no significant differences between treatment and control (p=0.662).[10]

In our sample, nearly all women are married, with almost a third in polygamous unions. Nearly two thirds have at least a primary school education, and 61% can read in English or Dagbani. The average respondent is 28 years old, lives in a household of nine, and has two children. At baseline, although 89% of women had children, only 61% had an eligible child already born, while the remaining 39% were currently pregnant.

As expected, and consistent with the qualitative background research that led us to conceive this study, baseline knowledge about the role of verbal engagement in early childhood development is limited. Table 2 presents baseline IDS beliefs and behavior for our sample. On average, respondents report parents should start talking to their baby at 11 months, but only in full sentences when the child is 2 years old, which is a few months after the age at which respondents believe children start saying meaningful words themselves (20 months). These reports demonstrate that the beliefs of many women in our sample diverge from evidence-supported practices for enhancing child development such as extensively conversing with newborn infants.

2.3 Endline measurements

We conducted the endline activities from September to December 2021, on average 6.4 months after the intervention. The endline consisted of a main survey conducted in person, at the home of the respondent, and, for a subset of respondents, one or two day-long LENA recording activities followed by short LENA-debrief surveys.

[10] We also cannot reject the null of the joint test for the baseline variables presented in Table 1 providing additional evidence that the randomization was implemented correctly.

Survey. We completed interviews with 89% of respondents, with no differential attrition between the treatment and the control group (see Table A.2). The endline survey measured parental beliefs about verbal engagement with infants, parental verbal inputs to the child, child language development, child gestural communication, recall of the treatment, and perceived barriers to IDS (see Sections 2.5 and 2.6 for details on these measures). In addition, to have a measure of child development that is not subject to experimenter demand effects, the endline survey included some direct measurements of child development outcomes as observed by the surveyor.

We describe the measurement tools used and how we combine them for analysis in section 2.6. All survey outcomes were collected in the main endline visit except for those on treatment recall and perceived barriers to IDS, which were measured at the very end of all endline-related activities (i.e., post-LENA measurements when applicable) for a given respondent.

LENA measurements. As an observed measure of parental verbal inputs, we gathered a day-long recording of parent-child interactions through the LENA system.[11] For the LENA, the child wears a specially-designed shirt with an attached recording device for at least 8 hours for one day. The device records all sounds produced around the child and the data are then processed using a speech recognition software to generate count-based metrics of words spoken by female adults and male adults to the child, child vocalizations, and conversational turns between the child and adults. A separate set of surveyors was tasked with dropping and picking up LENA devices at respondents' homes in the days following the endline survey. On average, respondents completed the LENA-activities 16 days after the main endline survey.

The LENA surveyors visited respondents before 10 a.m. on the day of the recording activity to give mothers the shirt with the LENA device, answer any questions about the device and/or instructions, and stayed as mothers dressed the child with the shirt. Surveyors asked

[11] This device was validated in Ghana by the Harvard Laboratory for Developmental Studies. See Appendix Section B for more information.

mothers to have the child wear the shirt until the next day. 97% of mothers who completed the endline survey consented after the surveyor described the LENA process. We restrict the LENA analysis to audio-data collected from 10 a.m. to 7 p.m. with no interruption (9 hours of recording).[12] The same surveyor came back the next day to pick up the LENA device and conduct a short survey on the respondent's experience with the LENA, barriers to conversing with babies, and (for treatment respondents) recall of the main intervention. Given the cost of the LENA devices and the LENA pickup surveyors, we could only afford to use the LENA measurements with 900 respondents. We randomly chose 900 respondents from our full sample, stratified by treatment status.[13] We obtained 785 LENA recordings (see Table A.2 for details on survey and LENA-activities participation rates).

2.4 Endline intervention

Had we found null effects at endline, it would have been important to understand if participants never adopted the recommended practices, or adopted them initially but then stopped. It is also possible the treatment effects could have grown over time, as participants gained experience and comfort with IDS. Thus, to assess how the immediate effect of the intervention differs from the 6-8 month effect, we also compare the child's auditory environment *the day before* their mother watches the video to the *the day after* through our 'endline intervention.'

We use the LENA to measure the effect of the endline intervention. Relative to our other options (self-reports or direct observation by the surveyor), the LENA should engender less of an experimenter demand effect. Moreover, a positive treatment effect on IDS that is driven by experimenter demand still enables us to rule out the existence of a 'technological'

[12] This 9-hour time window ensures we have comparable data for all observations as households received the LENA device between 6 and 10 a.m. depending on when the LENA surveyor arrived. The LENA device could record 16 hours of audio, but after 7 p.m. a few LENA devices turned off (either because they ran out of battery, or households turned them off (purposely or not) to bath the child). 99% of recordings had 9 interrupted hours of audio and were kept in the analysis.

[13] We originally sampled 900 respondents, but discovered at endline that one respondent had been interviewed twice at baseline. The respondent also appeared twice in the sample selected for the LENA-activities, so the final sample for the LENA was composed of 899 individuals (N=450 from the treatment group, N=449 from the control group).

barrier or contextual factor that prevents mothers from talking to their infant even when they would like to do so.

We selected participants for the endline intervention from among the 450 control group respondents slated to provide a day-long LENA recording at endline (see Section 2.3 for details on how respondents were selected for the LENA subsample). This sampling frame allows us to use their first LENA recording as a pre-endline-intervention measure of the child's auditory environment. For budget reasons, we chose 225 of these women for the endline intervention, randomly selecting them with stratification by child age and baseline self-reports of behavior.[14]

After we collected their first LENA recording, endline intervention participants (unexpectedly) were shown the intervention video and asked to record for a second day.[15] 99% of respondents who had consented to a first recording agreed to keep the device for a second day. After a day, the surveyor picked up the device and administered a short debrief survey and the last set of questions on barriers to conversing with babies.[16] We compare the LENA data collected from endline intervention participants one day before and one day after the video was shown to them to estimate the short-run effects of watching the video. That is, we estimate the effects using a before-after design applied to the endline intervention sample, not by comparing them to a control group.[17]

Our approach to measuring short-run effects—by delivering the intervention at endline to

[14] We stratified by whether the focal child would be under or over 1 month old at baseline, and for households with a focal child over 1 month old at baseline, whether they scored above or below median on the baseline self-reported mother behavior index. Of the 225 selected households, 24 were ineligible or not available at endline. 6 refused to participate in the follow-up survey or LENA-activity, and 195 participated in the first day of LENA recording. 193 of those then received the endline intervention and participated in a second day of recording (see Table A.2 for further details).

[15] When we picked up the first recording, they completed the debrief survey but were not asked questions on barriers to conversing with babies.

[16] 112 of the 225 endline intervention participants were also sampled to watch a 1-minute video of a Ghanaian mother verbally engaging with their infant just prior to watching the intervention video. After this video, respondents answered questions about what they or others would think of the mother in the video. These questions allow us to assess social norms around infant-conversation practices.

[17] For cost reasons, we did not also collect a second day of LENA recordings for the "pure control" group. For use of the LENA, we did not expect any learning effects between the two days of LENA use that might bias the before-after analysis.

some control households—could potentially be useful in other studies. There are at least two advantages over the standard approach of delivering the intervention to a single treatment group and then measuring outcomes twice, once in the short run and again in the longer run. First, in our approach, the environment is held constant (e.g., same economic conditions) when the short-run and long-run outcomes are measured, because the measures are collected at the same calendar time. This prevents a confound such as the endline measurement occurring during the lean season and treatment effects being smaller in the lean season, which could cause treatment effects to only be observed in the short run. With the standard approach, environment-contingent treatment effects like this could be misinterpreted as fade-out. Second, the approach reduces study costs if outcome measurement has a fixed cost component (e.g., to train a team of surveyors on how to deliver the LENA device to study participants). Since the short- and long-run measurements occur simultaneously, fixed costs are incurred only once. There are also drawbacks, such as the short-run effects being estimated from children who are older than those who identify the long-run effects.[18]

2.5 Treatment recall, social norms, and perceived barriers

To avoid inducing experimenter demand effects among our respondents, we did not mention the treatment or discuss barriers to parent-infant conversations until after all other endline measures for a given respondent had been gathered. Respondents who were not selected to use the LENA device answered questions on these topics at the end of the endline survey, while the LENA subsample answered them only after the LENA measurement had been collected, in a short survey administered during the surveyor's visit to pick up the device. For respondents sampled for two LENA recordings (the endline intervention sample), the questions were asked after their final (second) LENA recording. Figure A.3 summarizes the study timeline and timing of the different endline questions.

At endline, we asked treatment respondents about their participation in the baseline survey

[18]Our approach also implies a smaller sample size for estimating the short-run effects, as the analysis is conducted within the original control group. However, if the study is powered to detect long-run effects and fade-out is expected, then a smaller sample size will often suffice to detect short-run effects.

to understand their susceptibility to experimenter demand effects and engagement with the intervention. When asked whether they "recall anything specific about" being interviewed by our survey organization, Innovations for Poverty Action (IPA), 71% of the treatment group associate the survey organization interview with receiving a calendar, 58% associate it with watching a video, and 21% say they only recall answering questions (Table 3). When prompted about the video/calendar, 91% report remembering the calendar and 93% report watching the video. However, only 52% can describe elements of the video and only 36% remember the message about talking to your child.[19] The calendar was quite popular, with 93% ever hanging it on their wall and 78% still using it 6-8 months later. The stars on the calendar were less popular, with only 36% of respondents reporting that they colored in the stars as encouraged by the baseline surveyor.

In our final set of endline questions for control and treatment respondents, we asked their opinions on the barriers to parent-infant conversations for families in their community (see Table A.3). Among respondents who did not watch the video, 43% do not mention any barriers. The most oft-reported barriers by these respondents are "it's hard to remember to do it, it takes effort to make it a habit" (35%), " it's mocked/frowned upon in the community" (32%), and "it's not clear that it makes any difference for the child" (28%).

2.6 Outcome measures

We combine the several measures we collected into summary indices corresponding to our outcomes of interest: mother's beliefs, mother-reported parental verbal inputs, mother-reported child language skills, mother-reported child gestural communication, and surveyor-measured child cognitive development. We follow Anderson (2008)'s suggested method for constructing variance-weighted summary indices.[20]

[19]Respondents were asked in order "*We interviewed you in March. Do you recall anything specific about that interview?*", then were probed specifically about the video and calendar "*Did you see a video and/or received a calendar?*". To measure how much respondents remembered from the video, we asked them "*In March, the surveyor should have shown you a video and given you a calendar. Could you tell me more about what you remember from the video?*. If they still did not mention anything related to talking to babies, surveyors asked *"Do you remember the overall message/idea of the video?"*.

[20]We do not impute missing index components when calculating the components' weights and the indices. Except for the observed child development index, missing index components are due to "Refuse to answer"

Mother's beliefs. We measured parental beliefs about verbal engagement using items from the Caregiver Knowledge of Child Development Inventory (CKCDI) (Ertem et al., 2007) and the Baby Survey of Parental Expectations And Knowledge (Baby SPEAK) (Suskind et al., 2016). We adapted these questions to the Ghanaian context through an extensive piloting process. The adapted CKDCI questions, shown in Table A.4, ask the caregiver when (in terms of the child's age) a parent should start doing activities such as singing songs to, telling stories to, or saying complete sentences to a child in order to promote the child's brain development. The adapted Baby SPEAK items present statements about child cognitive and language development to respondents and asked them to rate their level of agreement with the statement on a Likert scale from 1 (strongly disagree) to 4 (strongly agree).

Mother-reported parental verbal inputs. For mother-reported parental verbal inputs, we used questions developed by the Harvard Laboratory for Developmental Studies designed specifically for Ghana (Duflo et al., 2024). This instrument consists of questions about whether the respondent and/or another adult engaged in a given activity with the focal child (see list in Table A.5).

Mother-reported child development. To assess children's language and communication skills, we relied on items from a version of the MacArthur-Bates Communicative Development Inventories Words & Gestures (MB-CDI-WG) adapted to Ghana by the Harvard Laboratory for Developmental Studies (Duflo et al., 2024).

Specifically, for language, mothers were asked about specific words and phrases their child understands and/or attempts to pronounce (see Appendix Table A.6). Mothers were also asked three questions about whether the child has started to talk. We do not use an Anderson index to compute the child language score; instead, we compute a score using Item Response

and "Don't know". Between 95-100% of respondents answered all components. For the observed child development index, missing components are due to the surveyor selecting the answer "Unable to assess" (because the infant became agitated, refused to participate, etc.). We drop observations missing more than 50% of components (N=17).

Theory (IRT) which involves estimating a one-parameter logistic model on the mother's responses to these questions, where the model assigns a difficulty level to each question and, then, a latent trait to each individual based on their answers to the questions adjusting for the question's difficulty level.[21]

For gestural communication skills, mothers were asked a series of questions on how their child communicates through gestures (see Appendix Table A.7 for questions).

Since the MB-CDI-WG was initially designed for children 8 to 18 months old, we test for floor and ceiling effects by age group. Unsurprisingly, there are substantial floor effects for children under 5 months old. For all other age groups (6-9 months old, 10-14 months old, or 15 months or older), we do not find substantial floor or ceiling effects (see Panels A and B of Appendix Figure A.4).

Surveyor-measured child development. We collect a direct observation (by surveyors) of child development by adapting questions from the Ages and Stages Questionnaire (ASQ) and the Oxford Neurodevelopment Assessment (Ox-NDA), shown in Table A.8.[22] Specifically, we use two ASQ-like items adapted to be administered by surveyors (as opposed to collecting mothers' reports): whether the child's eyes follow the mother when she moves, and whether the child's eyes follow a toy placed in front of her.

Most items in the original Ox-NDA ask the surveyor to take an action (e.g., placing a spoon, cup, plate, ball and pen in front of the child and asking the child "Which one is the spoon?") and record observations about the child's response (e.g., whether they pointed to/picked up the spoon, pointed to/picked up a different object, or did not respond). While the Ox-NDA is designed for children 10-14 months old, 49% of our focal children were under 9 months

[21] We follow the MacArthur-Bates CDI Advisory Board in using IRT (Marchman and Dale, 2023). For the child language score, in addition to the vocabulary checklist, we include the responses to the three additional language questions.

[22] We considered using other child cognitive tests/assessments such as the Bayley Scales of Infant and Toddler Development (BSITD), the Denver Developmental Screening Test, and others. However, some of these other tests are too costly (the BSITD costs around $120 per child according to Attanasio (2015)) or need to be administered by a trained psychologist. In addition, these tests have not been piloted in and adapted to the Ghanaian context.

or younger at endline. In July 2022, we piloted the items, after adapting them to the local context, and identified the most promising seven items (in terms of expected variation) to include in our measure.

We combine both sets of items into one "surveyor-observed child development index". Not surprisingly given how difficult it is to measure outcomes for very young infants, the score is not positively correlated with age prior to 3 months (see Figure A.5).

When prompted to do some of the Ox-NDA items, some children disengaged from the test or refused to perform tasks they had performed at other instances in the survey, such as babbling. Two weeks into the data collection, we added a question on whether the surveyor observed the child babbling (at least one syllable) *at some point* during the home visit. As a result, we have two observed measures of babbling: one as part of the OX-NDA test and one for whether the child babbled at any point during the survey. As the latter measure was added to the survey two weeks after starting data collection, it is only available for 71% (888/1,258) of respondents who completed the endline survey.

LENA outcomes. We focus on two LENA measures: female adult words per minute, a measure of parental verbal input; and child vocalization count per minute, a measure of child verbal output.

To measure the impact of the endline intervention, we use the second day-long LENA recording of children's auditory environment (i.e., recorded parental verbal inputs and child vocalizations) and perceived barriers to conversing with babies (recorded in the debrief survey after the second LENA recording).

2.7 Pre-registration

We registered the RCT first in the AEA RCT Registry (ID AEARCTR-0007161) and subsequently with more details in ClinicalTrials.gov (ID NCT04807907).

The main deviation from the (more detailed) study protocol registered in ClinicalTrials.gov occurred due to a budget-induced reduction in the LENA sample, which created concern about statistical power that we sought to alleviate by adding the surveyor-observed outcomes and the endline intervention. We had to reduce the size of the LENA sample because the cost per LENA recording was higher than originally projected. With the smaller LENA sample, we estimated a minimum detectable effect size of 0.196 SDs with 80% power.[23] Given the light-touch nature of the intervention and the possibility of fade-out over 6-8 months, we thought this level of statistical power might be insufficient to estimate policy-relevant effects. To increase our power, we added the direct observation measures of child cognitive development, which were administered to the entire sample. To understand whether there was an immediate effect that faded out over time, we added the endline intervention to quantify the immediate effects of the video.

In our analysis of the LENA recordings, we present female adult words per minute and child vocalization count per minute as the primary outcomes rather than the pre-specified outcomes of adult word count and number of conversational turns (which are still presented as secondary outcomes). We made this change based on feedback received after presenting preliminary results. The change allows us to more cleanly measure the mother's input (the behavior of the potential intervention participant) and the child's output, as female adult word count provides the best measure of a mother's verbal input and child vocalization count provides the best measure of the focal child's verbal output.[24]

Besides these deviations, we followed our pre-registered sampling criteria, randomization procedure, primary outcome measurement, and main analysis. We did not pre-register any heterogeneity analyses or robustness checks.

[23] Assuming a LENA sample size of 900 with 9% attrition and 100% take-up rate.

[24] Conversational turn count is the number of alternations between the child and adults in the vicinity, so it combines parental input and child output.

2.8 Descriptive analyses

We collected an array of outcome measures because it was unknown which ones would be reliable in our context. In Appendix C, we present some descriptive analyses of our outcome variables and how they correlate with each other, and provide insights for what types of measurement appear the most promising. We highlight three findings here. First, surprisingly, child vocalizations measured by the LENA do not increase strongly with age. A pattern in the data that might offer a partial explanation is that parental words directed at the child actually decrease with child's age, perhaps due to a decline in time spent with the mother. As such, to the extent that child vocalizations are *responses* to parental inputs, they are not necessarily a good proxy for child language development. Second, there is a positive correlation between LENA-measured female adult verbal inputs and mother-reported IDS behavior—our objective and self-reported measures of the mother's behavior. However, the correlation is weak, which could stem from the mother spending limited time with the child as mentioned, the LENA being a noisy measurement tool in this context, or the unreliability of self-reports.[25] Future research to disentangle these possible explanations would be useful. Third, neither LENA-measured verbal inputs to the child nor the child's verbal outputs are increasing in maternal education or socioeconomic status, in contrast to the strong positive correlations seen in wealthier countries (Hart and Risley, 1995; Hoff, 2003; List et al., 2021).[26] This suggests that, in LMICs, the need for interventions to promote IDS exists across the entire socioeconomic spectrum.

3 Empirical Framework

3.1 Treatment effects of main intervention

We identify the impact of the video-plus-calendar intervention on our outcomes of interest at endline (6-8 months after the intervention) by estimating the following equation via ordinary

[25] Child vocalizations are also weakly positively correlated with surveyor-observed and mother-reported child language development.

[26] Parental beliefs about the importance of IDS or self-reported behavior are also not positively correlated with maternal education or socioeconomic status.

least squares (OLS):

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \eta_i + \epsilon_i \tag{1}$$

where i denotes a household, Y_i is the outcome of interest measured at endline, and T_i is a dummy variable that equals 1 if the mother received the intervention at baseline and 0 otherwise. X_i is a vector of controls including the child's age in days, date of the survey, and an indicator for the surveyor being female. For outcomes derived from the LENA recording, we control for the child's age in days, household size, the day of the week the audio was recorded, and interruptions to the LENA's recordings.[27] η_i represents clinic fixed effects. ϵ_i is the error term, and the estimated standard errors are robust to heteroskedasticity. We adjust for multiple hypothesis testing among our primary outcomes using Romano and Wolf (2016).[28]

The experimental variation in T_i generated by the RCT enables us to estimate the causal effect of the intervention on our outcomes of interest as long as the stable unit treatment value assumption (SUTVA) holds. Our estimate of this effect will be captured by β_1, so we will be interested in testing whether β_1 is significantly different from zero.

One threat to our identifying assumption (SUTVA) is that control respondents may learn about the intervention from treatment respondents. These spillovers would likely downward bias our estimates by improving the outcomes of the control respondents. To explore the magnitude of spillovers, we gathered rough measures of the extent of social diffusion of our intervention. At baseline, we asked treated respondents after they saw the video "Do you know anyone who has already seen this video?" (enrollment in the study was on a rolling basis over 3 weeks). 7% reported knowing someone who had seen the intervention video. At endline, 8% of 195 control respondents who received the endline intervention and were asked the same question reported knowing someone who had seen the video and 16% of 615 treatment respondents (who had received the 'main intervention' at baseline) had discussed

[27]Interruptions include the device being removed or the child being on someone's back where the sound might be muffled.

[28]We assess the robustness of our results to alternative specifications in Appendix Table A.9.

the video with friends (see Table 3); a subset of these friends could be in the control group. We did not probe control group respondents further on what they had heard about the video prior to their baseline survey nor what they discussed with others post-baseline. Hence, we are unable to provide further details on how much control respondents learned about the video or its key message through social spillovers. Such spillovers would lead us to underestimate the treatment effect, but the magnitude of this underestimation is likely very small.

3.2 Before-after immediate effect of the endline intervention

To measure the effect of the endline intervention on the measures recorded by the LENA, we estimate the following equation using all respondents who received the endline intervention:

$$Y_{it} = \gamma_0 + \gamma_1 EndlineIntervention_t + \gamma_2 Z_t + \omega_i + \mu_{it} \tag{2}$$

where t denotes whether this was the first or second day-long LENA recording for a given focal child and $EndlineIntervention_t = 1$ if $t = 2$, i.e., for the observation collected after the respondent received the endline intervention, and 0 otherwise. Z_t represents a vector of LENA-recording specific characteristics such as the day of the week the audio was recorded and interruptions to the LENA's recordings. ω_i represents household fixed effects. The estimated standard errors are robust to heteroskedasticity.

The use of household fixed effects means that our coefficient of interest, γ_1, identifies the effect from within-household (i.e., within-focal-child) changes over time in treatment status. More precisely, γ_1 estimates the change in the child's verbal inputs or outcomes between the *day before* the endline intervention and the *day after* it (i.e., the day after their mother watches the video). We interpret γ_1 as the immediate effect of their mother watching the video on the child's verbal environment.[29]

[29]The mother knows she is being recorded and we cannot speak to whether the results generalize to when she is not being recorded. Another potential confound is that the post-treatment measure is always the second day of recording, while the pre-treatment measure is always the first day of recording. Past studies using the LENA did not report that LENA outcomes differed substantively between the first and second

4 Results and Discussion

4.1 Effects of the intervention after 6-8 months

In Table 4, we present our main results on the effects of the intervention, measured 6 to 8 months after it was delivered. We analyze outcomes reported by the mother, observed by the surveyor, and recorded by the LENA device.

Mother-reported outcomes. Columns 1-4 of Table 4 report effects on outcomes reported by the mother. The intervention increased mother's beliefs in the importance and efficacy of conversing with infants by 0.126 standard deviations (SD) (p =0.030). We next analyze whether it changed her behavior. The intervention increased an index of the parental verbal behavior toward the child by 0.124 SD (p =0.025). Treatment mothers also report significantly higher child language skills and non-verbal communication (0.102 SD and 0.097 SD increases, respectively; language p =0.003; gestural communication p =0.018). Thus, based on mothers' reports, the full theory of change materialized: They realized IDS was important, they started practicing it more, and this translated into improved child language development.

The magnitude of these effects are lower than heavier-touch interventions but encouraging given the light-touch nature of the intervention. Relative to the pooled effect sizes computed by Jeong et al. (2021) in their meta-analysis of around two dozen randomized interventions with children under 12 months at baseline, our treatment effects on parental beliefs are 26.3% of the average effect of heavier-touch interventions; for parental behaviors and infant language development, our effects are 42.8% and 48.2% of the average effects, respectively.

Using indices has the disadvantage that the outcomes are less concrete, so we also present the effects for a few concrete outcomes.[30] Treated mothers are 10.4 percentage points (pp)

usage.

[30] Appendix Tables A.4, A.5, A.7 and A.8 present the disaggregated results for each Anderson index individual outcomes. Appendix Table A.10 presents the results for other available LENA outcomes. We do not run Equation 1 on the individual components of the language score as this index is computed through Item Response Theory, but we present the results when using easily interpretable outcomes computed from the score components (such as the number of words in the list the child understands) in Table A.6.

more likely to report that parents should start talking to their infant at birth, a key message of the video and calendar ($p < 0.001$). The effect size is 31.5% of the control group mean. Treated mothers also report 6.5 pp ($p = 0.019$; 14.8% of control mean) more that an adult read to or looked at a book with the child and 4.5 pp more that an adult told stories to the child ($p = 0.089$; 14.5% of the control mean) in the last 4 weeks. Based on their reports, their children understand 0.48 more words ($p = 0.008$; 6.8% of control mean) and can say 0.31 more words ($p = 0.014$; 29.3% of control mean) from the 16 words listed by the surveyor (see Table A.6 for the list of words and treatment effects on easier to interpret/more concrete outcomes).

Surveyor-observed outcomes. The treatment effects on the surveyor-observed measure of child development are positive but smaller than the effects for mother-reported outcomes and generally insignificant. We estimate a 0.038 SD increase ($p = 0.430$) in child development (Table 4; column 5). The 95% confidence interval on this estimate (-0.058 to 0.134) includes both negative effects and the mother-reported effects on child language/gestural communication.

There is a significant positive effect on whether the child babbled at any point during the survey. As mentioned earlier, we added this question after two weeks of data collection. Among the 71% of respondents for whom the measure was recorded, the intervention led to a 5.4 percentage point increase ($p = 0.054$) in the infant babbling, which is 26.1% of the control mean (Table 4; column 6).

Heterogeneity by child age. We also estimate the treatment effects on the mother-reported and surveyor-observed outcomes separately by age group, reported in Table 5.[31] Heterogeneity by age is interesting per se, but we show this breakdown mainly because the various instruments we use are more appropriate for certain age groups. In particular, the child development measures are not designed for very young infants. For mother-reported outcomes, the positive pooled effects reported above are driven by children aged 6 months and

[31] See Appendix Figure A.6 for full distribution of infant age at endline.

older. This is reassuring since these are the ages when, a priori, the measures should be more reliable. The treatment effect on surveyor-observed child development, which is statistically insignificant in the full sample, shows no clear age gradient (see Appendix Figure A.7 for effects by child's age in months). The treatment effect on the probability of babbling is concentrated among young infants. Among young infants, the intervention doubles the proportion who babbled, from 10% to 20% ($p = 0.049$; Table 5; column 4).

4.1.1 LENA-recorded outcomes

The treatment effects on our primary LENA-recorded outcomes are positive but insignificant. We estimate an insignificant 1% increase in female adult words per minute and 3% increase in child vocalizations per minute (0.131 words w/ $p = 0.851$ and 0.052 w/ $p = 0.377$ respectively; (Table 4; columns 7-8).[32]

Examining the more detailed breakdown of the audio recording available in the LENA data, we find weak evidence for increases in exposure to speech. There are insignificant positive effects on the percentage of the recording that is female adult speech, male adult speech, other child speech, focal child vocalizations, focal child non-vocalizations, and faint/overlapping sounds (see Appendix Table A.10). There are negative effects on the amount of the recording with silence or background noise, with the largest effect on background noise (-1.25 pp; 6.4% of the control mean; $p = 0.056$). We interpret these effects as weak evidence that increased IDS is crowding out background noise and silence. The negative correlations between background noise and child vocalizations, conversational turns, female adult word count, male adult word count, and other child speech in the control group support this interpretation.

One limitation of our LENA results is that the estimates are noisy since our sample size is half as large as for the self-reported outcomes due to budget constraints. As anticipated in our power calculations (see Section 2.7), the treatment effects on LENA measures have larger

[32]The distributions for the treatment and control groups are shown in Appendix Figures A.8 and A.9. Treated children seem to produce more vocalizations at the bottom of the distribution. For the full distribution, the Kolmogorov-Smirnov test of equality yields $p = 0.169$.

standard errors (about 0.07 when using standardized outcomes) than the treatment effects on the other outcomes. This could contribute to the insignificant effects on the LENA outcomes while we observe significant and substantial effects on the mother-reported outcomes. Future research could attempt to survey more households or record households for multiple days to see if the reported maternal behavior changes translate into true small, hard-to-detect changes.

4.1.2 Experimenter demand effects

Experimenter demand effects could arise if the treatment group associated the IPA surveyors with the intervention and, thus, felt pressure to report believing in, practicing, and seeing positive results from conversing with infants when they spoke to another IPA surveyor 6-8 months later. We included the surveyor-observed measures to avoid experimenter demand effects, but given the inconclusive results on these measures, we attempt to explore this hypothesis further.

We exploit the fact that 131 treatment respondents (21% of treatment respondents), when asked at the end of the endline activities, did not associate the intervention video or intervention calendar with IPA's baseline interview (see Table 3). This group is unlikely to be subject to experimenter demand effects. We test whether this group has smaller treatment effects, which one could interpret as evidence of experimenter demand driving the results (Appendix Table A.11 reports the results). Note that this test might over-conclude that there is experimenter demand because this group may have forgotten about the intervention and thus not applied the information, so they might truly have smaller treatment effects.[33]

For the belief index, we find results consistent with experimenter demand effects. Those who associate the intervention with the survey organization have a significantly larger treatment effect. This pattern is not mirrored for any of the other mother-reported outcomes. The differential treatment effects among those who associate the survey with the intervention

[33] Since we are conditioning on an endogenous variable, differences in baseline characteristics may also be driving differences in endline outcomes.

are close to 0 and statistically insignificant. Experimenter demand could also cause LENA inputs to be higher (the mother uses IDS more when her behavior is being recorded), but, reassuringly, we find no evidence for this.

With the caveat that our test is imperfect, we tentatively conclude that experimenter demand effects may drive the treatment effect on the belief index but are unlikely to drive the effects on mothers' behavior and children's outcomes that we observe.

4.1.3 Robustness checks

To test the robustness of our results, we estimate the treatment effects using alternative specifications: excluding control variables other than clinic fixed effects, including clinic-day (stratum) fixed effects instead of just clinic fixed effects, adding surveyor fixed effects, and using double Lasso to select control variables. We describe the tests in more details below, but Figure 1 graphically summarizes them by showing the treatment coefficient and 95% and 90% confidence intervals from each robustness specification for our eight main outcomes (see Appendix Tables A.9 and A.12 for regression tables.). In brief, the results remain significant at the 10% level for mother-reported mother's beliefs, mother's behavior, and child gestural communication and language scores, and they remain insignificant but positive for child development and LENA outcomes, across the different robustness checks.[34] The significance of the effect on observed child babbling outcome is not robust to the inclusion of surveyor fixed effects or Lasso-selected controls.

When we include only clinic fixed effects or add clinic-day fixed effects, the coefficients change only slightly relative to our main specification in Table 4. The standard errors increase when excluding control variables, as expected, but barely change when replacing clinic fixed effects by clinic-day fixed effects.

Surveyors might be idiosyncratic in how they interpret and record mothers' responses or assess the child's behavior, so surveyor fixed effects could be correlated with outcome measures,

[34]The treatment effects on LENA adult words and child vocalizations per minute are noisy throughout, so we do not discuss them further in this section.

in which case including them could improve precision. However, every surveyor surveyed both treatment and control respondents, and, by design, surveyor fixed effects should not be systematically correlated with treatment. This is indeed the case in the data: Surveyor fixed effects are significantly correlated with all of the outcome measures except the LENA outcomes (reassuringly), but they are not correlated with treatment status (see Appendix Table A.13). When we include surveyor fixed effects, there are nonetheless some changes in the treatment coefficients. The effect on observed child babbling remains positive but is no longer statistically significant (3.2 percentage points; $p = 0.232$). The other results that are significant in the main specification remain significant with surveyor fixed effects.

We use the double Lasso approach of Belloni et al. (2013) as implemented by Ahrens et al. (2019) to flexibly choose control variables, separately for each outcome. Overall, we see little change in the coefficients and small reductions in standard errors. Significance levels do not change except for the effect on infant babbling where the p-value increases from 0.054 to 0.135 and child gestural communication index where the p-value increases from 0.018 to 0.083.

We also compute Romano-Wolf stepdown adjusted p-values to adjust for multiple hypothesis testing and report them in Table 4 (Romano and Wolf, 2016). Focusing on the results with a p-value below 0.1 using the conventional t-test, the effect on reported child language score remains significant (stepdown $p = 0.023$), while the stepdown p-values are above 0.1 for reported maternal beliefs ($p = 0.16$), parental behavior ($p = 0.16$), child gestural communication ($p = 0.13$), and observed child babbling ($p = 0.21$).

4.2 Effects on newly-informed mothers

We next present the effects of the endline intervention, or the immediate effects of the mother watching the video on the child's verbal inputs and outputs. The treatment effects (estimated using Equation 2) are reported in Table 6. On average, the child hears 1.83 more adult words ($p = 0.036$) the day after the endline intervention compared to the day before. In our context, women speak far more words to infants than men do, so the effect is primarily driven by a rise in female adult words (accounts for 76% of the effect). However, there is also an impact on male adult words (0.51 words; $p = 0.076$) suggesting spillovers of the intervention to other members of the household who did not view the video. As expected, we do not see increases in measures of child verbal output such as child vocalizations per minute; one would expect these gains to only materialize in the longer run as child language skills accrue from increased IDS. We see a modest but insignificant increase in "conversational turns" per minute (0.015 turns; $p = 0.35$), which requires engagement between adults' verbal inputs and the focal child's verbal output.

The positive, significant, and substantial impacts on parental verbal inputs show that mothers do not face a "technological barrier" in verbally engaging infants. After watching only a 3-minute video, mothers know how to significantly increase their verbal engagement and persuade other household members to do so too. The difference between these substantial effects and the positive but noisy effects of the main intervention on the outcomes measured by the day-long recording suggests that there are barriers to sustaining this level of behavioral change.

4.3 Self-reported barriers to IDS

When we asked respondents about potential barriers preventing families from talking to their babies, the most common answers for those who never watched the video were "it's hard to remember/make a habit" (35%) or "it's mocked/frowned upon in the community" (32%) (see Appendix Table A.3). Comparing this group to those who received the endline intervention (i.e., the newly-informed mothers), we find that the endline intervention increased reporting of mocking/social scorn as the main barrier by 19.7 pp ($p < 0.001$; see Appendix Table A.14 column 3). However, such a reaction appears short-lived. Mothers who received the treatment 6-8 months earlier (the treatment group who received the main intervention) are much less likely to report social scorn as the main barrier ($p = 0.02$). In contrast, habit formation is equally likely to be cited as a main barrier among the main treatment group and the endline intervention group ($p = 0.967$).

To summarize this evidence, after just one day of experimentation with the encouraged behavior, there are substantial social norms-related concerns about engaging in it, but the concerns seem to fade over the subsequent 6-8 months, while the challenge of habit formation persists as a barrier. One interpretation of these results is that people quickly get over the initial awkwardness of departing from traditional parenting practices, possibly because they realize that just explaining the benefits of IDS to others is sufficient to generate social acceptance. But departing from traditional parenting practices takes more than social courage: it also requires adopting new habits, which is notoriously difficult.

As further suggestive evidence that habit formation inhibited infant verbal engagement for some mothers, first-time mothers—so women who have not established their "typical" parenting practices—are 5.9 pp less likely to report habit formation as a barrier, and they experience larger treatment effects on the child language score and the child gestural communication index (see Appendix Table A.15).[35]

In addition, we test whether the treatment effects are larger among mothers who followed the recommendation to fill in the stars on the calendar if they conversed with their infant every day in a given week.[36] We find that filling in the stars is associated with larger effects on all of our main outcomes except child vocalizations (see Appendix Table A.18, effects only significantly greater for parental verbal inputs and child gestural communication index). While this finding could be driven by selection, since adherence is a choice, it is consistent with filling in the calendar helping participants to form and stick to a habit of conversing with their child.

5 Cost-effectiveness

As a measure of cost-effectiveness, we estimate the cost per child's SD improvement in cognitive and language outcomes. We assume the at-scale costs of the intervention would

[35] Balance on baseline characteristics (F-test p-value $p = 0.431$) and IDS beliefs and behavior (F-test p-value $p = 0.361$) for the sample of first-time mothers is shown in Appendix Tables A.16 and A.17.

[36] We added questions on calendar use mid-survey so do not have this data for 90 treatment respondents.

only include printing out the calendars and delivering them to health clinics. As our measure of the benefits of the intervention, we average the effects on child language and gestural communication (from Table 4).[37] Using these estimates, we calculate that the intervention delivers a 1 SD improvement in child development for $4.52. This would be lower than any of the interventions included in the meta-analysis of 'responsive caregiving' intervention by Verguet et al. (2022) (the majority of which are home visiting interventions).[38]

Cost-effectiveness under various alternative assumptions is shown in Table 7. When we use observed measures of language and cognitive development instead of mother-reported measures, the estimate rises to $10.01. If we assume that the effect sizes at scale-up are only half (Panel B) or a quarter (Panel C) as large as our estimated intervention effects, the cost per SD gain in child development rises to $9.05 and $18.09, respectively. Using observed measures, the estimates rise to $20.01-$40.02 per SD.

Using the intervention costs in our RCT, we estimate a cost of $30.26 per SD improvement using our estimated effect size on mother-reported measures. In the RCT, we paid trained surveyors to stay at each clinic during the clinic's working hours to show the video and give out calendars which drives up costs relative to an at-scale model where existing health clinic staff could perform these tasks. Even with this inefficient use of labor, our intervention would still be more cost-effective than the 12 interventions evaluated by Verguet et al. (2022).[39]

The treatment effects on objective measures are mostly insignificant, and we cannot fully rule out experimenter demand effects influencing the results for mother-reported outcomes, so there is a possibility that the intervention does not truly have positive effects. This

[37] Cost-effectiveness is similar when we use Cohen's d as the measure of the intervention effect, following Verguet et al. (2022), instead of our regression coefficients.

[38] When we compare to Verguet et al. (2022), we are referencing their 'standardised cost-effectiveness' estimates. They standardize costs and prices by using 2010 USD as their unit and the average gross domestic product per capita of LMICs in 2010 as their wage rate. While we present the cost-effectiveness estimates in 2021 USD with local labor costs in Table 7, we harmonize the price level and labor costs when comparing to Verguet et al. (2022).

[39] After adjusting for inflation, the most cost-effective intervention in Verguet et al. (2022) delivers a 1 SD improvement for $35.96.

possibility of null effects combined with the high upside if the intervention does have positive effects means that a policymaker's decision on whether to scale up this policy will depend on their risk preferences. Future research using more precise observed measures or larger sample sizes could remove the uncertainty around cost-effectiveness, enabling even risk-averse policymakers to scale up the policy. Furthermore, combining the intervention with support for mothers in forming a parent-infant conversation habit could be even more cost-effective than the intervention alone, given our finding that habit formation is the main barrier to mothers sustaining their short-run levels of verbal input.

6 Conclusion

This paper provides experimental evidence on a 'light-touch' intervention aimed at encouraging mother-infant conversations in northern Ghana. The short-term effects of the intervention show a sizable increase in the number of words that mothers spoke to their infants, while effects after 6-8 months are present but more mixed, likely due to challenges in mothers sustaining this new habit.

Even the 6-8 month point estimates suggest that the intervention is cost-effective, delivering a 1 SD improvement in child development for $5 to $10. We administered the intervention to women visiting public health clinics for prenatal or postnatal checkups. This setting and sample mirrors how the intervention could be implemented at scale: the 3-minute video could be shown in waiting rooms of prenatal care centers, and health workers could hand out the calendars to patients during their visit. Even if the treatment effects fall by 50-75% at scale, relative to what we estimate, the intervention would still be more cost-effective than alternative policies such as home-visiting programs. However, the possibility of experimenter demand effects when we analyze self-reported measures and the noisiness of the objective measures we used mean that further research is needed to be confident in these policy conclusions.

We identify local norms and habit formation as the main remaining barriers to parent-infant conversations once the intervention has conveyed the importance of IDS. While our evidence suggests that local norms are mutable, difficulties with habit formation seem to be more persistent. Future research could focus on complementary interventions to increase retention and compliance with the intervention by helping mothers form an infant-directed speech habit.

References

Ahrens, Achim, Christian B. Hansen, and Mark E Schaffer, "PDSLASSO: Stata module for post-selection and post-regularization OLS or IV estimation and inference," *Statistical Software Components*, 1 2019.

Anderson, Michael L, "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495.

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

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "Inference on treatment effects after selection among high-dimensional controls," *Review of Economic Studies*, 2013, *81* (2), 608–650.

Black, Maureen M, Susan P Walker, Lia C.H. Fernald, Christopher T Andersen, Ann M. DiGirolamo, Chunling Lu, Dana C. McCoy, Günther Fink, Yusra R Shawar, Jeremy Shiffman, Amanda E Devercelli, Quentin T Wodon, Emily Vargas-Barón, and Sally Grantham-McGregor, "Early childhood development coming of age: science through the life course," *The Lancet*, 2017, *389* (10064), 77–90.

Duflo, Esther, Spelke Elizabeth Dupas Pascaline, and Mark Walsh, "The intergenerational effects of secondary education: Experimental Evidence from Ghana," 2024.

Ertem, IO, G Atay, DG Dogan, A Bayhan, BE Bingoler, CG Gok, S Ozbas, D Haznedaroglu, and S Isikli, "Mothers' knowledge of young child development in a developing country," *Child: care, health and development*, 2007, *33* (6), 728–737.

Farran, Lama K, Chia-Cheng Lee, Hyunjoo Yoo, and D Kimbrough Oller, "Cross-cultural register differences in infant-directed speech: An initial study," *PLOS ONE*, 2016, *11* (3), e0151518.

Fernandes, Michelle, "The Oxford Neurodevelopment Assessment (OX-NDA) for Infants aged 10 to 14 months. THE Ox-NDA MANUAL – V2.1," 2021.

Gilkerson, Jill and Jeffrey A Richards, "The LENA natural language study," *LENA Found. Tech. Rep.*, 2008, pp. 1–26.

_ **and** _ , "A Guide to Understanding the Design and Purpose of the LENA® System," *LENA Foundation Technical Report*, 2020.

Hart, Betty and Todd R Risley, *Meaningful differences in the everyday experience of young American children.*, Paul H Brookes Publishing, 1995.

Hoff, Erika, "The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech," *Child development*, 2003, *74* (5), 1368–1378.

Jeong, Joshua, Emily E Franchett, Clariana V. Ramos de Oliveira, Karima Rehmani, and Aisha K. Yousafzai, "Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis," *PLOS Medicine*, 2021, *18* (5).

_ , Helen O Pitchik, and Aisha K Yousafzai, "Stimulation Interventions and Parenting in Low-and Middle-Income Countries: A Meta-analysis," *PEDIATRICS*, 2018, (4).

Lally, Phillippa and Benjamin Gardner, "Promoting habit formation," *Health psychology review*, 2013, *7* (sup1), S137–S158.

List, John A., Julie Pernaudet, and Dana Suskind, "It All Starts with Beliefs: Addressing the Roots of Educational Inequities by Shifting Parental Beliefs," *NBER Working Paper*, 2021, (No. 29394).

Marchman, Virginia A and Philip S Dale, "The MacArthur-Bates Communicative Development Inventories: updates from the CDI Advisory Board," *Frontiers in Psychology*, 2023, *14*, 1170303.

Monnot, Marilee, "Function of infant-directed speech," *Human nature*, 1999, *10* (4), 415–443.

Romano, Joseph P and Michael Wolf, "Efficient computation of adjusted p-values for resampling-based stepdown multiple testing," *Statistics & Probability Letters*, 2016, *113*, 38–40.

Rothman, Alexander J, Peter M Gollwitzer, Adam M Grant, David T Neal, Paschal Sheeran, and Wendy Wood, "Hale and hearty policies: How psychological science can create and maintain healthy habits," *Perspectives on Psychological Science*, 2015, *10* (6), 701–705.

Shrestha, Merina, Manjeswori Ulak, Tor A. Strand, Ingrid Kvestad, and Mari Hysing, "How much do Nepalese mothers know about child development?," *Early Child Development and Care*, 2019, *189* (1), 135–142.

Suskind, Dana L, Christy YY Leung, Robert J Webber, Alison C Hundertmark, Kristin R Leffel, Iara E Fuenmayor Rivas, and William A Grobman, "Educating parents about infant language development: a randomized controlled trial," *Clinical pediatrics*, 2018, *57* (8), 945–953.

_ , Kristin R Leffel, Eileen Graf, Marc W Hernandez, Elizabeth A Gunderson, Shannon G Sapolich, Elizabeth Suskind, Lindsey Leininger, Susan Goldin-Meadow, and Susan C Levine, "A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study," *Journal of child language*, 2016, *43* (2), 366–406.

Verguet, Stéphane, Sarah Bolongaita, Anthony Morgan, Nandita Perumal, Christopher R Sudfeld, Aisha K Yousafzai, and Günther Fink, "Priority setting in early childhood development: an analytical framework for economic evaluation of interventions," *BMJ Global Health*, 2022, *7*, 8926.

Webb, Thomas L and Paschal Sheeran, "Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence.," *Psychological bulletin*, 2006, *132* (2), 249.

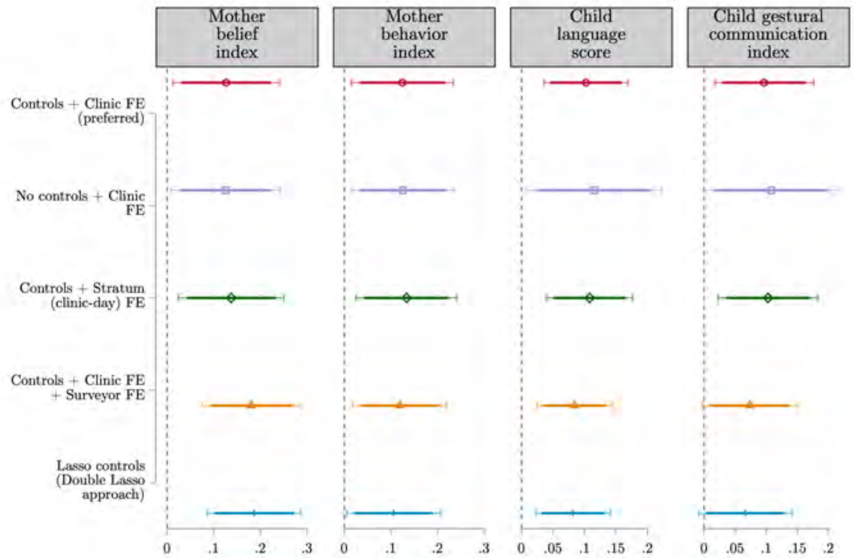
Weisleder, Adriana and Anne Fernald, "Talking to children matters: Early language experience strengthens processing and builds vocabulary," *Psychological science*, 2013, *24* (11), 2143–2152.

Xu, Dongxin, Umit Yapanel, and Sharmi Gray, "Reliability of the LENA Language Environment Analysis System in young children's natural home environment," *Boulder, CO: Lena Foundation*, 2009, pp. 1–16.

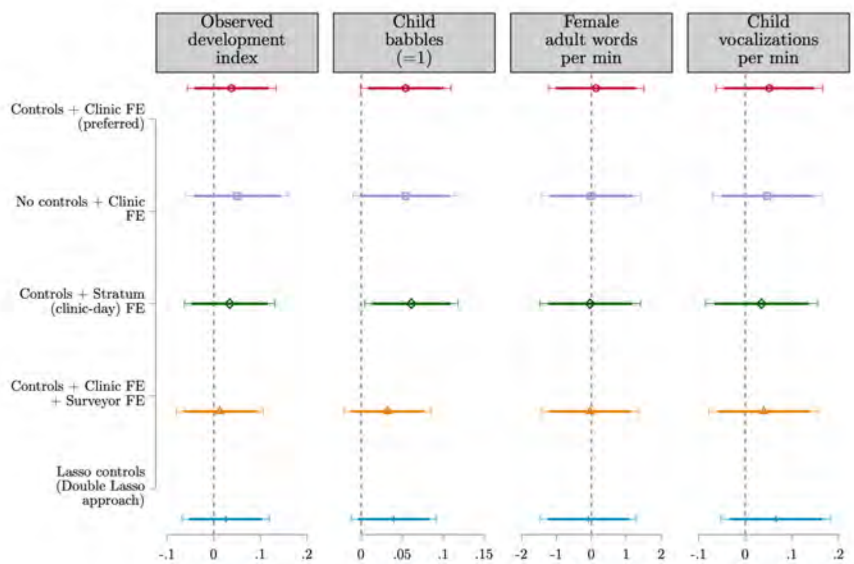
Zellman, Gail L., Rita Karam, and Michal Perlman, "Predicting child development knowledge and engagement of Moroccan parents," *Near and Middle Eastern Journal of Research in Education*, 10 2014, *2014* (1), 5.

Figure 1: Robustness Analysis

(a) Panel A: Mother-reported Measures



(b) Panel B: Observed and LENA Measures



Note: Endline and LENA day 1 recording data. The figure plots the treatment coefficients and 95% (lighter lines) and 90% (darker lines) confidence intervals for regressions of the outcome variable listed at the top on a treatment dummy and the fixed effects (FEs) and controls indicated on the y-axis. The 1st specification is the one used to estimate treatment effects in Table 4. The 2nd to 4th rows change the main specifications by removing control variables other than clinic FEs; replacing clinic FEs by clinic-day FEs; or adding surveyor FEs (also reported in Table A.9). The estimate in the 5th row uses the double Lasso approach of Belloni et al. (2013) as implemented by Ahrens et al. (2019) to choose control variables (also reported in Table A.12). All specifications use robust standard errors. For further details on outcomes, please see Table 4.

Table 1: Baseline Characteristics and Balance

	Full Sample			Treatment		Control		T=C
	Mean	SD	N	Mean	SD	Mean	SD	P-value
Age (years)	27.75	5.17	1,403	27.93	5.15	27.57	5.19	0.194
Dagomba ethnie	0.82	0.38	1,407	0.83	0.38	0.82	0.38	0.825
Main langage spoken: Dagbani	0.88	0.33	1,408	0.88	0.33	0.88	0.33	0.951
Highest level of education:								
None	0.37	0.48	1,406	0.38	0.49	0.36	0.48	0.500
Primary school	0.28	0.45	1,406	0.28	0.45	0.28	0.45	0.879
Secondary school	0.22	0.42	1,406	0.21	0.41	0.23	0.42	0.463
Can read (English/Dagbani)	0.61	0.49	1,408	0.59	0.49	0.62	0.48	0.205
Housewive/no occupation	0.23	0.42	1,408	0.23	0.42	0.22	0.41	0.501
Married	0.99	0.09	1,408	1.00	0.05	0.99	0.12	0.020
Polygamous	0.30	0.46	1,304	0.28	0.45	0.32	0.47	0.214
Partner is home whole month	0.77	0.42	1,399	0.78	0.42	0.76	0.43	0.392
Partner passed primary school	0.75	0.43	1,399	0.75	0.43	0.75	0.43	0.889
Household size	8.62	5.72	1,400	8.71	5.74	8.53	5.71	0.542
# of household members: under-5	1.90	1.60	1,407	1.92	1.66	1.87	1.54	0.537
# of household members: 5-15 y/o	1.88	2.08	1,405	1.96	2.08	1.80	2.08	0.142
# of household members: above-16	4.85	3.22	1,400	4.83	3.23	4.86	3.21	0.877
Has children	0.89	0.31	1,408	0.90	0.29	0.88	0.32	0.138
Has child 6 years or younger	0.75	0.43	1,408	0.77	0.42	0.74	0.44	0.229
Has child older than 1 month	0.69	0.46	1,408	0.70	0.46	0.68	0.47	0.284
Has child older than 3 months	0.64	0.48	1,408	0.64	0.48	0.63	0.48	0.516
Age at first child (years)	22.23	3.50	1,242	22.14	3.37	22.33	3.62	0.327
# of children	2.21	1.55	1,408	2.28	1.54	2.15	1.57	0.105
Age youngest child (months)	15.31	20.74	1,182	15.12	20.12	15.50	21.36	0.754
Youngest child eligible	0.61	0.49	1,408	0.62	0.49	0.60	0.49	0.389
Pregnant with an eligible child	0.39	0.49	1,408	0.38	0.49	0.40	0.49	0.360
Focal child is first born	0.28	0.45	1,408	0.26	0.44	0.30	0.46	0.129
F-test p-value								0.662
Observations	1,408			705		703		

Note: Baseline data. Treatment is a dummy equal to 1 if the respondent received the intervention at baseline. The question on polygamy was added after the start of the data collection, hence is missing for some observations. The F-test p-value reported at the bottom of the table is for the joint significance of the differences between the treatment and control groups for all of the variables reported in the table. For the F-test, missing values (due to refusal/don't know or a logic skip (e.g., age of youngest child when no children)) are replaced by the variable average value and flagged by a dummy.

Table 2: Baseline IDS Beliefs and Behavior

	Full Sample			Treatment		Control		T=C
	Mean	SD	N	Mean	SD	Mean	SD	P-value
Beliefs on IDS and Child Development:								
Time/attention is more important than money to a child's success	0.37	0.48	1,408	0.36	0.48	0.38	0.48	0.517
Child's age (in mo) when:								
a child starts responding with noise/babbles	7.51	9.13	1,364	7.30	7.85	7.72	10.26	0.398
a child starts saying meaningful words	19.99	12.42	1,344	19.67	12.24	20.31	12.59	0.341
it becomes clear a child is smart	35.53	25.93	1,365	34.94	24.54	36.14	27.26	0.391
Child's age (in mo) when parents should start:								
talking to their child	10.90	11.49	1,376	11.03	11.19	10.77	11.78	0.676
talking in full sentences to their child	24.08	17.97	1,282	23.55	17.20	24.60	18.69	0.297
telling stories to their child	21.33	15.93	1,305	21.32	15.23	21.34	16.62	0.985
Self-Reported IDS Behavior:								
Tells stories to youngest child	0.51	0.50	1,059	0.50	0.50	0.52	0.50	0.366
Ask youngest child to repeat words	0.61	0.49	1,059	0.60	0.49	0.61	0.49	0.764
When child was 1m/o: Described objects when cleaning/organizing	0.40	0.49	972	0.36	0.48	0.43	0.50	0.036
When child was 3m/o: Described things to child when walking	0.64	0.48	895	0.65	0.48	0.63	0.48	0.594
Inequality Aversion:								
It is best to treat/invest in children equally	0.48	0.50	1,408	0.47	0.50	0.49	0.50	0.457
A mother should feel bad for 1st child if she provides better care to 2nd child	0.69	0.46	1,408	0.68	0.47	0.69	0.46	0.758
F-test p-value								0.765
Observations	1,408			705		703		

Note: Baseline data. Treatment is a dummy equal to 1 if the respondent received the main intervention (at baseline). Child's age outcomes are in months. In the panel "Beliefs on IDS and Child Development", questions "child's age (in months) when parents should start..." were only asked to respondents who reported that the respective activities were important to a child's brain development. In the panel "Self-Reported IDS Behavior", questions were only asked to a subset of respondents based on their youngest child's age. "Tell stories to youngest child" and "Asks youngest child to repeat words" were only asked to respondents with a child aged 6 years or less, and the two subsequent questions to those with a child aged between 1 month and 6 years, and between 3 months and 6 years. The F-test p-value reported at the bottom of the table is for the joint significance of the differences between the treatment and control groups for all the variables reported in the table. For the F-test, missing values (due to refusal/don't know or a logic skip (e.g., age of youngest child when no children)) are replaced by the variable average value and flagged by a dummy.

Table 3: Treatment Recall & Self-Reported Behavior Change

	Mean	SD	Count	N
Main Intervention Sample (Treatment)				
Without prompting				
Mentions receiving a calendar at baseline	0.71	0.45	436	615
Mentions watching a video at baseline	0.58	0.49	357	615
Mentions neither video nor calendar at baseline	0.21	0.41	131	615
After prompting				
Remembers video	0.93	0.26	490	529
Remembers calendar	0.91	0.28	482	529
Remembers elements of the video	0.52	0.50	273	529
Remembers IDS message	0.36	0.48	192	529
Discussed video with anyone	0.61	0.49	375	615
Discussed video with husband	0.44	0.50	269	615
Discussed video with friends	0.16	0.36	97	615
Calendar use duration				
Still hung up on wall	0.78	0.42	412	529
Hung up at first but not anymore	0.15	0.36	80	529
Never hung up	0.07	0.26	37	529
Calendar use				
Look at date	0.39	0.49	208	529
Color weekly IDS stars	0.36	0.48	188	529
No use of calendar	0.17	0.38	90	529
Number of respondents	615			
Endline Intervention Sample				
Since you saw the video, did you talk to your child:				
More than usual	0.65	0.48	125	191
As much as usual	0.16	0.37	31	191
Less than usual	0.18	0.39	35	191
If talked more to child since seeing the video: how likely are you to continue talking more to your child?				
Very likely	0.60	0.49	73	121
Likely	0.37	0.49	45	121
Number of respondents	191			

Note: Endline data. In the panel "Main Intervention Sample (Treatment)", the sample is restricted to respondents who received the main intervention (at baseline), 6-8 months earlier. Respondents either answered questions at the end of the endline survey, or, if they were sampled to receive a LENA recording device, after the day of recording (7/625 treatment respondents reached for the endline survey did not answer those questions because they did not finish the survey). See Figure A.3 for further details on the study design. Questions on recall after prompting and on calendar use were added mid-data collection, which explains the higher number of missing values. Respondents were asked in order "*We interviewed you in March. Do you recall anything specific about that interview?*" ("Without prompting" panel outcomes), then were probed specifically about the video and calendar "*Did you see a video and/or received a calendar?*" ("After prompting" panel outcomes). To measure how much respondents remembered from the video, we asked them "*In March, the surveyor should have shown you a video and given you a calendar. Could you tell me more about what you remember from the video?*. If they still did not mention anything related to talking to babies, surveyors asked *"Do you remember the overall message/idea of the video?"*. "Remembers IDS message" is a dummy equal to 1 if the respondent mentions talking to infants/children is good for their brain development or that it is good to talk to children from birth. "Color weekly IDS star" is a dummy equal to 1 if respondents reported coloring the stars printed next to each week on the calendar respondents were given at baseline. Respondents were encouraged to fill in the stars next to each week in the calendar if they conversed with their child each day that week. In the panel "Endline Intervention Sample", the sample is restricted to respondents who were sampled to receive the intervention at endline.

Table 4: Treatment Effects, 6 to 8 Months After Intervention

	Mother's Interview				Observed		LENA	
	Mother's belief index	Mother's behavior index	Child language score	Child gestural communication index	Child development index	Child babbles (= 1)	Female adult words per min	Child vocalizations per min
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.126	0.124	0.102	0.097	0.038	0.054	0.131	0.052
	(0.058)	(0.056)	(0.034)	(0.041)	(0.049)	(0.028)	(0.697)	(0.058)
	{0.030}	{0.025}	{0.003}	{0.018}	{0.430}	{0.054}	{0.851}	{0.377}
Stepdown P-values	0.160	0.158	0.023	0.127	0.740	0.207	0.836	0.740
Control mean	0.00	0.00	0.00	0.00	0.00	0.51	16.61	1.53
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,258	1,258	1,258	1,258	1,184	888	774	775

Note: Endline and LENA day 1 recording data. For columns 1 to 6, regressions include controls for the child's age in days, day of the survey, and surveyor gender. In columns 7 and 8, the regressions include controls for the child's age in days, the day of the week the audio was recorded (dummies), the total time (min) the shirt/LENA device was removed from the child, the total time (min) the child was held on someone's back while wearing the device, and the household size. All regressions include baseline clinic fixed effects. All indices are Anderson indices except for the Child language score (column 3) which is calculated using Item Response Theory. All are normalized over the control group. See tables in Appendix for details on the variables included in each index. Mother's interview outcomes: Indices are from measures self-reported by the respondent. Observed outcomes: The Observed child development index (column 5) is based on a selection of items adapted from the Ages and Stages Questionnaire (ASQ) and the Oxford Neurodevelopment Assessment (Ox-NDA). The assessment was administered by the surveyor to the child during the survey. 1,203/1,258 children were available and received parental consent to participate. Children for which the surveyors were unable to asses more than 50% of the test items are dropped (N=17). "Child Babbles" (column 6) is a dummy equal to 1 if the surveyor observed the child babbling (at least one syllable) at some point during the home visit. The outcome was added mid-data collection, hence, it is missing for some households. LENA outcomes: Given financial constraints, only a random subset of households could be included in the LENA measurement. 900 households were sampled to receive a LENA for a day, and 225 of those were sampled for a second day of recording. For households which kept the LENA device for two days, only the first day of recording is kept in the analysis presented in this table. The analysis is further restricted to recording times between 10 a.m. to 7 p.m. (this excludes 10/785 LENA day 1 recordings which have less than 9 hours (rounded up) of recording). 1 audio did not have the breakdown of adult words per gender. Child vocalizations are estimated by the LENA software and include words, babbles, and pre-speech communicative sounds or "protophones" such as squeals, growls, or raspberries. We report Romano-Wolf stepdown adjusted p-value to adjust for multiple hypothesis testing. Robust standard errors in parenthesis, p-values in curly brackets.

Table 5: Treatment Effects by Age Groups, 6 to 8 Months After Intervention

	Mother's Interview		Observed	
	Child language score	Child gestural communication index	Child development index	Child babbles (=1)
	(1)	(2)	(3)	(4)
Panel A: 5 months or younger				
Treatment	0.012	-0.027	0.092	0.100
	(0.067)	(0.056)	(0.109)	(0.051)
	{0.859}	{0.636}	{0.398}	{0.049}
Control mean	-0.96	-0.72	-0.82	0.10
Observations	334	334	308	183
Panel B: 6-9 months				
Treatment	0.118	0.154	0.076	0.062
	(0.060)	(0.063)	(0.076)	(0.054)
	{0.050}	{0.015}	{0.321}	{0.256}
Control mean	-0.25	-0.35	-0.00	0.40
Observations	389	389	368	296
Panel C: 10-14 months				
Treatment	0.067	0.157	-0.116	0.021
	(0.062)	(0.096)	(0.083)	(0.054)
	{0.280}	{0.101}	{0.163}	{0.697}
Control mean	0.60	0.45	0.42	0.74
Observations	335	335	317	248
Panel D: 15 months or older				
Treatment	0.261	0.073	0.136	0.047
	(0.088)	(0.127)	(0.127)	(0.056)
	{0.003}	{0.567}	{0.283}	{0.402}
Control mean	1.23	1.24	0.73	0.86
Observations	200	200	191	161
Control mean (all)	0.00	0.00	0.00	0.51
Observations (all)	1,258	1,258	1,184	888
Controls	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes

Note: Endline data. Each column presents the results from regressing the outcomes (top) on dummies for treatment interacted with age groups (coefficients presented in the table rows), child age group (5 months or younger, 6-9 months old, 10-14 months old, 15 months or older dummies), controls as listed in Table 4, and clinic fixed effects. See Table 4 for details on each outcome, Appendix Figure A.6 for distribution of children's ages, and Appendix Figure A.4 for the distribution of each outcome per age group. Robust standard errors in parenthesis, p-values in curly brackets.

Table 6: Treatment Effects on Newly-Informed Mothers: Evidence From the 'Endline' Intervention

	Adult words per min	Female adult words per min	Male adult words per min	Conversational turn count per min	% meaningful speech
	(1)	(2)	(3)	(4)	(5)
2nd day (post-intervention)	1.827	1.389	0.509	0.015	0.927
	(0.864)	(0.728)	(0.285)	(0.016)	(0.487)
	{0.036}	{0.058}	{0.076}	{0.352}	{0.058}
Mean Pre-intervention (Day 1)	20.12	16.46	3.67	0.42	16.65
Mean Post-intervention (Day 2)	21.81	17.67	4.14	0.44	17.62
Controls	Yes	Yes	Yes	Yes	Yes
Household Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	372	371	371	372	372

Note: LENA days 1 and 2 recording data. Unit: recording. The sample is restricted to recordings from control households sampled to keep a LENA device for two days at endline. Those households have two recordings. Before the 2nd day of recording, households were shown the intervention video. 192 of the 225 households sampled for a 2nd day of recording consented to the recording and saw the intervention video. The analysis is restricted to recording times between 10 a.m. to 7 p.m.. 186/192 households had 2 complete audio recordings. Adult words per minute by gender is not available for one recording. Regressions include controls for the day of the week the audio was recorded (dummies), the total time (min) the shirt/LENA device was removed from the child, and the total time (min) the child was held on someone's back while wearing the device. Household fixed effects are included. Conversational turn count is the number of alternations between the focal child and adults in the vicinity.% meaningful speech is the share of the audio categorized as sounds from the focal child or speech from adults or other children near the focal child. For further details on the LENA outcomes, please refer to Appendix, Section A.2. Robust standard errors in parenthesis, p-values in curly brackets.

Table 7: Cost-Effectiveness Calculations

		At-Scale Costs		RCT costs	
	SD Effects	Unit Cost ($)	Cost/SD ($)	Unit Cost ($)	Cost/SD ($)
Panel A: Full intervention effects size					
1 Mother-reported measures	0.10	0.45	4.52	3.01	30.26
2 Observed measures	0.04	0.45	10.01	3.01	66.93
Panel B: $\frac{1}{2}$ intervention effects size					
3 Mother-reported measures	0.05	0.45	9.05	3.01	60.52
4 Observed measures	0.02	0.45	20.01	3.01	133.85
Panel C: $\frac{1}{4}$ intervention effects size					
5 Mother-reported measures	0.02	0.45	18.09	3.01	121.03
6 Observed measures	0.01	0.45	40.02	3.01	267.71

Note: At-scale costs would only include the cost of printing each calendar and delivering them to health clinics ($0.45). The RCT costs include the labor cost of hiring an IPA surveyor to go to clinics and only give out calendars and show the video on their tablet and attendant management costs. Mother-reported and observed outcomes are reported in main text Table 4. For mother-reported measures of development, we use the child language score and gestural communication index. For the observed measures of development, we use the observed child cognitive index and the LENA-measured child vocalizations per minute. Following Verguet et al. (2022), we take the average of the language and cognitive/gestural development effects to get the average standard deviation effect of the intervention.

Appendix A

A.1 Appendix Figures

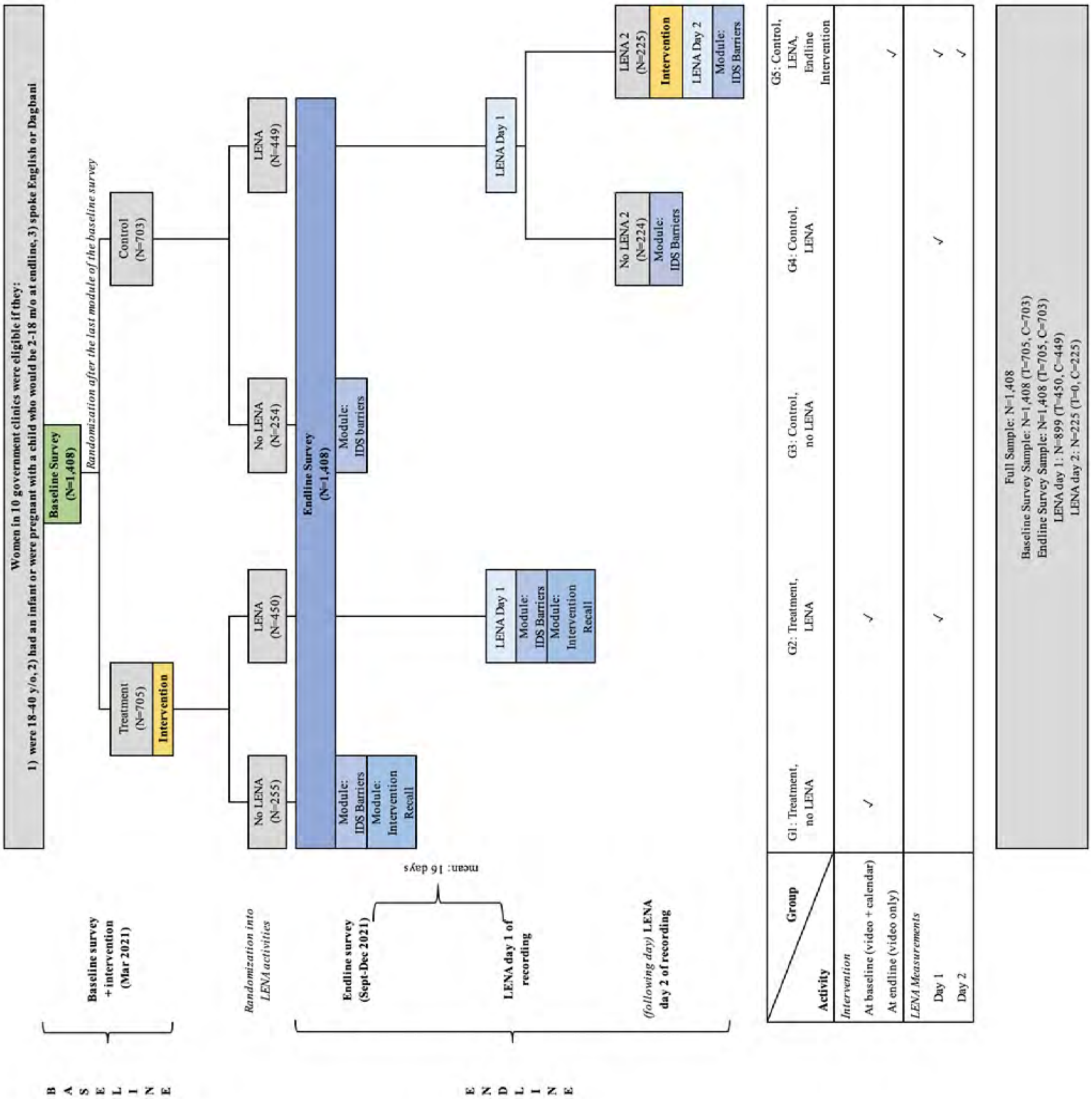
Figure A.1: Calendar for Treated Respondents

Note: 50% of the sample (N=705) watched the video and received an IDS-themed calendar at the end of the baseline survey. The calendar displays a star at the end of each week. Respondents were encouraged to fill in the stars next to each week in the calendar if they conversed with their child each day that week.

Figure A.2: Calendar for Control Respondents

Note: 50% of the sample (N=703) received a regular calendar at the end of the baseline survey as a token of gratitude for participating in the survey. Control respondents did not see the IDS-themed video.

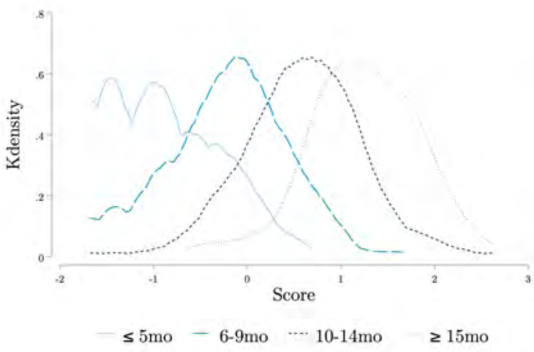
Figure A.3: Experimental Design and Timeline



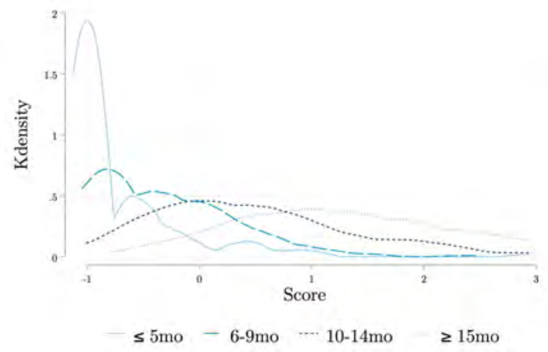
Note: See main text Section 2 for further details on the study design and timeline. On average, 6.4 months elapsed between the baseline and endline surveys.

Figure A.4: Distribution of Child Language and Development Scores (Control Group)

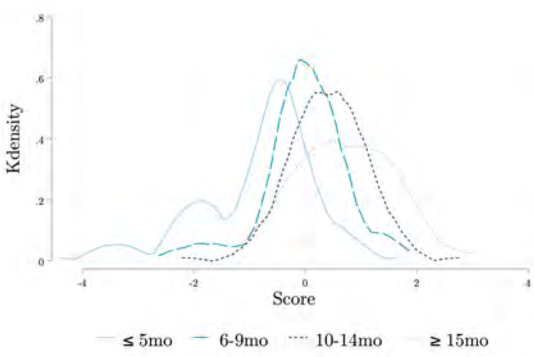
(a) Panel A: Reported Child Language Score



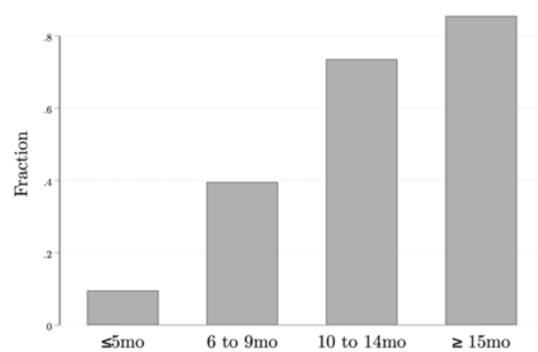
(b) Panel B: Reported Child Gestural Communication Index



(c) Panel C: Observed Child Development Index

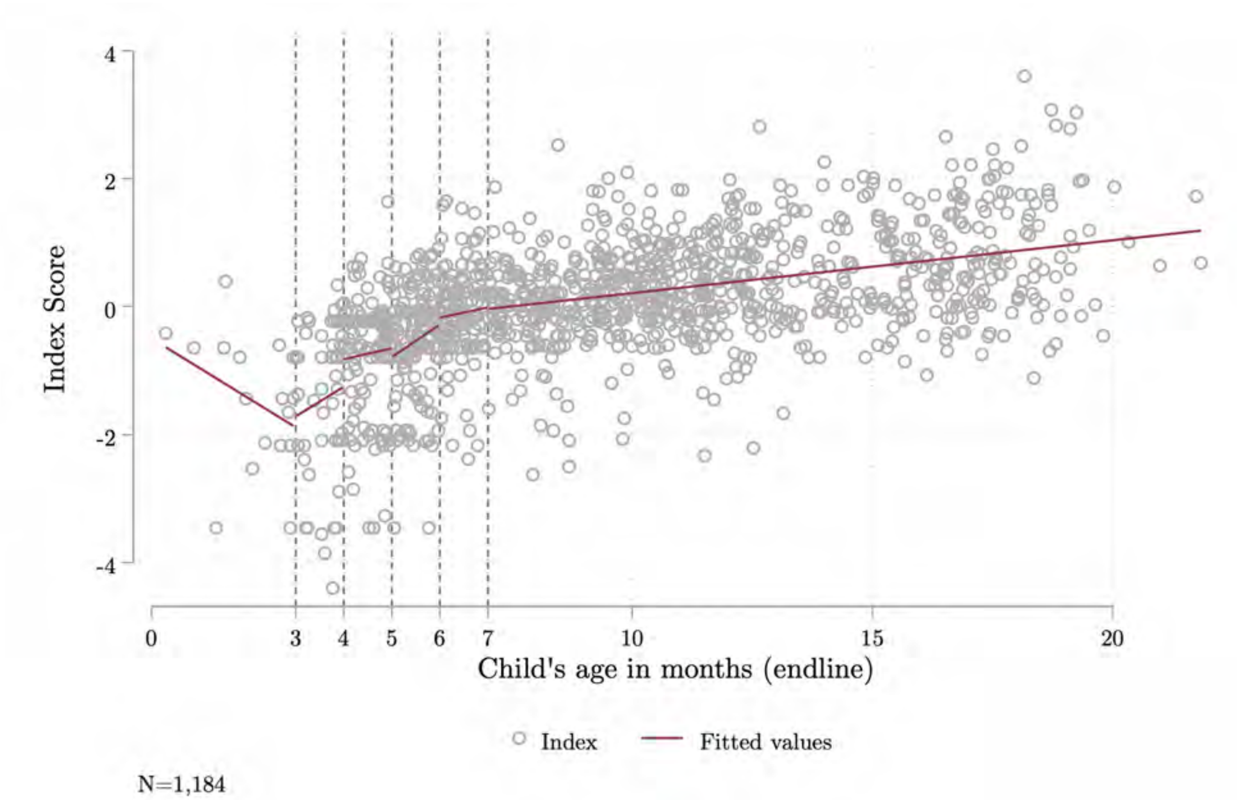


(d) Panel D: Observed Child Babbles (=1)



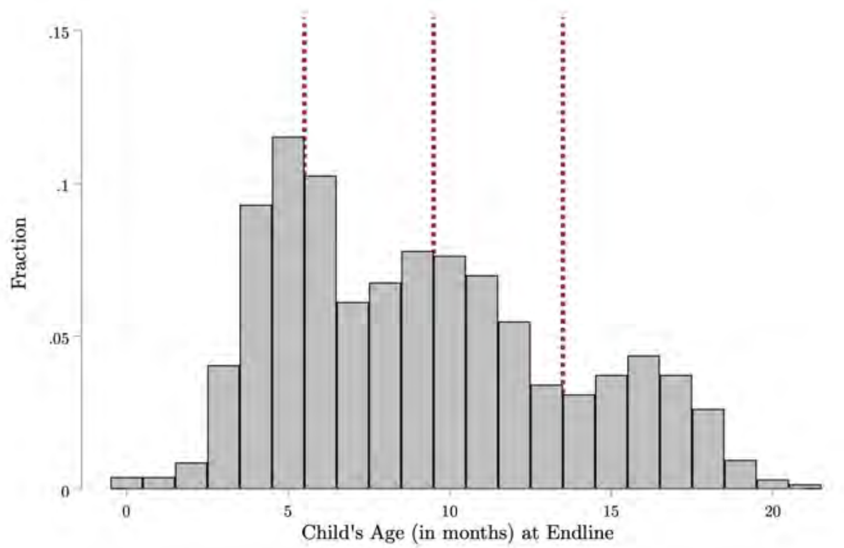
Note: Endline data. Distributions are plotted for children in the control group only. For details on outcomes, please refer to Table 4. Treatment effects for each age groups are shown in Table 5.

Figure A.5: Observed Child Development Index by Child Age



Note: Endline data. The index is an Anderson index, normalized over the control group. It is based on a selection of items adapted from the Ages and Stages Questionnaire (ASQ) and the Oxford Neurodevelopment Assessment (Ox-NDA). The assessment was administered by the surveyor to the child during the survey. See Table A.7 for the list of components included in the index.

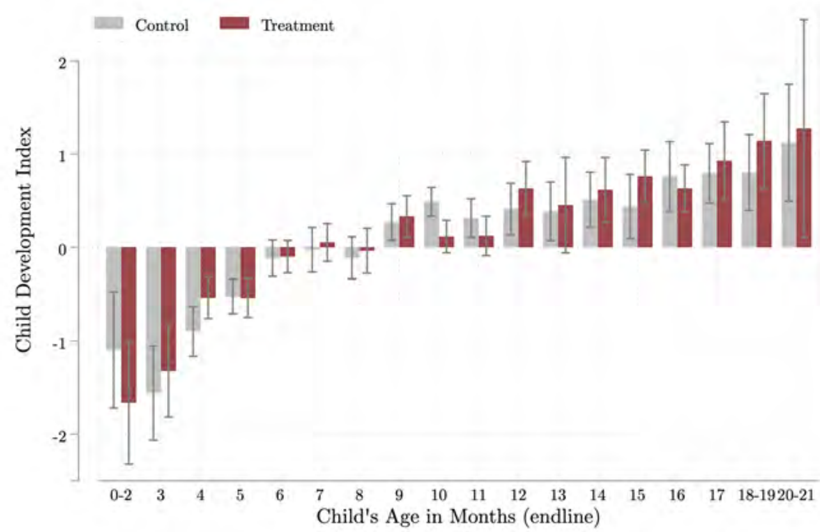
Figure A.6: Distribution of Children's Ages at Endline



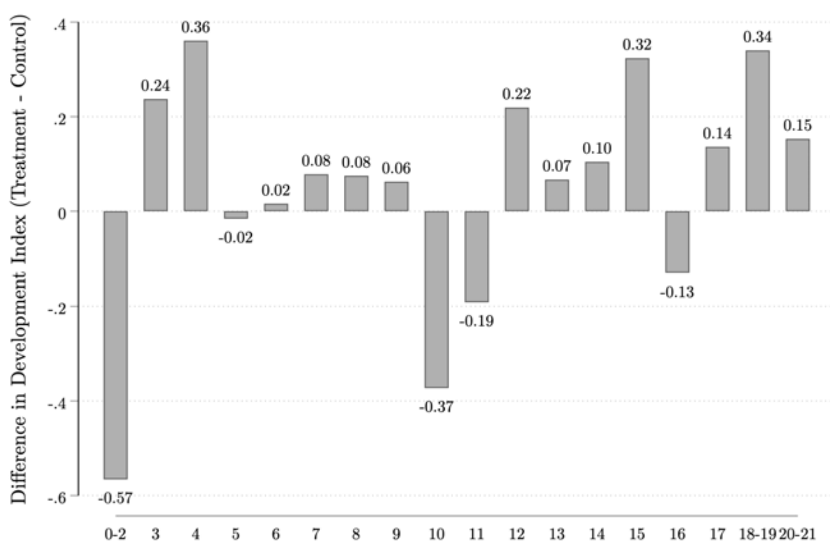
Note: Endline data. N=1,258/1,408 households participated in the endline survey. Average infant age is 9.6 months in the treatment group and 9.5 months in the control group. The dashed red lines delimit the 4 groups for which we present the disaggregated treatment effects in Table 5: 5mo or less, 6 to 9mo old, 10 to 14mo old, and 15 months or more.

Figure A.7: Observed Child Development Index by Child's Age (in Months)

(a) Panel A: Mean by Child's Age (in Months) and Group

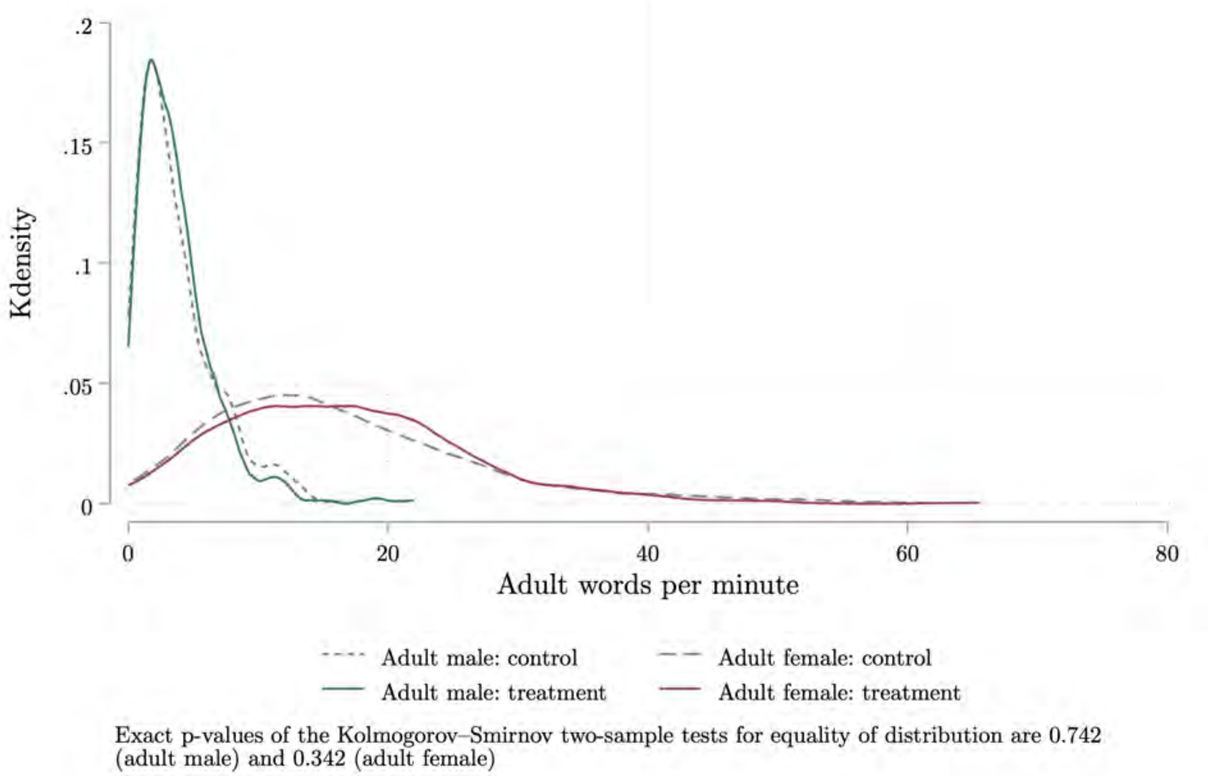


(b) Panel B: Difference in Means (Treatment - Control)



Note: Panel A: the bars show the control and treatment group means by age group with 95% confidence intervals. Panel B: the bars show the group difference in means (treatment minus control) by child's age. We pool children aged 0-2 months, 18-19 months, and 20-21 months as there are few observations in those groups (see Figure A.6 for children's age distribution at endline). The Observed child development index is an Anderson index, normalized over the control group. It is based on a selection of items from the Ages & Stages Questionnaire (ASQ) and the Oxford Neurodevelopment Assessment (Ox-NDA) (see Table A.8 for further details).

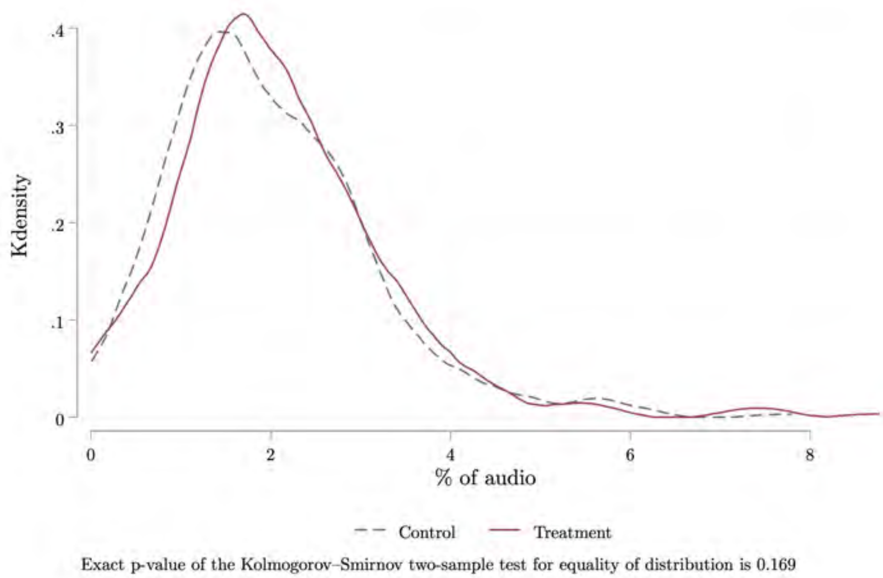
Figure A.8: LENA-measured Adult Words per Minute by Speaker Gender and Treatment



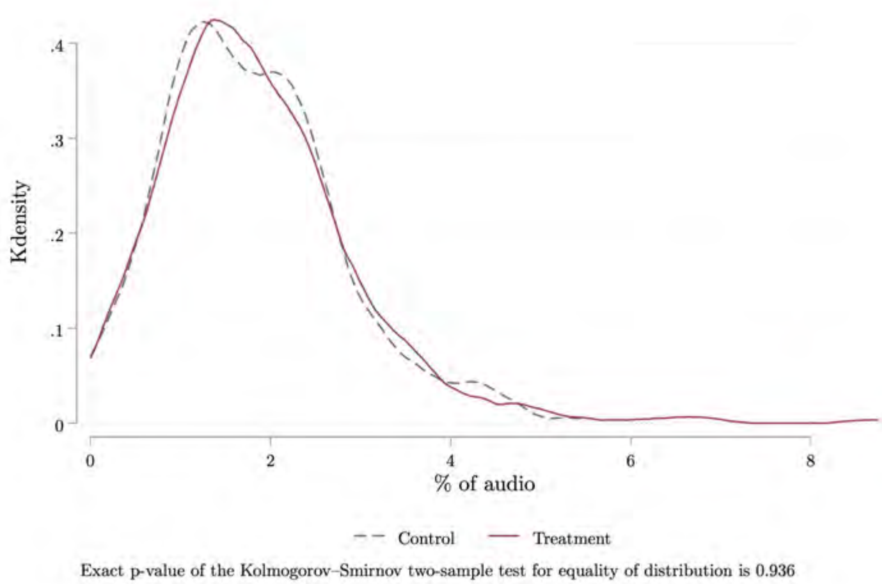
Note: LENA day 1 recording data. N=774 recordings (1 recording does not have the breakdown of adult words by gender). Please refer to Table A.2 for details on the sample. The LENA software estimates the number of words spoken by post-pubescent males and females in the child's vicinity. Adult word per minute is the estimated total number of words spoken by adults during the recording divided by the length of the recording. For further details on the LENA outcomes, please refer to Section A.2.

Figure A.9: LENA Measurements: Child Sounds by Treatment

(a) Panel A: Vocalizations (% of total audio time)



(b) Panel B: Non-Speech Sounds (% of total audio time)



Note: LENA day 1 data. N=775 recordings. Please refer to Table A.2 for details on the sample. The LENA software categorizes the focal child sound segments into 1) **vocalizations** (including words, babbles, and pre-speech communicative sounds or "protophones" such as squeals, growls, or raspberries) and 2) **non-speech** sounds (including fixed signals and vegetative sounds such as breathing or crying). For further details on the LENA outcomes, please refer to Section A.2.

A.2 Appendix Tables

Table A.1: List of Government Health Facilities

Name	District
Choggu RCH	Sagnarigu Municipal
Kalpohini Health Centre	Sagnarigu Municipal
Kanvilli Health Centre	Sagnarigu Municipal
Malshegu CHPS	Sagnarigu Municipal
Sagnarigu Health Centre	Sagnarigu Municipal
Bilpela Health Centre	Tamale Metropolitan
Moshie Zongo Health Centre	Tamale Metropolitan
Tamale Central Hospital	Tamale Metropolitan
Tamale SDA Hospital	Tamale Metropolitan
Tamale West Hospital	Tamale Metropolitan

Note: List of health facilities in Tamale (Northern Ghana) where women were recruited when coming for prenatal or postnatal checkups.

Table A.2: Attrition and Endline Survey Status

	All				Control			Treatment			T=C
	Mean	SD	Count	N	Mean	SD	Count	Mean	SD	Count	P-value
Endline Survey											
Dead	0.00	0.03	1	1,408	0.00	0.04	1	0.00	0.00	0	0.317
Had COVID symptoms	0.00	0.04	2	1,408	0.00	0.00	0	0.00	0.05	2	0.157
Refused to participate	0.00	0.07	7	1,408	0.00	0.07	3	0.01	0.08	4	0.708
Moved temporarily	0.01	0.08	8	1,408	0.00	0.07	3	0.01	0.08	5	0.481
Unavailable (other reason)	0.01	0.12	20	1,408	0.02	0.12	11	0.01	0.11	9	0.648
Ineligible	0.01	0.12	21	1,408	0.01	0.10	7	0.02	0.14	14	0.125
Moved permanently	0.02	0.15	33	1,408	0.03	0.16	18	0.02	0.14	15	0.592
Not found	0.04	0.20	58	1,408	0.04	0.19	27	0.04	0.21	31	0.600
Completed survey	0.89	0.31	1,258	1,408	0.90	0.30	633	0.89	0.32	625	0.398
Age of child at endline (months)	9.58	4.41		1,258	9.52	4.35		9.64	4.47		0.651
Number of Respondents				1,258			633			625	
Child Assessment											
Consented to child test	1.00	0.07	1,252	1,258	1.00	0.06	631	0.99	0.08	621	0.406
Child available (if consented)	0.96	0.19	1,203	1,252	0.97	0.18	611	0.95	0.21	592	0.172
LENA Recording Day 1											
Refusal (survey or LENA)	0.02	0.15	22	899	0.03	0.17	14	0.02	0.13	8	0.194
Not available/eligible main survey	0.10	0.30	92	899	0.09	0.29	41	0.11	0.32	51	0.277
Complete	0.87	0.33	785	899	0.88	0.33	394	0.87	0.34	391	0.698
Number of Respondents				899			449			450	
If complete: kept in analysis	0.99	0.11	775	785	0.99	0.10	390	0.98	0.12	385	0.517
LENA Recording Day 2											
Missing/Lost	0.00	0.07	1	225	0.00	0.07	1				
Refusal (survey or LENA)	0.04	0.19	8	225	0.04	0.19	8				
Not available/eligible main survey	0.11	0.31	24	225	0.11	0.31	24				
Complete	0.85	0.35	192	225	0.85	0.35	192				
Number of Respondents				225			225				
If complete: kept in analysis	0.98	0.14	188	192	0.98	0.14	188				

Note: Endline data. Due to monetary constraints, only a sub-sample of respondents were randomized to receive a LENA device (N=900). A subsample of the control group was randomized to keep the LENA device for two days instead of only one (N=225). Before the start of the second day of recording, those respondents were shown the intervention video (see Section 2 and Figure A.3 for further details on the study design and timeline). In the panels "LENA Day 1" and "LENA Day 2", "If complete: kept in analysis" is a dummy equal to 1 if the audio has 9 hours (rounded up) of recording between 10 a.m. and 7 p.m., and, hence, is kept in the analysis.

Table A.3: Reported Barriers to IDS

	(1) Pure Control	(2) Main Intervention (6-8 mo ago)	(3) Endline Intervention (day before)	(4) P-value Main Int = Endline Int	(5) P-value Control = Endline Int
=1 if it could be a barrier to other families					
It's hard to remember/make a habit	0.35	0.35	0.29	0.130	0.185
It's mocked/frowned upon in the community	0.32	0.30	0.28	0.661	0.311
It's not clear it makes a difference	0.28	0.20	0.07	<0.001	<0.001
Too busy/Not enough time	0.08	0.06	0.01	<0.001	<0.001
Parents are too preocuppied or unhappy	0.01	0.02	0.01	0.417	0.884
Other	0.02	0.00	0.00	0.083	0.003
Lack of patience	0.00	0.01	0.01	0.915	0.478
Laziness	0.01	0.00	0.00	0.157	0.083
Child may grow to be disrespectful	0.00	0.00	0.00	0.083	0.318
Parent's personality: shy, not talkative	0.00	0.00	0.00	0.318	0.318
Lack of reaction/responsiveness from the child	0.00	0.00	0.00	0.318	0.318
No barriers to IDS cited	0.43	0.46	0.48	0.660	0.251
=1 if could be the main barrier to other families					
It's hard to remember/make a habit	0.16	0.19	0.18	0.828	0.448
It's mocked/frowned upon in the community	0.18	0.17	0.29	0.002	0.003
It's not clear it makes a difference	0.08	0.07	0.02	0.002	<0.001
Too busy/Not enough time	0.06	0.04	0.01	<0.001	<0.001
Parents are too preocuppied or unhappy	0.01	0.01	0.01	0.915	0.688
Other barrier (specify)	0.13	0.11	0.03	<0.001	<0.001
=1 if it's a barrier to respondent and her family					
It's hard to remember/make a habit			0.31		
It's mocked/frowned upon in the community			0.32		
It's not clear it makes a difference			0.16		
Parents are too preocuppied or unhappy			0.02		
No barriers to IDS cited			0.37		
Observations	424	615	191		

Note: Endline data. Respondents were asked about barriers that may prevent families from talking to their babies. Questions were asked at the end of the endline survey if the household did not receive a LENA device, or after the last day of recording if the household received a LENA (see timing of the "Module: IDS Barriers" in the design chart Figure A.3). The endline intervention sample received the intervention between the 1st and 2nd day of LENA recording and the IDS barrier questions were asked after the second day of recording. Column 1 presents the means for respondents in the pure control group (who never received the intervention), column 2 those for respondents who received the main intervention, at baseline (their views incorporate their experience with IDS over the past 6 to 8 months between the intervention and the endline survey), and column 3 for respondents who received the intervention at endline (their views incorporate their experience with IDS over the past 24 hours). The last two columns report the p-values from t-tests comparing the means between the respondents who received the intervention at baseline vs at endline (column 4) and respondents who did not receive the intervention (pure control group) vs those who received it at endline. Questions in the last panel "=1 if it's a barrier to respondent and her family" were only asked to those who received the endline intervention.

Table A.4: Treatment Effects on Mother-reported Parental Beliefs

	Treatment Effect			Control Group		
	Coefficient (1)	SE (2)	P-value (3)	Mean (4)	SD (5)	N (6)
Believes should talk to child from birth	0.104	0.027	0.000	0.33	0.47	1,257
Outcomes in the index						
Age (in mo) when babbles/makes noise in response	-0.145	0.483	0.764	7.37	8.11	1,256
Age (in mo) when says meaningful words	-1.638	0.638	0.010	19.96	12.22	1,248
Age (in mo) for talking to child	-1.170	0.486	0.016	5.55	7.94	1,257
Age (in mo) for telling stories to child	-2.530	0.858	0.003	18.30	15.90	1,228
Age (in mo) for talking to child in full sentences	-1.507	1.346	0.263	26.00	24.98	1,250
Importance to brain development of talking in full sentences to a child (/10)	0.024	0.112	0.832	8.72	2.02	1,252
How strongly do you agree with: (1=strongly disagree to 4= strongly agree)						
Intelligence is set at birth	0.053	0.071	0.451	3.08	1.27	1,250
Infants learn little language in their 1st year	-0.005	0.053	0.931	3.54	0.94	1,255
Parents shouldn't talk back to babble	0.002	0.074	0.981	3.15	1.31	1,257
Children learn more from overhearing than being spoken to	-0.100	0.060	0.094	3.49	1.00	1,249
Adults can't have conversations with babies who can't talk	-0.097	0.074	0.191	2.03	1.36	1,253

Note: Each line reports the result of a different regression for which the outcome (variable indicated on the left) is regressed on a dummy equal to 1 if the household received the main intervention, at baseline (treatment group). Column 1 reports the coefficient on treatment, column 2 the standard error, and column 3 the p-value. Columns 4 and 5 report the control group mean and standard deviation. Column 6 reports the number of observations. "Believes should talk to child from birth" is not included in the mother's belief index in Table 4, but all other outcomes are. For outcomes in the panel "How strongly do you agree with the following statements:", respondents were asked to choose from a 4-point Likert scale (strongly disagree (1), somewhat disagree, somewhat agree, strongly agree (4)). As for the main text Table 4, all regressions include clinic fixed effects and controls for the child's age (in days), survey date, and surveyor gender.

Table A.5: Treatment Effects on Mother-reported Parental Verbal Inputs

	Treatment Effect			Control Group		
	Coefficient (1)	SE (2)	P-value (3)	Mean (4)	SD (5)	N (6)
Outcomes in the index						
In the last 4 wks, how often did you: (0=never to 5=daily)						
Talk to child while doing an activity w/ child around	0.165	0.103	0.109	2.05	1.88	1,256
Describe things to child when walking	0.152	0.096	0.112	2.11	1.76	1,256
Pointed, named object and asked child to repeat	0.141	0.096	0.142	1.65	1.85	1,256
In the last 4 weeks, did any adult:						
Sang to child	-0.012	0.019	0.521	0.88	0.32	1,254
Read to/looked at book with child	0.065	0.028	0.019	0.44	0.50	1,256
Told story to child	0.045	0.026	0.089	0.31	0.46	1,251
Played with child	0.005	0.007	0.500	0.98	0.14	1,256
Decribed things to child	0.017	0.025	0.489	0.69	0.46	1,257
As percent of total play time:						
% of time playing w/ adult	0.921	0.903	0.308	31.37	16.29	1,258

Note: Each line reports the result of a different regression for which the outcome (variable indicated on the left) is regressed on a dummy equal to 1 if the household received the main intervention, at baseline (treatment group). Column 1 reports the coefficient on treatment, column 2 the standard error, and column 3 the p-value. Columns 4 and 5 report the control group mean and standard deviation. Column 6 reports the number of observations. For outcomes in the panel "In the last 4 weeks, how often did you...", respondents were asked to choose from a 6-point Likert scale (never (0), rarely, a few times, once a week, multiple times a week, daily (5)). As for the main text Table 4, all regressions include clinic fixed effects and controls for the child's age (in days), survey date, and surveyor gender.

Table A.6: Treatment Effects on Mother-reported Child Language Score Components

	Treatment Effect			Control Group		
	Coefficient (1)	SE (2)	P-value (3)	Mean (4)	SD (5)	N (6)
# of words in list child understands	0.475	0.180	0.008	6.95	4.99	1,258
# of words in list child says	0.305	0.124	0.014	1.04	2.42	1,258
# of phrases in list child understands	0.096	0.055	0.078	1.40	1.44	1,258
Attempt to say words (yes/no)	0.044	0.021	0.038	0.42	0.49	1,257
How often does child: (1=not yet to 3=often)						
Repeat/imitate words	0.046	0.032	0.154	1.49	0.76	1,258
Name/label things	0.016	0.025	0.535	1.18	0.50	1,256

Note: Each line reports the result of a different regression for which the outcome (variable indicated on the left) is regressed on a dummy equal to 1 if the household received the main intervention, at baseline (treatment group). Column 1 reports the coefficient on treatment, column 2 the standard error, and column 3 the p-value. Columns 4 and 5 report the control group mean and standard deviation. Column 6 reports the number of observations. The Child language score is computed using questions derived from a version of the MacArthur-Bates Communicative Development Inventories Words and Gestures (MB-CDI-WG) adapted to Ghana by the Harvard Laboratory for Developmental Studies (Duflo et al., 2024). The adapted inventory includes a list of words and sentences across various domains. Mothers were asked if the child understands and/or can say (either spontaneously or upon prompting) the items in this list: *words*: "Ouch (wa aish)", "Toy (biebi)", "Bread (paa nu)", "Toffee (toffe)", "Biscuit (biscuit)", "Shoe (namda)", "Chair (ku'g)", "Cup (copu)", "Feed (dima/dibu)", "Beautiful (vilem)", "Egg (galili)", "Baby (bilegu)", "Bath (kom subu)", "Don't (dining)", "Shh (shh)", "Thank you (npahiya)", *phrases*: "Give me a hug (awa wa tuu)", "Go get _ (chamtikpahi)", "Good girl/boy (bi sung)", and "Spit it out (tuhi bahi)". Additionally, mothers were asked questions about whether the child started to talk. For the child language score, we use binary versions of the variables presented above (dummies for each word/phrase indicating whether the child understands/says it and for whether the child attempts to speak, to imitate words, and to name/label things). We compute the child language score using Item Response Theory which involves estimating a one-parameter logistic model on the mother's responses to the adapted MB-CDI-WG, where the model assigns a difficulty level to each question and, then, a latent trait to each individual based on their answers to the questions adjusting for the question's difficulty level. As for the main text Table 4, all regressions include clinic fixed effects and controls for the child's age (in days), survey date, and surveyor gender.

Table A.7: Treatment Effects on Mother-reported Child Gestural Communication

	Treatment Effect			Control Group		
	Coefficient (1)	SE (2)	P-value (3)	Mean (4)	SD (5)	N (6)
Outcomes in the index						
How often does child: (1=not yet to 3=often)						
Give toy when holding it	0.057	0.038	0.133	1.82	0.79	1,256
Point at interesting things	0.022	0.040	0.585	1.61	0.84	1,258
Wave when someone leaves	0.026	0.035	0.446	1.60	0.82	1,257
Shake head for no	0.081	0.038	0.035	1.72	0.84	1,258
Gesture shh	0.049	0.029	0.096	1.22	0.55	1,257
Blows kisses	0.044	0.034	0.195	1.40	0.65	1,257

Note: Each line reports the result of a different regression for which the outcome (variable indicated on the left) is regressed on a dummy equal to 1 if the household received the main intervention, at baseline (treatment group). Column 1 reports the coefficient on treatment, column 2 the standard error, and column 3 the p-value. Columns 4 and 5 report the control group mean and standard deviation. Column 6 reports the number of observations. Components of the Mother-reported child gestural communication index are based on a selection of items from the "First Communicative Gestures" from the MB-CDI-WG. Respondents were asked to choose from a 3-point Likert scale (not yet (1), sometimes, and often (3)). As for the main text Table 4, all regressions include clinic fixed effects and controls for the child's age (in days), survey date, and surveyor gender.

Table A.8: Treatment Effect on Observed Child Development

	Treatment Effect			Control Group		
	Coefficient (1)	SE (2)	P-value (3)	Mean (4)	SD (5)	N (6)
Outcomes in the index						
Child assessment: (1=worst to 3=best)						
Watches mother move	0.031	0.032	0.326	2.78	0.59	1,172
Watches toy placed in front	0.009	0.023	0.706	2.90	0.42	1,179
Child assessment: (1=worst to 4=best)						
Identifies spoon correctly when asked	-0.034	0.059	0.562	2.25	1.04	1,141
Imitates or tries to imitate bi-syllabic words	-0.010	0.035	0.763	1.28	0.62	1,132
Reacts to name when playing	0.091	0.071	0.202	2.74	1.27	1,174
Stops reaching for toy when told no	-0.014	0.061	0.823	1.76	1.09	1,116
Uses or mimics words in play context	0.007	0.029	0.814	1.19	0.49	1,139
Babbles or attempts to when prompted	0.058	0.046	0.207	1.40	0.81	1,142
Combines word and gesture (correctly or not)	-0.043	0.046	0.351	1.45	0.86	1,139

Note: Each line reports the result of a different regression for which the outcome (variable indicated on the left) is regressed on a dummy equal to 1 if the household received the main intervention, at baseline (treatment group). Column 1 reports the coefficient on treatment, column 2 the standard error, and column 3 the p-value. Columns 4 and 5 report the control group mean and standard deviation. Column 6 reports the number of observations. The assessment was administered to the child by the surveyor during the survey. The first two tasks were adapted from the problem-solving ASQ-3 (2 months) module and the others were adapted from the language and cognitive Ox-NDA modules. For the ASQ-like items, each task was evaluated from 1 (lowest score) to 3 (highest score). For the Ox-NDA-like items, each task was evaluated using a scale from 1 (lowest score) to 4 (highest score). Instead of giving a score, surveyor could indicate they were unable to assess the child (because the infant was out of sight, sleeping, crying, became too agitated, etc.) and those are recoded as missing. As for the main text Table 4, all regressions include clinic fixed effects and controls for the child's age (in days), survey date, and surveyor gender.

Table A.9: Robustness of Treatment Effects, 6 to 8 Months After Intervention

	Mother's Interview				Observed		LENA	
	Mother's belief index	Mother's behavior index	Child language score	Child gestural communication index	Child development index	Child babbles (= 1)	Female adult words per min	Child vocalizations per min
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Without Controls								
Treatment	0.125	0.126	0.115	0.108	0.051	0.054	-0.008	0.047
	(0.060)	(0.056)	(0.054)	(0.055)	(0.056)	(0.033)	(0.720)	(0.060)
	{0.036}	{0.026}	{0.034}	{0.051}	{0.368}	{0.098}	{0.991}	{0.428}
Control mean	0.00	0.00	0.00	0.00	0.00	0.51	16.61	1.53
Controls	No	No	No	No	No	No	No	No
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,258	1,258	1,258	1,258	1,184	888	774	775
Panel B: With Clinic-Day Fixed Effects								
Treatment	0.137	0.133	0.108	0.103	0.034	0.061	-0.037	0.035
	(0.058)	(0.055)	(0.035)	(0.041)	(0.049)	(0.029)	(0.733)	(0.061)
	{0.018}	{0.016}	{0.002}	{0.012}	{0.491}	{0.035}	{0.959}	{0.570}
Control mean	0.00	0.00	0.00	0.00	0.00	0.51	16.61	1.53
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	No	No	No	No	No	No	No	No
Clinic-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,258	1,258	1,258	1,258	1,184	888	774	775
Panel C: With Surveyor Fixed Effects								
Treatment	0.180	0.119	0.085	0.073	0.013	0.032	-0.044	0.040
	(0.054)	(0.051)	(0.031)	(0.039)	(0.047)	(0.027)	(0.710)	(0.060)
	{0.001}	{0.021}	{0.006}	{0.059}	{0.783}	{0.232}	{0.951}	{0.508}
Control mean	0.00	0.00	0.00	0.00	0.00	0.51	16.61	1.53
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Surveyor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,258	1,258	1,258	1,258	1,184	888	774	775

Note: Endline and LENA day 1 recording data. This table presents the results from running the main specification (presented in Table 4) without any control (Panel A); replacing clinic FE by clinic-day fixed effects (Panel B); and adding surveyor fixed effects (Panel C). Since 3 surveyors completed 20 or fewer endline surveys each, we grouped them as one surveyor when including Surveyor Fixed Effects (in Panel C). Robust standard errors in parenthesis, p-values in curly brackets.

Table A.10: Treatment Effects on Observed Child and Parental Behavior (LENA Outcomes)

	Treatment Effect			Control Group		
	Coefficient (1)	SE (2)	P-value (3)	Mean (4)	SD (5)	N (6)
Count per minute						
Adult words per min	0.070	0.786	0.929	20.37	11.75	775
Female adult words per min	0.131	0.697	0.851	16.61	10.39	774
Male adult words per min	-0.058	0.215	0.788	3.76	3.00	774
Focal child vocalizations per min	0.052	0.058	0.377	1.53	0.82	775
Conversational turns per min	0.004	0.017	0.815	0.41	0.25	775
Length as % of audio						
Female adult speech	0.094	0.302	0.754	7.32	4.49	774
Male adult speech	0.021	0.102	0.835	1.88	1.42	774
Other children speech	0.133	0.153	0.383	3.46	2.13	774
Focal child sounds	0.163	0.148	0.272	4.14	2.03	774
Focal child vocalizations	0.090	0.082	0.274	2.00	1.14	774
Focal child non-vocalizations	0.061	0.074	0.412	1.79	0.98	774
Faint or overlapping speech	0.750	1.108	0.499	39.07	15.67	775
Silence	-0.336	1.120	0.764	29.79	15.31	775
Background noise	-1.245	0.651	0.056	19.32	9.63	775
Electronic media (TV, radio, etc.)	0.157	0.663	0.813	9.37	10.39	775

Note: LENA day 1 recording data. Each line reports the result of a different regression for which the outcome (variable indicated on the left) is regressed on a dummy equal to 1 if the household received the main intervention, at baseline (treatment group). Column 1 reports the coefficient on treatment, column 2 the standard error, and column 3 the p-value. Columns 4 and 5 report the control group mean and standard deviation. Column 6 reports the number of observations. Given financial constraints, only a random subset of households could be included in the LENA measurement (N=900 households sampled). For households sampled to keep the LENA device for two days, only the first day recording is kept in the analysis. The analysis is further restricted to recording times between 10 a.m. to 7 p.m. (this excludes 10/785 LENA day 1 recordings which have less than 9 hours (rounded up) of recording). Some variables are missing for one recording. The focal child sounds as well as adults' and other children's speech only include sounds from those sources categorized as near by the LENA algorithm. Faint or overlapping speech includes faint female and male adults' sounds, other children's vocalizations, and overlapping speech (both near and faint). Background noise and electronic media each include both near and far sounds from those sources. As in columns 7 and 8 of main text Table 4, regressions include clinic fixed effects and controls for the child's age in days, the day of the week the audio was recorded (dummies), the total time (min) the shirt/LENA device was removed from the child, the total time (min) the child was held on someone's back while wearing the device, and the household size. For further details on the LENA outcomes, please refer to Section A.2.

Table A.11: Treatment Effect Split by Susceptibility to Experimenter Demand Effects

	Mother's Interview				LENA
	Mother's belief index	Mother's behavior index	Child language score	Child gestural communication index	Female adult words per min
	(1)	(2)	(3)	(4)	(5)
Treatment	0.181	0.143	0.112	0.104	-0.344
	(0.060)	(0.059)	(0.037)	(0.044)	(0.745)
	{0.003}	{0.015}	{0.003}	{0.019}	{0.645}
Treatment x did not associate intervention w/ surveyor	-0.260	-0.033	-0.030	-0.013	2.498
	(0.121)	(0.103)	(0.052)	(0.069)	(1.124)
	{0.032}	{0.749}	{0.566}	{0.850}	{0.027}
P-val total effect for did not associate	0.509	0.278	0.101	0.172	0.051
Control mean	-0.00	-0.00	-0.00	0.00	16.61
Controls	Yes	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes	Yes
Observations	1,248	1,248	1,248	1,248	773
Observations: did not associate	131	131	131	131	72

Note: Endline and LENA day 1 recording data. See main text Table 4 for details on specifications and outcomes. "Did not associate the intervention with survey" is a dummy equal to 1 if the respondent mentioned neither the video nor the calendar when asked about the baseline survey (without prompting). The dummy is always equal to 0 for control respondents as they did not receive the main intervention and were not asked those questions. N=1,248 instead of1,258 because 10/625 treatment respondents who consented to the endline survey did not reach the intervention recall module at the end of the endline activities, hence are dropped from the sample in this table. See main text Table 3 for further details on recall questions and sample. Robust standard errors in parentheses, p-values in curly brackets.

Table A.12: Robustness of Treatment Effects to Double Lasso Approach

	Mother's Interview				Observed		LENA	
	Mother's belief index	Mother's behavior index	Child language score	Child gestural communication index	Child development index	Child babbles (= 1)	Female adult words per min	Child vocalizations per min
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.187	0.106	0.082	0.066	0.026	0.040	-0.085	0.066
	(0.051)	(0.051)	(0.030)	(0.038)	(0.048)	(0.027)	(0.703)	(0.060)
	{0.000}	{0.040}	{0.007}	{0.083}	{0.593}	{0.135}	{0.904}	{0.273}
Control mean	0.00	0.00	0.00	0.00	0.00	0.51	16.61	1.53
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	No	No	No	No	No	No	No	No
Observations	1,258	1,258	1,258	1,258	1,184	888	774	775

Note: Endline and LENA day 1 recording data. We use the double Lasso approach of Belloni et al. (2013) as implemented by Ahrens et al. (2019) to flexibly choose control variables for each regression. See main text Table 4 for further details on outcomes. Robust standard errors in parenthesis, p-values in curly brackets.

Table A.13: Surveyors: Balance Check and Influence on Outcomes

	Mother's Interview				Observed		LENA		
	Mother's belief index	Mother's behavior index	Child language score	Child gestural communication index	Child development index	Child babbles (= 1)	Female adult words per min	Child vocalizations per min	Treatment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Surveyor 2	0.650	0.519	0.128	0.200	0.025	0.002	-0.771	-0.287	-0.028
	(0.134)	(0.165)	(0.086)	(0.096)	(0.170)	(0.083)	(1.922)	(0.162)	(0.083)
	{0.000}	{0.002}	{0.139}	{0.038}	{0.883}	{0.984}	{0.689}	{0.076}	{0.735}
Surveyor 3	0.781	-1.128	-0.442	-0.406	-0.061	-0.014	0.999	-0.140	-0.031
	(0.147)	(0.103)	(0.072)	(0.091)	(0.114)	(0.064)	(1.760)	(0.145)	(0.067)
	{0.000}	{0.000}	{0.000}	{0.000}	{0.593}	{0.830}	{0.570}	{0.333}	{0.642}
Surveyor 4	0.081	-0.223	-0.228	-0.162	0.059	0.068	1.672	-0.028	0.032
	(0.151)	(0.108)	(0.075)	(0.095)	(0.116)	(0.065)	(1.712)	(0.147)	(0.064)
	{0.594}	{0.038}	{0.003}	{0.090}	{0.611}	{0.297}	{0.329}	{0.849}	{0.623}
Surveyor 5	-0.845	0.575	1.137	0.925	0.103	-0.042	1.545	-0.272	-0.005
	(0.143)	(0.167)	(0.079)	(0.108)	(0.148)	(0.080)	(2.025)	(0.144)	(0.078)
	{0.000}	{0.001}	{0.000}	{0.000}	{0.488}	{0.601}	{0.446}	{0.059}	{0.945}
Surveyor 6	-0.566	-0.215	0.665	0.381	0.243	0.405	1.376	-0.079	0.022
	(0.147)	(0.186)	(0.083)	(0.096)	(0.127)	(0.071)	(1.814)	(0.144)	(0.076)
	{0.000}	{0.249}	{0.000}	{0.000}	{0.056}	{0.000}	{0.449}	{0.586}	{0.772}
Surveyor 7	-1.001	0.293	0.592	0.781	0.291	0.452	1.782	-0.115	0.135
	(0.125)	(0.159)	(0.081)	(0.104)	(0.113)	(0.065)	(1.696)	(0.155)	(0.070)
	{0.000}	{0.065}	{0.000}	{0.000}	{0.010}	{0.000}	{0.294}	{0.457}	{0.055}
Surveyor 8	-0.271	1.078	0.947	0.510	0.701	0.024	-0.449	0.128	0.047
	(0.133)	(0.172)	(0.087)	(0.099)	(0.133)	(0.070)	(1.691)	(0.174)	(0.075)
	{0.041}	{0.000}	{0.000}	{0.000}	{0.000}	{0.734}	{0.791}	{0.461}	{0.532}
Surveyor 9	-0.589	0.421	0.475	0.185	0.189	0.234	0.669	-0.159	0.013
	(0.139)	(0.171)	(0.084)	(0.096)	(0.133)	(0.068)	(1.655)	(0.158)	(0.072)
	{0.000}	{0.014}	{0.000}	{0.054}	{0.157}	{0.001}	{0.686}	{0.313}	{0.852}
Surveyor 10	0.790	-0.819	-0.108	-0.096	-0.287	-0.151	0.510	-0.037	-0.018
	(0.153)	(0.107)	(0.071)	(0.096)	(0.122)	(0.066)	(1.676)	(0.149)	(0.064)
	{0.000}	{0.000}	{0.127}	{0.317}	{0.019}	{0.023}	{0.761}	{0.803}	{0.777}
Surveyors 11-13	-0.090	0.994	0.027	0.246	-0.518	-0.027	-5.469	-0.368	-0.214
	(0.199)	(0.176)	(0.139)	(0.114)	(0.249)	(0.084)	(2.470)	(0.224)	(0.112)
	{0.652}	{0.000}	{0.846}	{0.032}	{0.038}	{0.743}	{0.027}	{0.101}	{0.055}
Surveyor 14							2.310	-0.065	
							(1.891)	(0.160)	
							{0.222}	{0.687}	
F-stat surveyor FE	35.65	29.73	39.31	14.90	6.49	14.02	1.40	1.35	1.48
P-val F-stat surveyor FE	0.000	0.000	0.000	0.000	0.000	0.000	0.167	0.192	0.143
Control mean	0.00	0.00	0.00	0.00	0.00	0.51	16.61	1.53	0.00
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,258	1,258	1,258	1,258	1,184	888	774	775	1,258

Note: Endline and LENA day 1 recording data. For columns 1-6 and 9, regressions include controls for the child's age in days, day of the survey, and surveyor gender (as in Table 4). In columns 7 and 8, regressions include controls for the child's age in days, the day of the week the audio was recorded (dummies), the total time (min) the shirt/LENA device was removed from the child, the total time (min) the child was held on someone's back while wearing the device, and the household size (as in Table 4). All regressions include baseline clinic fixed effects (not shown) and surveyor fixed effects (shown). The dummy for one surveyor ("Surveyor 1") is omitted. Since 3 surveyors completed 20 or fewer endline surveys each, we grouped them ("Surveyors 11-13"). Note that we still use endline surveyors for the LENA outcomes (as opposed to LENA surveyors) since, at the time the LENA measures were collected, the LENA surveyors had minimal contact with respondents (only dropping off and setting up the LENA devices). See main text Table 4 for further details on outcomes. Robust standard errors in parenthesis, p-values in curly brackets.

Table A.14: Treatment Effect on Perceived Barriers to Parent-Infant Conversations

	Agreed with main barrier			Does not
	Beliefs	Hard to form habit	Risk of social scorn	agree there is any barrier
	(1)	(2)	(3)	(4)
Main intervention	-0.019	0.043	0.086	-0.104
	(0.028)	(0.045)	(0.041)	(0.058)
	{0.497}	{0.341}	{0.037}	{0.074}
Endline intervention	-0.050	0.041	0.197	-0.191
	(0.021)	(0.038)	(0.039)	(0.049)
	{0.018}	{0.275}	{<0.001}	{<0.001}
Pure control mean	0.07	0.15	0.10	0.66
P-value Endline=Main Int	0.150	0.967	0.020	0.142
Controls	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes
Observations	780	780	780	780

Note: Endline data. Sample restricted to respondents who received a LENA device hence answered the IDS barrier module after 1 or 2 days of recording (see Figure A.3 for further details on the experimental design and timing of the "Module: IDS Barriers"). Respondents were asked about barriers that may prevent families from talking to their babies. Respondents were specifically asked about three barriers: "it's hard to remember to do it, it takes effort to make it a habit" (habit), "it's not clear that it makes any difference for the child" (belief), and "it's frowned upon /mocked in the community" (social sanctions/scorn). Respondents could also suggest other barriers. Regressions include baseline clinic fixed effects and controls for the child's age (in days), endline survey date, days elapsed between the endline survey and the LENA activity, and LENA randomization strata. Robust standard errors in parenthesis, p-values in curly brackets.

Table A.15: Treatment Effect Split by Child Birth Order

	Mother's Interview				Observed		LENA	
	Mother's belief index	Mother's behavior index	Child language score	Child gestural communication index	Child development index	Child babbles (= 1)	Female adult words per min	Child vocalizations per min
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.143	0.090	0.066	0.047	0.040	0.030	0.272	0.050
	(0.070)	(0.064)	(0.040)	(0.047)	(0.057)	(0.033)	(0.821)	(0.068)
	{0.041}	{0.156}	{0.099}	{0.315}	{0.482}	{0.362}	{0.741}	{0.459}
Treatment x 1st-time mother	-0.051	0.132	0.129	0.186	-0.012	0.092	-0.672	-0.001
	(0.128)	(0.127)	(0.077)	(0.094)	(0.110)	(0.063)	(1.521)	(0.131)
	{0.691}	{0.300}	{0.092}	{0.048}	{0.914}	{0.143}	{0.659}	{0.997}
1st-time mother	0.116	-0.000	-0.079	-0.037	-0.043	-0.023	-0.482	-0.027
	(0.084)	(0.090)	(0.054)	(0.064)	(0.075)	(0.044)	(1.052)	(0.096)
	{0.166}	{0.997}	{0.140}	{0.560}	{0.565}	{0.609}	{0.647}	{0.778}
P-val total effect for 1st-time mother	0.385	0.045	0.003	0.004	0.765	0.021	0.757	0.656
Control mean not 1st-time mother	-0.02	-0.01	-0.06	-0.06	-0.04	0.49	17.02	1.55
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,258	1,258	1,258	1,258	1,184	888	774	775
Observations: 1st-time mother	347	347	347	347	335	240	213	214

Note: Endline and LENA day 1 recording data. See main text Table 4 for details on specifications and outcomes. "1st-time mother" is a dummy equal to 1 if the focal child is the first born of the respondent. Robust standard errors in parenthesis, p-values in curly brackets.

Table A.16: Baseline Characteristics and Balance for Sample with First-Time Mothers

	Full Sample			Treatment		Control		T=C
	Mean	SD	N	Mean	SD	Mean	SD	P-value
Age (years)	23.25	3.37	392	23.25	3.34	23.25	3.39	1.000
Dagomba ethnie	0.81	0.40	393	0.81	0.39	0.80	0.40	0.882
Main langage spoken: Dagbani	0.86	0.35	393	0.88	0.33	0.85	0.36	0.422
Highest level of education:								
None	0.21	0.41	393	0.21	0.41	0.22	0.41	0.832
Primary school	0.32	0.47	393	0.33	0.47	0.31	0.46	0.665
Secondary school	0.33	0.47	393	0.33	0.47	0.33	0.47	0.988
Can read (English/Dagbani)	0.75	0.43	393	0.76	0.43	0.74	0.44	0.660
Housewive/no occupation	0.35	0.48	393	0.30	0.46	0.39	0.49	0.067
Married	0.98	0.12	393	1.00	0.00	0.97	0.17	0.014
Polygamous	0.20	0.40	361	0.23	0.42	0.18	0.38	0.229
Partner is home whole month	0.76	0.43	388	0.75	0.43	0.77	0.42	0.573
Partner passed primary school	0.86	0.35	388	0.86	0.35	0.86	0.34	0.909
Household size	7.56	5.43	390	7.82	5.74	7.33	5.14	0.375
# of household members: under-5	1.44	1.46	393	1.51	1.49	1.38	1.44	0.387
# of household members: 5-15 y/o	1.29	1.81	391	1.43	2.01	1.16	1.60	0.149
# of household members: above-16	4.84	3.14	390	4.90	3.27	4.79	3.03	0.739
Has children	0.66	0.47	393	0.69	0.46	0.63	0.48	0.221
Has child 6 years or younger	0.51	0.50	393	0.53	0.50	0.49	0.50	0.432
Has child older than 1 month	0.50	0.50	393	0.51	0.50	0.48	0.50	0.586
Has child older than 3 months	0.40	0.49	393	0.40	0.49	0.41	0.49	0.841
Age at first child (years)	23.03	3.39	259	22.80	3.43	23.27	3.35	0.266
# of children	0.74	0.71	393	0.81	0.82	0.67	0.60	0.056
Age youngest child (months)	5.47	3.43	259	5.25	3.34	5.69	3.51	0.305
Youngest child eligible	0.66	0.47	393	0.69	0.46	0.63	0.48	0.221
Pregnant with an eligible child	0.34	0.47	393	0.31	0.46	0.37	0.48	0.221
Focal child is first born	1.00	0.00	393	1.00	0.00	1.00	0.00	.
F-test p-value								0.431
Observations	393			184		209		

Note: Baseline data. The sample is restricted to mothers whose child enrolled in the study was their first (alive) child. Treatment is a dummy equal to 1 if the respondent received the intervention at baseline. The question on polygamy was added after the start of the data collection, hence is missing for some observations. Note that some mothers had twins or adopted children from their relatives, hence # of children is on average slightly higher than the % of women who have children. The F-test p-value reported at the bottom of the table is for the joint significance of the differences between the treatment and control groups for all of the variables reported in the table. For the F-test, missing values (due to refusal/don't know or a logic skip (e.g., age of youngest child when no children)) are replaced by the variable average value and flagged by a dummy.

Table A.17: Baseline IDS beliefs and Behavior for First-Time Mothers Sample

	Full Sample			Treatment		Control		T=C
	Mean	SD	N	Mean	SD	Mean	SD	P-value
Beliefs on IDS and Child Development:								
Time/attention is more important than money to a child's success	0.35	0.48	393	0.36	0.48	0.34	0.48	0.686
Child's age (in mo) when:								
a child starts responding with noise/babbles	7.83	10.40	366	8.01	9.67	7.68	11.02	0.762
a child starts saying meaningful words	20.97	14.54	351	21.12	14.40	20.84	14.71	0.858
it becomes clear a child is smart	32.67	21.34	369	32.42	22.54	32.90	20.24	0.830
Child's age (in mo) when parents should start:								
talking to their child	10.83	11.37	375	12.08	12.22	9.74	10.48	0.049
talking in full sentences to their child	25.66	19.11	345	27.01	20.86	24.53	17.50	0.237
telling stories to their child	23.32	18.63	356	23.74	17.72	22.94	19.44	0.684
Self-Reported IDS Behavior:								
Tells stories to youngest child	0.43	0.50	201	0.39	0.49	0.47	0.50	0.264
Ask youngest child to repeat words	0.45	0.50	201	0.45	0.50	0.46	0.50	0.917
When child was 1m/o: Described objects when cleaning/organizing	0.36	0.48	195	0.34	0.48	0.39	0.49	0.509
When child was 3m/o: Described things to child when walking	0.58	0.49	158	0.64	0.48	0.53	0.50	0.147
Inequality Aversion:								
It is best to treat/invest in children equally	0.46	0.50	393	0.45	0.50	0.48	0.50	0.516
A mother should feel bad for 1st child if she provides better care to 2nd child	0.72	0.45	393	0.77	0.42	0.68	0.47	0.054
F-test p-value								0.361
Observations	393			184		209		

Note: Baseline data. Sample restricted to mothers whose child enrolled in the study was a first (alive) child. In the panel "Beliefs on IDS and Child Development", questions "child's age (in months) when parents should start..." were only asked to respondents who reported that the respective activities were important to children's brain development. Child's age outcomes are in months. In the panel "Self-Reported IDS Behavior", questions were only asked to a subset of respondents based on their youngest child's age. "Tell stories to youngest child" and "Asks youngest child to repeat words" were only asked to respondents with a child aged 6 years or less, and the two subsequent questions to parents with a child aged between 1 month and 6 years, and between 3 months and 6 years. The F-test p-value reported at the bottom of the table is for the joint significance of the differences between the treatment and control groups for all of the variables reported in the table. For the F-test, missing values (due to refusal/don't know or a skip pattern such as age of youngest child when there are no children) are replaced by the variable average value and a missing flag is included.

Table A.18: Treatment Effect Split by Use of Calendar to Track Habit

	Mother's Interview				Observed		LENA	
	Mother's belief index	Mother's behavior index	Child language score	Child gestural communication index	Child development index	Child babbles (= 1)	Female adult words per min	Child vocalizations per min
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.102	0.075	0.088	0.053	0.061	0.033	-0.055	0.094
	(0.066)	(0.068)	(0.041)	(0.048)	(0.056)	(0.032)	(0.835)	(0.070)
	{0.123}	{0.267}	{0.032}	{0.270}	{0.275}	{0.314}	{0.947}	{0.180}
Treatment x Colored stars	0.135	0.170	0.044	0.127	0.011	0.056	0.851	-0.097
	(0.100)	(0.086)	(0.057)	(0.072)	(0.075)	(0.042)	(1.066)	(0.095)
	{0.178}	{0.049}	{0.443}	{0.079}	{0.878}	{0.176}	{0.425}	{0.309}
P-val total effect for colored stars	0.012	0.002	0.012	0.007	0.290	0.020	0.418	0.970
Control mean	-0.01	-0.00	0.02	0.02	0.02	0.51	16.62	1.53
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,136	1,136	1,136	1,136	1,076	881	723	724
Observations: colored stars	188	188	188	188	174	172	117	117

Note: Endline and LENA day 1 recording data. See main text Table 4 for details on specifications and outcomes. "Colored stars" is a dummy equal to 1 if the respondent reported keeping track of her IDS-practice by coloring stars on the calendar given at baseline to treated respondents. The dummy is always equal to 0 for control respondents as they did not receive the main intervention and were not given the IDS-themed calendar nor asked those questions. The sample size is smaller than in other tables as we added question on calendar use mid-survey so do not have this data for 90 treatment respondents (see Table 3). See Figure A.1 for further details on the calendar. Robust standard errors in parenthesis, p-values in curly brackets.

References

Ahrens, Achim, Christian B. Hansen, and Mark E Schaffer, "PDSLASSO: Stata module for post-selection and post- regularization OLS or IV estimation and inference," *Statistical Software Components*, 1 2019.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "Inference on treatment effects after selection among high-dimensional controls," *Review of Economic Studies*, 2013, 81 (2), 608–650.

Appendix B: LENA Technology Description

What is a LENA device?

A LENA device is a small recorder children wear for a day in the front pocket of a "LENA shirt" (see Figure B.1). It functions as a sort of "talk pedometer". The audio is processed by a cloud-based LENA software which provides detailed information on the child's audio environment. Available outcomes include lengths of the audio capturing sounds coming from 8 different sources: focal child, other children, male and female adults, overlapping speech, electronic media,non-speech noise, and silence. The LENA software further categorizes those segments as near or far/faint (those with a lower probability of being attributed to the right source of origin) and extracts additional outcomes such as lengths of meaningful speech (including near sounds emitted by the focal child, adults, or other children), faint or overlapping speech (including overlapping speech or far sounds emitted by the focal child, adults, or other children), lengths of focal child vocalizations and non-vocalizations segments, number of focal child vocalizations, adult words, and conversational turn counts. Those are further described in the next section.

Figure B.1: LENA device and shirt

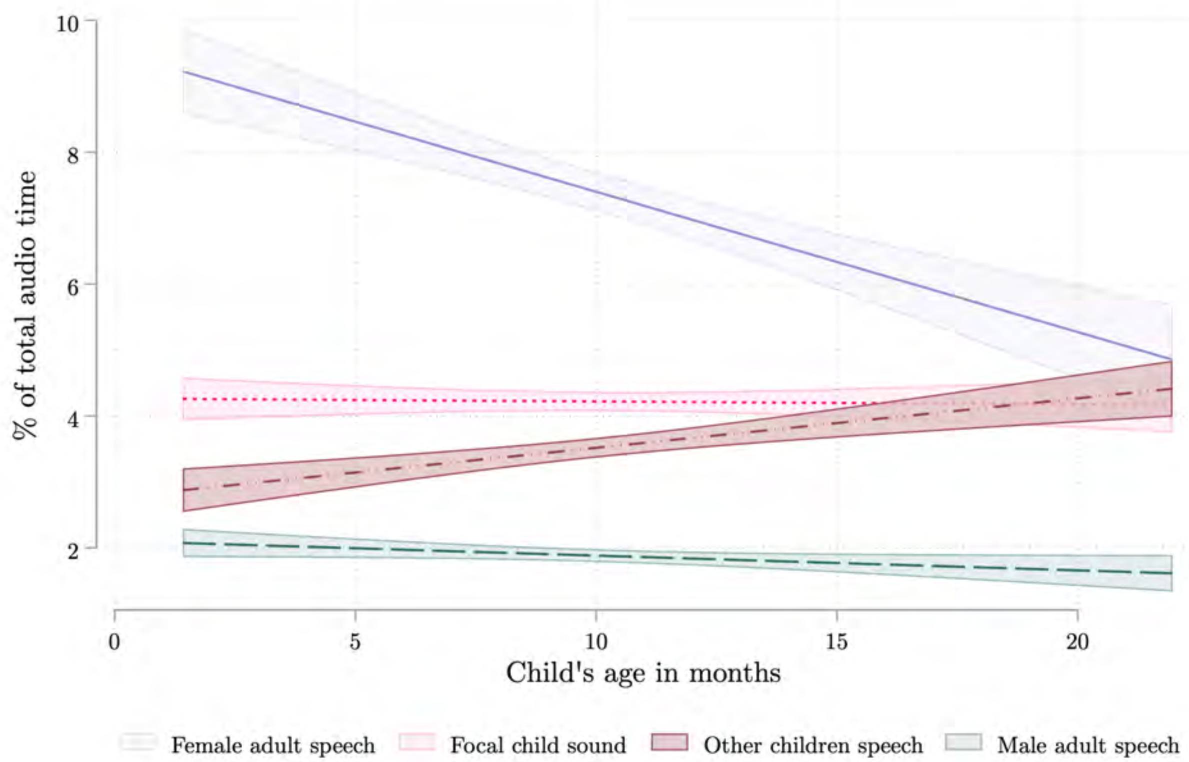
Note: A child participating in the study and wearing the LENA shirt with the device inserted in the front pocket. The LENA device was purchased from the LENA Foundation and the shirt was designed by the research team.

Description of LENA outcomes

In the paper, we focus on the two LENA outcomes summarized below. For further details and more complete descriptions of all LENA outcomes, please refer to the LENA technical reports LTR 12 (available at https://www.lena.org/wp-content/uploads/2020/07/LTR-12_How_LENA_Works.pdf) and LTR-05-2 (available at https://www.lena.org/wp-content/uploads/2016/07/LTR-05-2_Reliability.pdf) (Gilkerson and Richards, 2020; Xu et al., 2009)

- **Adult words count:** estimated number of words spoken by post-pubescent males and females in the child's vicinity.
- **Focal child sounds:** any sound from the child wearing the device. The LENA software further categorizes the focal child sound segments into 1) **vocalizations** (including words, babbles, and pre-speech communicative sounds or "protophones" such as squeals, growls, or raspberries) and 2) **non-speech** sounds like fixed signals and vegetative sounds (such as breathing, burping, crying, etc.).
- **Vocalization count:** estimated number of speech-like utterances the focal child wearing the device emits (including words, babbles, and pre-speech communicative sounds or "protophones" such as squeals, growls, or raspberries). It excludes non-speech sounds like fixed signals or vegetative sounds (such as breathing, burping, crying, etc.).
- **Conversational turn count:** estimated number of back-and-forth alternations between the focal child wearing the device (any alternation including a vocalization) and an adult.
- **% audio of meaningful speech:** includes all segments of the audio labeled as near sounds by adults or focal children as well as near vocalizations from other children.

Figure B.2: Proportion of Speech by Speaker Over Time



Note: LENA data. N=775 recordings. Given financial constraints, only a random subset of households could be included in the LENA measurement (N=900 households sampled). For households sampled to keep the LENA device for two days, only the first day recording is kept in the analysis. The analysis is further restricted to recording times between 10 a.m. to 7 p.m. (this excludes 10/785 LENA day 1 recordings which have less than 9 hours (rounded up) of recording). Outcomes are % of total audio time. Focal child sound (as % of total audio) is the share of the recording tagged by the LENA software as emitted by the child, including both vocalizations and non-vocalizations (cry, fixed signals, vegetative sounds). Lines indicate linear best fit, and shaded areas indicate 95% confidence intervals.

Table B.1: LENA Debrief Survey by Treatment Status (Only 1st Day of LENA Recording)

	Treatment			Control			T=C
	Mean	SD	N	Mean	SD	N	P-value
Shirt/device removed during recording	0.93	0.25	385	0.94	0.24	390	0.625
# times shirt/device removed	1.42	0.76	385	1.44	0.74	390	0.707
Device removed [10h,18h]	0.52	0.50	385	0.50	0.50	390	0.540
# times deviced removed [10h,18h]	0.66	0.72	385	0.64	0.74	390	0.759
Total min device removed [10h,18h]	42.79	78.89	346	39.85	70.85	347	0.606
Device removed during [10h,18h] but invalid duration	0.10	0.30	385	0.11	0.31	390	0.686
Child carried on someone's back with device	0.52	0.50	385	0.51	0.50	390	0.853
# times child carried on back with device	0.80	0.93	385	0.82	0.98	390	0.765
Held on back [10h,18h]	0.47	0.50	385	0.47	0.50	390	0.922
# times held on back [10h,18h]	0.66	0.83	385	0.67	0.83	390	0.875
Total min held on back [10h,18h]	29.18	53.23	361	35.10	69.68	351	0.204
Child held on back during [10h,18h] but invalid duration	0.06	0.24	385	0.10	0.30	390	0.055
Day was unusual for child	0.09	0.29	385	0.09	0.28	390	0.856
Reason why day was unusual for child:							
Child was sick	0.26	0.44	35	0.29	0.46	34	0.736
Child cried throughout day for no reason	0.29	0.46	35	0.24	0.43	34	0.639
Child was uncomfortable with LENA	0.37	0.49	35	0.41	0.50	34	0.736
Child took immunizations	0.06	0.24	35	0.06	0.24	34	0.977
Number of LENA 1 recordings	385			390			775

Note: LENA debrief survey (collected the day after the LENA recording). The sample is restricted to recordings with data from 10 a.m. to 7 p.m. (this excludes 12/785 LENA day 1 audio which have less than 9 hours (rounded up) of recording). Variables in the "LENA Debrief Survey" come from questions asked to primary caregiver the day following the recording.

References

Gilkerson, Jill and Jeffrey A Richards, "A Guide to Understanding the Design and Purpose of the LENA System," *LENA Foundation Technical Report*, 2020

Xu, Dongxin, Umit Yapanel, and Sharmi Gray, "Reliability of the LENA Language Environment Analysis System in Young Children's Natural Home Environment" *LENA Foundation Technical Report*, 2009

Appendix C: Descriptive analyses of outcome variables

This appendix highlights cross-sectional patterns in our data to offer insight into our outcome measures. We focus on data from the control group, i.e., a group not informed about the benefits of IDS. In Table C.1, we estimate multivariate regressions to understand whether socioeconomic status predicts our outcome variables. In Figures C.1 to C.8, we present scatter plots and bivariate correlations between the LENA-recorded measures, the mother-reported measures, the surveyor-observed measures, and child age. We discuss cross-sectional patterns in the data revealed by these analyses.

C.1 Associations with socioeconomic status

Table C.1 shows that maternal education and indicators of wealth (having a home with concrete walls rather than mud walls) are *negatively* associated with both female adult words per minute and child vocalizations per minute as recorded by the LENA. This is the exact opposite of what has been observed in higher-income countries. A key question is whether this is indicative that the LENA provides poor measures of the true outcomes in our context, or whether the gradient is truly different across higher-income and lower-income contexts.

The lack of a positive SES gradient appears common to all measures of child development (whether LENA, reported by the mother or observed by the surveyor), which strongly suggests that the underlying relationship is truly different in lower-income contexts. (It is only for mother-reported measures of parental behavior that we do observe a positive SES gradient, but this may be due to social desirability bias.) Overall, this evidence suggests that even the more educated and/or wealthier parents may under-invest in the early cognitive development of their children.

Table C.1: Correlation Between Outcome Variables and Socioeconomic Status

	LENA		Observed		Mother's Interview	
	Female adult words per min	Child vocalizations per min	Child development index	Child babbles (= 1)	Child language score	Mother's behavior index
Panel A: Maternal education						
Mother's years of education	-0.220	-0.015	0.007	0.007	0.000	0.022
	(0.089)	(0.007)	(0.006)	(0.003)	(0.004)	(0.007)
	{0.014}	{0.038}	{0.248}	{0.042}	{0.909}	{0.002}
Control mean	16.61	1.53	0.00	0.51	-0.06	0.00
Observations	388	389	598	435	628	628
Panel B: Indicator of wealth						
Concrete walls	-1.319	-0.095	0.057	0.057	-0.014	0.068
	(1.030)	(0.087)	(0.069)	(0.041)	(0.048)	(0.078)
	{0.201}	{0.277}	{0.403}	{0.166}	{0.766}	{0.384}
Control mean	16.61	1.53	0.00	0.51	-0.06	0.00
Observations	389	390	603	439	633	633
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	No	No	No	No	No	No

Note: Endline and LENA day 1 recording data. Sample is restricted to control group respondents. Panel A and Panel B report results from separate regressions. Regressions include controls for the child age group (under 5 months old, 6-9 months old, 10-14 months old, over 15 months old dummies). See Table 4 notes for further details about the sample and Appendix A.2 for details on the LENA device and outcomes. "Mother's years of education" is the number of years of education completed by the child's mother. "Concrete walls" is equal to 1 if the child's home has concrete walls and 0 otherwise. Robust standard errors in parenthesis, p-values in curly brackets.

C.2 Reliability of measures

Figure C.1 shows a weak relationship between child verbal output and child age. This also contrasts with findings from higher-income countries (Gilkerson and Richards, 2008). We also observe falling mother verbal input with child age (Figure C.1). This latter result may explain the former: lower mother verbal input means the child has less reason to vocalize even if their language skills have improved. As such, this would indicate that the LENA measure of child vocalizations is not a good proxy for child language development. Consistent with this, we find that child vocalizations per minute has weak correlations with the surveyor-observed child development index or child babbling (Figure C.3) as well as the mother-reported child language score or child gestural communication index (Figure C.4). Figure C.2 also shows a weak correlation between female adult word count and mother-reported behavior.

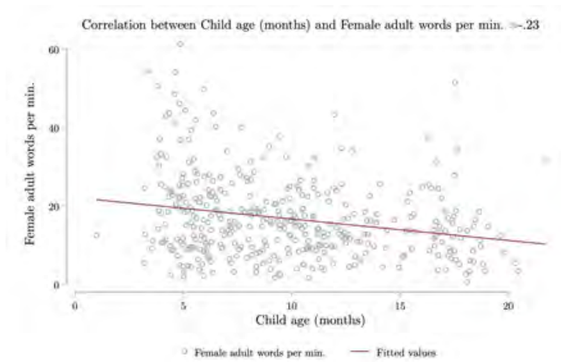
The correlation between LENA-recorded child verbal outputs and mother-reported/surveyor-observed measures is stronger for children 15 months or older. In Figure C.5, we estimate bivariate correlations between our main outcomes for four age groups: 5 months or younger, 6-9 months old, 10-14 months old, and 15 months or older. Children 15 months or older exhibit the strongest relationships between child vocalizations per minute and mother-reported measures (e.g., the child language score) as well as surveyor-observed measures (e.g., the child development index).

As expected, there are strong relationships between surveyor-observed/mother-reported measures of child development and child age. The bivariate correlations for child age and the child development index, child babbling, the child language score, or the child gestural communication index range from 0.54 to 0.77 (Figure C.6 and Figure C.7).

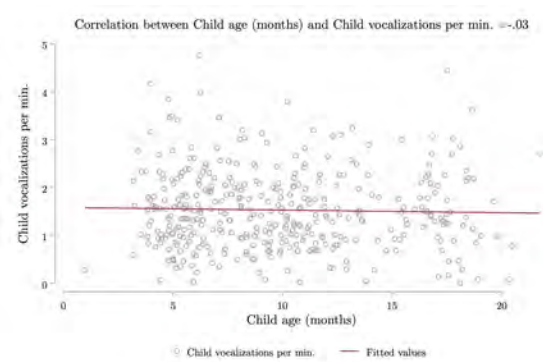
The surveyor-observed and mother-reported child development measures are also predictive of each other. Figure C.8 estimates correlations of 0.44 to 0.52 between the surveyor-observed measures (child development index/child babbling) and the mother-reported measures (child language score/the child gestural communication index). These correlations are present across age groups, though they are weaker for children under 5 months old.

Figure C.1: Correlation Between LENA-recorded Outcomes and Child Age

(a) LENA-recorded Parental Verbal Inputs



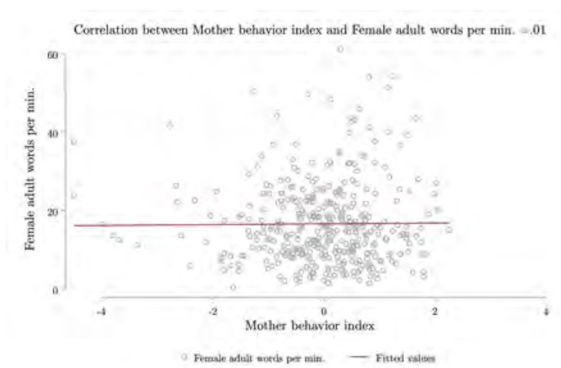
(b) LENA-recorded Child Verbal Outputs



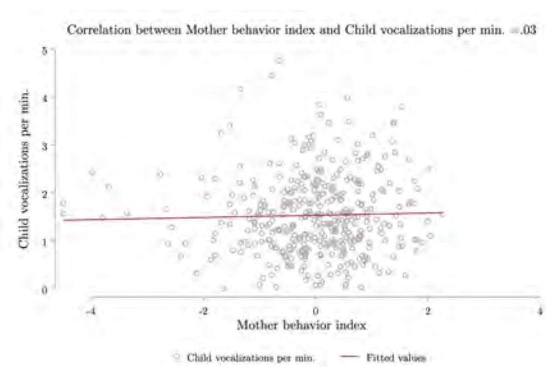
Notes: Endline and LENA data. Using data from the control group only. For details on outcomes and samples, please refer to Table 4. All correlations presented are bivariate correlations.

Figure C.2: Correlation Between LENA-recorded Outcomes and Mother-reported Behavior Index

(a) LENA-recorded Parental Verbal Inputs



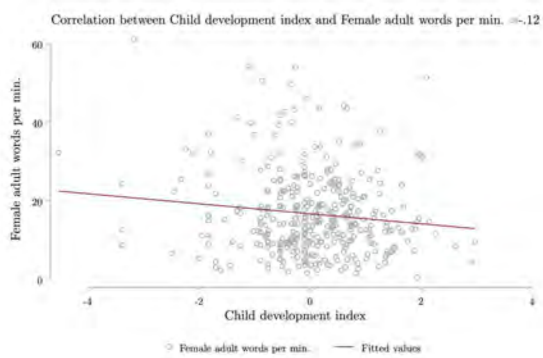
(b) LENA-recorded Child Verbal Outputs



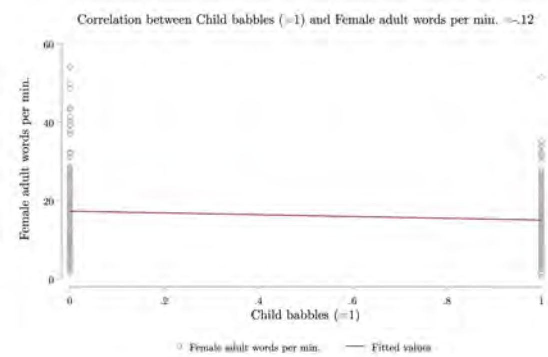
Notes: Endline and LENA data. Using data from the control group only. For details on outcomes and samples, please refer to Table 4. All correlations presented are bivariate correlations.

Figure C.3: Correlation between LENA-recorded Outcomes and Surveyor-observed Child Development

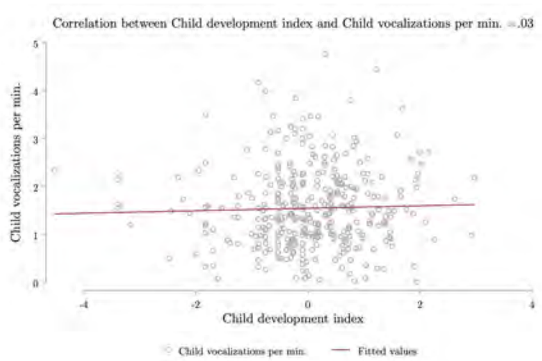
(a) LENA-recorded Parental Verbal Inputs and Surveyor Observed Child Development



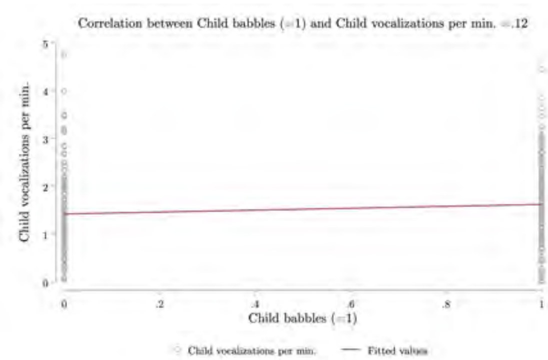
(b) LENA-recorded Parental Verbal Inputs and Surveyor-observed Child Babbling



(c) LENA-recorded Child Verbal Outputs and Surveyor-observed Child Development



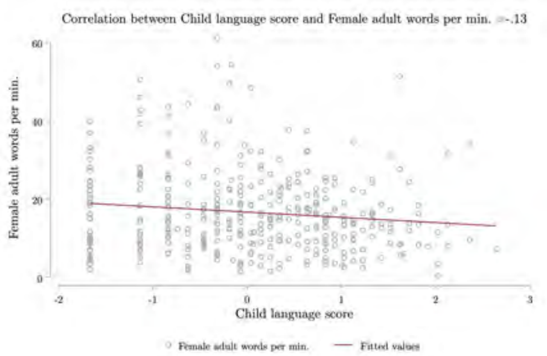
(d) LENA-recorded Child Verbal Outputs and Surveyor-observed Child Babbling



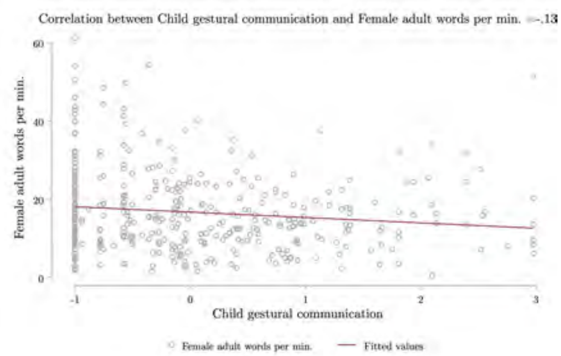
Notes: Endline and LENA data. Using data from the control group only. For details on outcomes and samples, please refer to Table 4. All correlations presented are bivariate correlations.

Figure C.4: Correlation Between LENA-recorded Outcomes and Mother-reported Child Communication

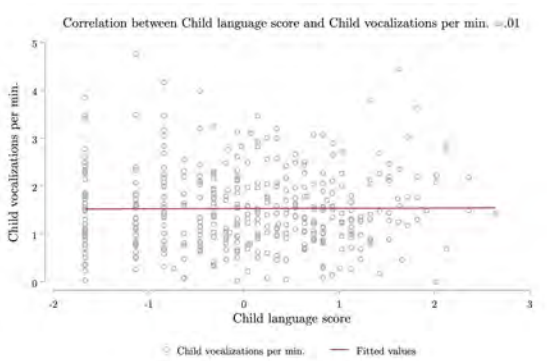
(a) LENA-recorded Parental Verbal Inputs and Mother-reported Child Language



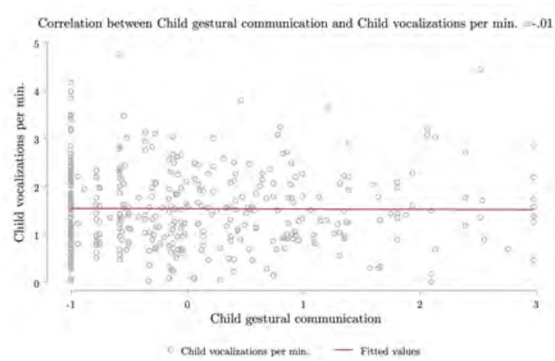
(b) LENA-recorded Parental Verbal Inputs and Mother-reported Child Gestural Communication



(c) LENA-recorded Child Verbal Outputs and Mother-reported Child Language



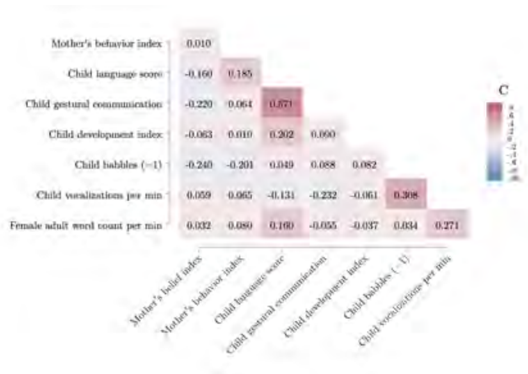
(d) LENA-recorded Child Verbal Outputs and Mother-reported Child Gestural Communication



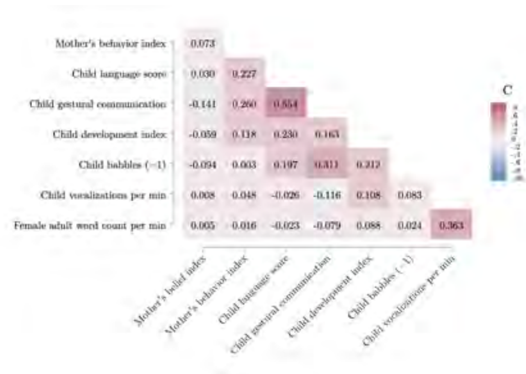
Notes: Endline and LENA data. Using data from the control group only. For details on outcomes and samples, please refer to Table 4. All correlations presented are bivariate correlations.

Figure C.5: Correlations of Outcome Variables by Age Group

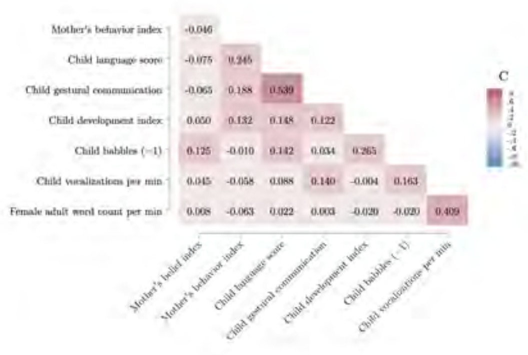
(a) 5 months or younger



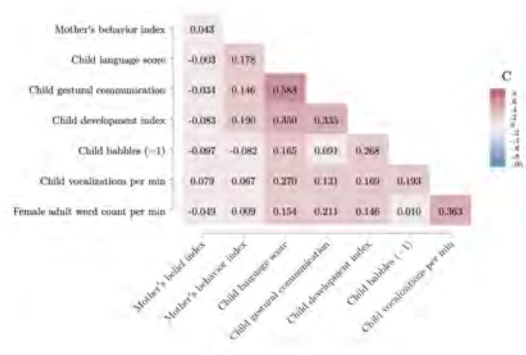
(b) 6-9 months old



(c) 10-14 months old



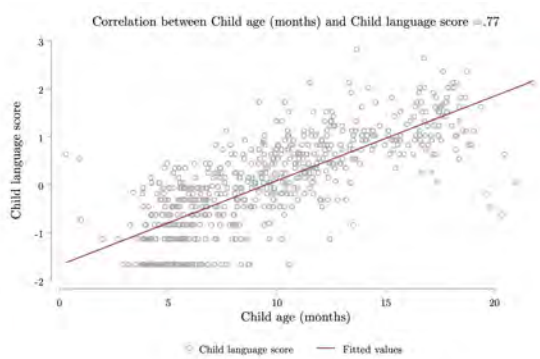
(d) 15 months or older



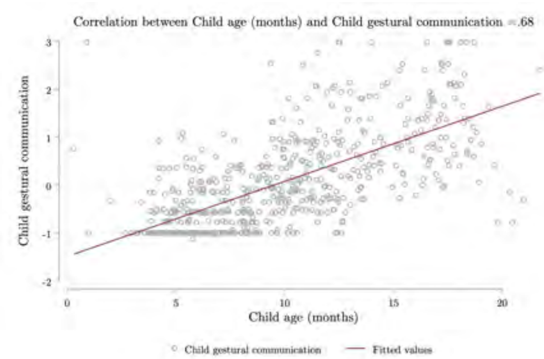
Notes: Endline and LENA data. Using data from the control group only. For details on outcomes and samples, please refer to Table 4. All correlations presented are bivariate correlations.

Figure C.6: Correlation Between Mother-reported Child Communication and Child Age

(a) Mother-reported Child Language



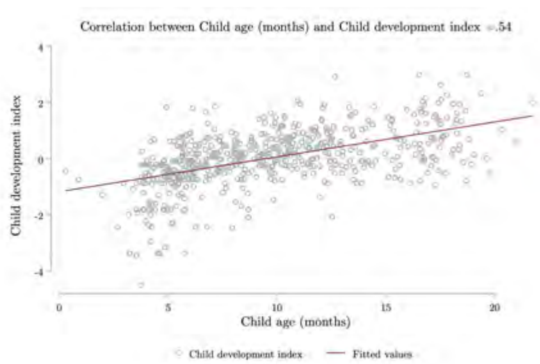
(b) Mother-reported Gestural Communication



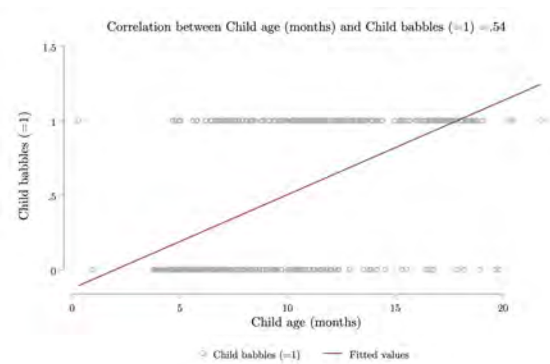
Notes: Endline. Using data from the control group only. For details on outcomes and samples, please refer to Table 4. All correlations presented are bivariate correlations.

Figure C.7: Correlation Between Surveyor-Observed Child Development and Child Age

(a) Surveyor-observed Child Development Index



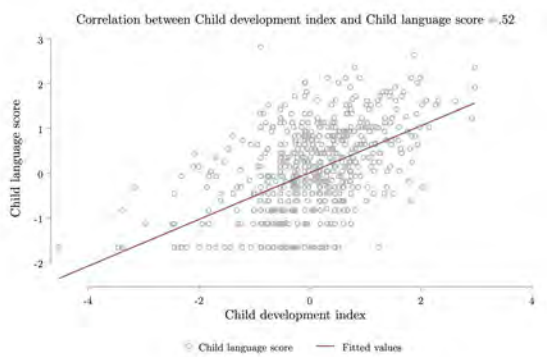
(b) Surveyor-observed Child Babbling



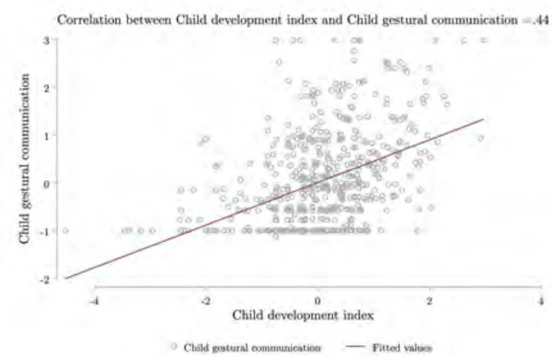
Notes: Endline. Using data from the control group only. For details on outcomes and samples, please refer to Table 4. All correlations presented are bivariate correlations.

Figure C.8: Correlation Between Surveyor-observed and Mother-reported Child Development Outcomes

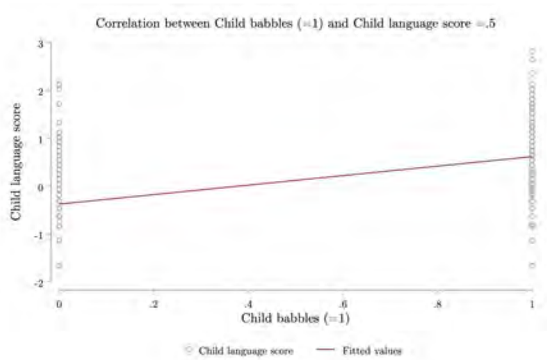
(a) Surveyor-observed Child Development Index and Mother-reported Child Language



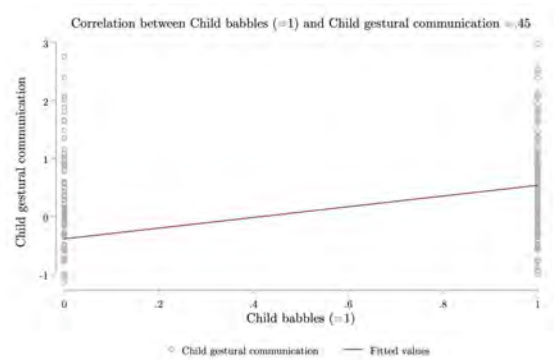
(b) Surveyor-observed Child Development and Mother-reported Child Gestural Communication



(c) Surveyor-observed Child Babbling and Mother-reported Child Language



(d) Surveyor-observed Child Babbling and Mother-reported Child Gestural Communication



Notes: Endline and LENA data. Using data from the control group only. For details on outcomes and samples, please refer to Table 4. All correlations presented are bivariate correlations.

References

Gilkerson, Jill and Jeffrey A Richards, "The LENA natural language study" *LENA Foundation Technical Report*, 2008