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

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

May 8, 2023

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. We report on a randomized field experiment evaluating a low-cost intervention that aims to raise verbal engagement with infants by showing recent or expectant mothers a 3-minute informational video and giving them a themed wall calendar. Six to eight months later, mothers selected for the intervention report greater belief in the benefits of verbally engaging with infants, more frequent parent-infant conversations, and that their infants have more advanced language and cognitive skills. We measure positive but noisy effects on parental verbal inputs in a day-long recording and on surveyor-observed infant cognitive skills. The intervention could be delivered to expectant mothers through existing health clinics at very low marginal cost so could be a highly cost-effective early childhood development policy in low-income contexts.

JEL: C93, D19, I21

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

*The authors are from Stanford University, NBER and CEPR (pdupas@stanford.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. This RCT was registered as AEARCTR-0007161 (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. These problems may be especially acute in developing 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 believe less in 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 mothers should start talking to their baby at birth,[1] and 61% reported that parents should begin talking when a child is over six months old. While one might have hoped that rising educational attainment in low-and-middle-income countries (LMICs) would narrow this gap, Duflo et al. (2022) 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 not surprising, 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-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 full intervention (or video-calendar 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 three-minute 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 instructed 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 calendar handed out to women visiting local government health clinics for an prenatal or postnatal checkup.

To evaluate this intervention, we randomly selected 705 Northern Ghanaian women to receive the intervention 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 infant cognitive development. 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 cognitive skills. The magnitude of the effects are remarkably aligned, at around 0.1 standard deviation.

To address concerns about experimenter demand effects, we use a number of strategies. We gathered observed measures of infant cognitive development and parental verbal inputs. For infant cognitive development, we adapted the Oxford Neurodevelopment Assessment (Ox-

NDA, a test for 10-14 month olds) to our sample of 2-18 month olds (Fernandes, 2021).[2] In addition, surveyors kept track of whether the infant was observed babbling at some point during the survey. To measure parental verbal inputs, we used LENA-recording devices to gather day-long recordings of the child's auditory environment. We used the LENA measures in only half of our sample due to budget constraints. We find an increase of 0.06 standard deviation in babbling and noisy but positive effects on the other observed outcomes. In addition, we home in on respondents who, at endline, did not associate the intervention with the survey organization; these respondents should be subject to less social desirability bias when talking with the endline surveyor. Among those who received the intervention, reports on mother behavior, child vocabulary, and child development differ little between those who associate the intervention with the survey organization and those who do not. Taken together, these results indicate that experimenter demand effects are unlikely to be driving our results.

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 fully embracing the recommended practices versus adopting them initially but not persisting, we also administered our informational video treatment to a randomly selected subset of control-group respondents *the morning after the endline measurements* (we refer to this as the 'endline intervention' to distinguish it from the full intervention administered just after the baseline). We use the LENA recording devices to measure their behavior *the following day they received the information.* The newly-informed mothers increase the words spoken to their child by 1.4 words per minute (≈9% of the mean, p=.052) in the day-long recording, which is eight times larger than the treatment effect observed after 6-8 months for the baseline intervention group (+0.17 words per minute, p=.805). The large immediate impact of the 'endline intervention' on parental behavior shows that mothers are willing

[2]We would have preferred to use a tool designed for 2-18 months, but we could not identify an infant cognitive development tool that was not based parental reports and could be implemented by survey staff at the homes of our Northern Ghanaian respondents.

and able to verbally engage with their children when (1) they are told that they should, and (2) they know their mother-child 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 baseline 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 full 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 initial information delivery, however. Mothers who received the full intervention at baseline (i.e., 6-8 months prior to being asked about barriers) are 30% less likely to report social scorn as the main barrier (p=.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 notoriously 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 infant cognitive 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 we identify 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] List et al. (2021) and other studies are also potentially more likely to engender social desirability bias compared to our light-touch approach which consisted of one short interaction that occurred 6-8 months prior to gathering outcome data.

There is less evidence on parenting interventions in 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. Many 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. Verguet et al. (2022) measures the cost-effectiveness of 12 early childhood interventions and finds that the median intervention costs $212 per child's standard deviation improvement in cognitive or language development. We estimate that our intervention delivers a 1 standard deviation improvement in a child's cognitive or language development for $4-$8 spent at scale ($29 to $49 under our research trial conditions), placing our intervention between the 85th-100th cost-effectiveness percentile of those evaluated by Verguet et al. (2022).[4] High-touch interventions are also more susceptible to the critique (made by Keller (2018) and Wang et al. (2021), for example) that outside meddling with traditional child-rearing practices in LMICs is potentially harmful because it shifts families' focus away from locally adapted skills. In short, our intervention is feasible to scale for LMICs and its success, despite being light-touch, provides evidence of LMIC families' latent demand for information on how to enhance their children's cognitive skills.

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

[4] When using average LMIC labor costs, our preferred estimate ($4.56) is lower than the most cost-effective study in Verguet et al. (2022), which cost $18 per standard deviation improvement.

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. While the Northern region is significantly poorer than the rest of Ghana,[5] Tamale has experienced significant growth in population and business activity over the past two decades.

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. IPA 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 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-

[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.

forth moments" are particularly important for infant cognitive 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 instructed 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]

2.2 Sample characteristics and baseline behavior

Table A.2 presents baseline characteristics for our sample. Reassuringly, only 1 of the 13 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

[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. One explanation for the low usage of the web page is that our sample was reluctant to use valuable internet data streaming a video.

[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 critical mechanism for the treatment effects. Unfortunately, our experiment cannot speak to this.

treatment and control (p=.927).[10]

In our sample, nearly all women are married, with almost a third in polygamous unions. Nearly two thirds of respondents have at least a primary school education, and 61% can read in English or in 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 1 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 infant cognitive development such as extensively conversing with newborn infants.

2.3 Endline survey

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. We completed interviews with 89% of respondents with no differential attrition between the treatment and the control group (refer to Table A.3). The most common reason for attrition was that the respondent moved to a family member's home outside the Tamale area. The endline survey measured parental beliefs about verbal engagement with infants, parental verbal inputs to the child, child language development, child cognitive development, recall of the treatment, and perceived barriers to IDS. All outcomes were collected in the main endline survey except for those on

[10] We also cannot reject the null of the joint test for the baseline variables presented in Table A.2 providing additional evidence that the randomization was implemented correctly.

treatment recall and perceived barriers to IDS which were measured at the very end of all endline-related activities (see more below).

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 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 presents statements about infant cognitive and language development to the respondent and asked them to rate their level agreement with the statement on a Likert scale from 1 (strongly disagree) to 4 (strongly agree).

For parental verbal inputs, we used questions developed by the Harvard Laboratory for Developmental Studies designed specifically for Ghana (Duflo et al., 2022). This measure consists of yes/no questions about whether the respondent and/or another adult engaged in a given activity with the target infant. 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 target 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, male adults, and other children to the child, child vocalizations, and conversational turns between the child and others. 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-activity 16 days after the main endline survey.

The LENA surveyors visited respondents before 10 a.m. on the day of the recording activity

[11] This device was validated in Ghana by the Harvard Laboratory for Developmental Studies. Refer to Appendix, Section A.1 for more information.

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 mothers to have the child wear the shirt until the next day. 97% of mothers consented after the surveyor described the LENA process. We restrict the LENA analysis to audio-data collected from 10am 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 baseline 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.[13] We obtained 785 LENA recordings (see Table A.3 for details on survey and LENA-activities participation rates).

To assess child language development, we relied on items from a version of the MacArthur-Bates Communicative Development Inventories (MB-CDI) adapted to Ghana by the Harvard Laboratory for Developmental Studies (Duflo et al., 2022). In this inventory, mothers answer a series of questions pertaining to how their child communicates through gestures and words. We measure a child's latent language ability using item response theory (IRT) which involves estimating a one-parameter logistic model on the mother's responses to the adapted MB-CDI, where the model assigns a difficulty level to each trial and, then, a latent trait to each individual based on their answers to the trials adjusting for the trial's difficulty level.

We measure infant cognitive development through items from the Ages and Stages Questionnaire (ASQ) and the Oxford Neurodevelopment Assessment (Ox-NDA).[14] The ASQ items

[12] This time window ensured we had comparable data for all observations as households received the LENA device between 6-10 a.m. depending 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).

[14] We considered using other child cognitive tests/assessments such as the Bayley Scales of Infant and

ask whether the target child reached certain developmental milestones (such as making cooing sounds). Most Ox-NDA items 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).[15] We include these Ox-NDA items to have a measure of infant cognitive development that is not be subject to social desirability bias. While the Ox-NDA is designed for children 10-14 months old, 49% of our target children were under 9 months or younger at endline. One of the developers of the Ox-NDA, Dr. Michelle Fernandes, generously worked with us to identify items that might work for our age group. In July 2022, we piloted these items, after adapting them to the local context, and identified the most promising ones (in terms of expected variation) to include in our measure. In the analysis of the Ox-NDA, we restrict our sample to children aged 3 months and older since the summary development index score is not positively correlated with age prior to 3 months (see Figure A.3). When prompted to do some tasks for the Ox-NDA test, some children disengaged from the test or refused to perform tasks they had performed at other instances in the survey (such as babbling, responding to no, etc.). As a result, we added an observed measure for whether the surveyor observed the child babbling or mimicking at least 1-syllable babble at some point during the home visit.

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 grow over time, as participants gain experience and comfort with conversing with their infant. Thus, to assess how the immediate effect of the intervention differs from the 6-8 month effect, we randomly selected 225 control respondents

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.

[15]The Ox-NDA items also include mother-reported questions that we group with the ASQ items in our analysis.

(among the 449 control respondents sampled for the LENA-activities) to be administered a video-only treatment *after* we had collected their endline measurements and a first set of LENA measurements.[16] The day after the LENA device was dropped off for a day-long recording, the LENA surveyor returned to respondents' homes to pick up the device. The sub-sample of control respondents selected for the 'endline intervention' completed the debrief survey but were not asked questions on barriers to conversing with babies. Instead, they watched the intervention video on the surveyor's tablet and were asked to use the LENA for one more day. 99% of selected respondents agreed to a second day of recording. 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.[17] For the 'endline intervention' subsample (all of whom watched the video), we use a before-after analysis to estimate the intervention's effects, comparing the LENA data collected one day before and one day after the video was shown to them.[18]

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

[16] We stratified households by whether the target child would be under or over 1 month old at baseline, and for household with a target child over 1 month old at baseline, whether they scored above or below median on a baseline self-reported IDS behavior score. 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.3 for further details).

[17] 112 of the 225 'endline intervention' respondents 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.

[18] 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.

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 fadeout.[19] 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.

2.5 Treatment recall, social norms, and perceived barriers

To avoid inducing social desirability bias 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 answered questions on these topics at the end of the endline survey, while the LENA subsample answered the questions on barriers to IDS only after the LENA measurement had been collected, in a short survey administered during a surveyor's visit to pickup the LENA device. For respondents sampled for two LENA recordings (the endline intervention sample), the questions were asked after their final (second) LENA recording. Figure A.4 summarizes the study timeline and timing of the different endline questions.

At endline, we asked treatment respondents about their participation in the baseline survey to understand their susceptibility to social desirability bias 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 2). 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%

[19] Conversely, a disadvantage of our approach is that the environment at the time the intervention is delivered is not held constant. Our approach also implies a smaller sample size to estimate the short run effects (because the analysis is conducted within the original control group); however, when one expects fadeout and the study is powered to detect long-run effects, a smaller sample size will often suffice to detect short-run effects.

remember the message about talking to your child. 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 instructed 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.4). 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 gathered in this experiment into summary indices corresponding to our outcomes of interest: mother's beliefs, parental verbal inputs, infant language, and infant cognitive development. We use different indices for mother-reported outcomes and observed outcomes in the domains of parental verbal inputs and child cognitive development to explore concerns around social desirability bias. Tables A.6 to A.10 present the outcomes included in each index. We follow Anderson (2008)'s proposed method for implementing variance-weighted summary indices. We do not impute missing index components when calculating the components' weights and the indices.[20]

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). With this limited set of measures, we do not need to create a summary index and instead directly estimate the effect on each outcome.

[20]Except for the observed infant cognitive development index, missing index components are due to "Refuse to answer" and "Don't know". Between 95-100% of respondents answered all components. For the observed infant cognitive 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% components (N=17).

3 Empirical Framework

3.1 Treatment effects

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 least squares (OLS):

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \eta_i + \epsilon_i \quad (1)$$

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, the household size, the day of the week the audio was recorded, and interruptions to the LENA's recordings.[21] η_i represents clinic fixed effects. ϵ_i is the error term, and the estimated standard errors are robust to heteroskedasticity.[22]

The experimental variation in T_i generated by the RCT enables us to causally identify the 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 raising 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 find that 7% of treatment respondents knew someone else who had seen the intervention video (enrollment in the study was on a rolling basis over 3 weeks). At endline, 8% of 195 control respondents who received the 'endline intervention' told us that they knew someone who had seen the video, and 16% of 615 treatment respondents (who

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

[22] We assess the robustness of our results to alternative specifications in Table A.5.

had received the 'baseline intervention') had discussed the video with friends (see Table 2); a subset of these friends could be in the control group.

3.2 Within-subject effects on newly-informed mothers

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 household 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.

Our coefficient of interest is γ_1. This coefficient captures the effect of receiving the intervention on child's verbal environment one day later, but may also capture the effect of the household using the LENA device for the second day (compared to the pre-treatment measure which was the first day the family had used the LENA).

4 Results and Discussion

4.1 Treatment effects

In Table 3, we present our main results from Equation 1.[23] Columns 1-4 show the outcomes of interest reported during the mother's interview. The intervention increases mother's beliefs in the efficacy of conversing with infants by .126 standard deviations (SD), based on the belief index (p =.030). The remarkably similar estimate of .124 SD (p =.025) in the reported parental verbal input index suggests that mothers altered their child-rearing practices as well.

[23] Tables A.6 to A.10 present the results from Equation 1 when applied to the individual outcomes that make up each index.

Treatment mothers also report significantly higher child language and cognitive development (.105 SD and .097 SD increases; language p =.003; cognitive p =.014). Taken at face value, these estimates indicate significant cognitive gains from a very low-cost intervention.

In our surveyor-observed and LENA-based measures, the treatment effects are also positive but noisily estimated. For the surveyor-observed child development index, we estimate a .057 SD increase ($p = .238$) (see Figure A.5 for mean scores by child's age for the treatment and control groups). The intervention seemed more effective when delivered before or very shortly after the child's birth. The intervention also led to a .059 SD increase ($p = .040$) in the surveyor-observed measure of the infant babbling at some point during the home visit.

The estimated treatment effect on the LENA-recorded parental inputs index is a .068 SD increase ($p = .335$). The divergence between the observed and mother-reported results could be the result of social desirability biasing the mother-report estimate upwards or noisy estimates obscuring the effects on observed outcomes. Recall that for the LENA results, the sample size is half as large as for the self-reported outcomes.[24]

4.1.1 Reported versus observed outcomes

In this section, we evaluate two potential explanations for the divergence of the observed and reported results in Table 3: social desirability bias and noisy measurement of observed outcomes. If social desirability bias drives our results, the estimates in Columns 1-4 of Table 3 would be biased upwards and our best estimate of the effect of the intervention on parental verbal inputs and infant cognition would be Columns 5-9 in Table 3. These estimates would only provide weak evidence of a positive impact of the intervention.[25] In contrast, if noisy measurement of the observed outcomes drives the differences, then our best estimates would be the more precise effects shown in Columns 1-4 of Table 3. These results would provide strong evidence that scaling up this intervention would be a highly cost-effective method of

[24]The distributions of key LENA measures by treatment group and Kolmogorov-Smirnov test of equality of the distributions are shown in Figures A.6 and A.7.

[25]Given the low cost of the program, this evidence might still justify scaling up the intervention.

improving early childhood development.

Social desirability bias could arise if the treatment group associates IPA surveyors with the video-calendar intervention and thus, feels pressure to report believing in, practicing, and seeing positive results from conversing with infants when they speak to another IPA surveyor 6-8 months later. To explore this hypothesis, we exploit the fact that 131 treatment respondents (21% of treatment respondents asked at the end of the endline activities) did not associate the intervention video or intervention calendar with IPA's baseline interview (see Table 2). This group is unlikely to be subject to social desirability bias, or experimenter demand effects, when interviewed by IPA at endline.

In Table A.11, we separately estimate the difference in outcomes between the control group and the treatment respondents who associated the intervention with the survey organization and those that did not. If social desirability bias drives our effects, we would expect to see larger effects for those who associate the survey organization with the intervention.[26] For the belief index, we find results consistent with social desirability bias. Those who associate the intervention with the survey organization have a .181 SD (p = .003) higher beliefs than the control respondents, while the rest of the treatment group has .079 SD (p = .509) lower beliefs. However, this pattern is not mirrored by the other mother-reported outcomes. The estimates for the parental verbal input index (.143 SD (p = .015) for those who associate the intervention with the survey organization to .110 SD (p = .278) for the others) and the child language score (.115 SD (p = .003) compared to .085 SD (p = .097)) are nearly equivalent, and the point estimate for the mother-reported child development index is *higher* for treatment respondents who did not associate the intervention with the survey organization (.098 SD (p = .020) compared to .109 SD (p = .087)). These patterns suggest that social desirability bias may drive the large estimated treatment effect on the belief index but is unlikely to drive the effects on the behavior index, child language score, or child development index.

[26]Since we are conditioning on an endogenous variable, differences in baseline characteristics may also be driving differences in endline outcomes. However, we would expect that this would bias us towards finding evidence of social desirability bias.

We have a priori reasons to worry about the noisiness of the observed outcomes. The Ox-NDA measures had to be adapted to our population (0-22 months old) from a test originally designed for 10-14 months olds. In addition, we only used 9 items rather than the suggested 37 due to time and budget constraints. For the LENA recordings, we could only survey half of our sample diminishing our ability to precisely estimate effects. Table 3 shows that the LENA index and observed child development index have higher standard errors than their mother-reported counterparts. We believe this noisiness in the observed measures is the likely reason why we failed to identify significant effects on the observed outcomes while we observe significant and substantial effects on the mother-reported outcomes.

4.2 Effects on newly-informed mothers

In Table 4, we measure the impacts of the video on newly-informed mothers using Equation 2. Compared to the recordings of the child's environment the day before (Day 1), the child hears more adult words (1.827 SD; $p = .036$). 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 77% of the effect). However, there is an impact on male adult words (.517 SD; $p = .071$) 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. We see a modest but insignificant increase in "conversation turns" per minute (.015 SD; $p = .352$), which require engagement between others' verbal inputs and the target 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 engender other household members to do so as well. The difference between these substantial effects and the positive but noisy effects of the 'baseline intervention' on the outcomes measured by the day-long recording suggests that there are barriers to sustaining this level of behavioral change.

4.3 Robustness checks

In Table A.5, we test the robustness of our results to the inclusion of different covariates. We present the results with only clinic fixed effects in Panel A and the results with clinic-day fixed effects[27] in Panel B. As expected when varying covariate inclusion in a randomized experiment, the coefficients change only slightly between Panels A and B and our main specification (Table 3). The standard errors increase in Panel A (because excluding other control variables reduces precision), while barely changing in Panel B.

In Panel C, we include surveyor fixed effects as well as the controls in our main specification. Comparing Panel C to Table 3, we see moderate changes in the coefficients, suggesting that surveyor idiosyncrasies may have influenced the measurement of our outcome variables. The significant effect on observed child babbling disappears while remaining positive (.038 SD; p = .170). However, the other results emphasized in the previous section are robust to surveyor fixed effects as there are still significant effects on reported beliefs, parental verbal inputs, and child development as well as noisy but positive effects on observed parental verbal inputs and child development.[28]

In Table A.12, we test the robustness of the results on the observed infant development index by imputing missing components with the lowest possible score. The coefficient remains positive and imprecise with similar standard errors (.045 SD; $p = .362$). In Table A.13, 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 equation.[29] Once again, we see little change in the coefficient values. We see gains in power (i.e. smaller standard errors) but the differences are quite small, except for the effect on observed infant babbling which becomes insignificant ($p = .103$).

[27] In addition to the controls in the main specification.

[28] The point estimates for observed child development are lower in Panel C indicating that surveyor judgment may have played a particularly important role in this measure. This is consistent with our hypothesis that noisiness in the measurement of observed child development caused the divergence in results between reported and observed outcomes.

[29] Roughly, the double Lasso approach selects controls from a large set of potential controls by identifying which ones best predict the outcome and treatment variables. This approach enables searching over many controls but does not necessitate corrections for multiple hypothesis testing.

4.4 Mechanisms

When we asked respondents what the main barrier that prevented families from talking to their babies was, 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%) (Table A.4). 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 pps ($p < .001$; Table A.14). However, this effect does not seem to persist. Mothers who received the treatment 6-8 months ago (the 'baseline intervention' group) are much less likely to report social scorn as the main barrier ($p = .002$; Table A.4). In contrast, habit formation is equally likely to be cited as a main barrier between the baseline and endline intervention groups ($p = .828$; Table A.4).

To summarize this evidence, after one day of experimentation with the newly encouraged behavior, there are substantial social norms-related concerns, but they 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.

Table A.15 and Table A.16 provide 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 pps less likely to report habit formation as a barrier and experience larger treatment effects on reported parental verbal inputs and infant cognition (Table A.15).[30] In Table A.16, we evaluate heterogeneity by adherence to our instructions to form a habit by filling in the stars on a calendar if they

[30]For balance tests on baseline characteristics and IDS beliefs and behavior for the sample of first-time mothers, refer to Appendix, Tables A.17 and A.18.

conversed with their infant everyday in a given week.[31] We see that filling in the stars is associated with larger effects on all outcomes of interest besides the observed infant cognition ($p = .144$) and LENA ($p = .343$) indices. While this result could be driven by selection, it is consistent with filling in the calendar helping the respondents form sticky habits.

5 Cost-effectiveness

Following Verguet et al. (2022) (a meta-analysis comparing the cost-effectiveness of 12 'responsive caregiving' interventions in LMICs), we estimate the cost per child's SD improvement in cognitive outcomes. We use the effects on mother-reported infant language development and cognitive development from Table 3 and estimate at-scale costs. We assume that the at-scale costs of the intervention would only include printing out the calendars and delivering them to health clinics. We average the effect on infant language development and the effect on infant cognitive development as our measure of the benefits of the intervention. Using these estimates, we calculate that the intervention delivers a 1 SD improvement in infant cognitive development for \$4.56. This would be lower than any of the interventions evaluated by Verguet et al. (2022) when using average labor costs, and lower than 11 out of 12 when using local labor costs.

Cost-effectiveness under various alternative assumptions is shown in Table 5. When we use observed measures instead of mother-reported measures, the estimate rises to \$7.34, which would still be more cost-effective than any intervention in Verguet et al. (2022). Using the intervention costs in our RCT yields an estimate of \$29.85 per SD improvement in infant cognitive development. 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 the costs at-scale 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 10 of the 12 interventions evaluated by Verguet et al. (2022) after standardizing labor costs.

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

6 Conclusion

This paper provides experimental evidence that a 'light-touch' intervention encouraging mother-infant conversations can generate substantial improvements in infant cognitive development. We estimate that this 'light-touch' approach is more cost-effective than alternative policies (such as home-visiting or community-based programs), delivering a 1 SD improvement in infant cognitive development for \$4 to \$8. We administered the intervention to patients in public health clinics after their prenatal or postnatal visit. 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. Finally, we identify local norms and habit formation as the main barriers to parent-infant conversations in our context. While our evidence suggests that local norms are mutable even over relatively, difficulties with habit formation could be more persistent. Future research could focus on complementary interventions that help mothers form an infant-directed speech habit after they have received a 'light-touch' informational intervention.

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," 2022.

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, "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).

Keller, Heidi, "Parenting and socioemotional development in infancy and early childhood," *Developmental Review*, 2018, *50* (January), 31–41.

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).

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

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.

Wang, Su Hua, Nora Lang, George C. Bunch, Samantha Basch, Sam R. McHugh, Salvador Huitzilopochtli, and Maureen Callanan, "Dismantling Persistent Deficit Narratives About the Language and Literacy of Culturally and Linguistically Minoritized Children and Youth: Counter-Possibilities," *Frontiers in Education*, 2021, *6* (July), 1–19.

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.

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.

Tables

Table 1: 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								0.330
Observations	1,408			705		703		

Note: Baseline data. The treatment group includes respondents who received the intervention at baseline, the control group all others. 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 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.

Table 2: Treatment Recall & Self-Reported Behavior Change

	Mean	SD	Count	N
Baseline 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 "Treatment Sample (Baseline Intervention)", the sample is restricted to respondents who received the 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.4 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. Surveyors first asked respondents *"We interviewed you in March. Do you recall anything specific about that interview?"* ("Without prompting" panel outcomes). Surveyors then reminded respondents they should have seen a video and received a calendar during the baseline interview and then asked if they recalled the content 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 told to color in the star for each week during which they had at least one conversation per day with their babies
In the panel 'Endline Intervention Sample', the sample is restricted to respondents who were sampled to receive the intervention at endline.

Table 3: 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 development index	Child development index	Child babbles	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.126	0.124	0.105	0.097	0.057	0.059	0.068
	(0.058)	(0.056)	(0.035)	(0.039)	(0.049)	(0.029)	(0.071)
	{0.030}	{0.025}	{0.003}	{0.014}	{0.238}	{0.040}	{0.335}
Control mean	0.00	0.00	0.00	0.00	0.00	0.52	0.00
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,258	1,258	1,258	1,258	1,167	867	775

Note: Endline and LENA day 1 recording data. For columns 1 to 6, regressions include controls for child's age in days, day of the survey, and surveyor gender. In columns 7 and 8, regressions include control 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 indexes are Anderson indices. 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 from the Oxford Neurodevelopment Assessment (Ox-NDA). The assessment was administered by the surveyor to the child during the survey. We restrict the sample to children aged 3 months and older since the development index score is not positively correlated with age prior to 3 months (see Figure A.3). "Child Babbles" (column 6) is a dummy equal to 1 if the surveyor observed the child babbling or mimicking at least 1-syllable babble at some point during the home visit. The outcome was added mid-data collection, hence is missing for a number of households. We restrict the sample to children 4 months old and older since no child in our sample babbles before age 4 months. 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 recording is kept in the analysis presented in this table. The analysis is further restricted to recording times between 10am to 7pm (this excludes 10/785 LENA day 1 recordings which have less than 9 hours (rounded up) of recording). "Child vocalization (% audio)" is the share of time the target child emitted vocalizations (including words, babbles, and pre-speech communicative sounds or "protophones" such as squeals, growls, or raspberries). Robust standard errors in parenthesis, p-values in curly brackets.

Table 4: Treatment Effects on Newly-Informed Mothers: Evidence From the 'Endline' Intervention

	Adult word per min (1)	Female adult word per min (2)	Male adult word per min (3)	Conversational turn per min (4)	% meaningful speech (5)
2nd day (post-intervention)	1.827	1.415	0.515	0.015	0.927
	(0.864)	(0.724)	(0.284)	(0.016)	(0.487)
	{0.036}	{0.052}	{0.071}	{0.352}	{0.058}
Mean Pre-intervention (Day 1)	20.12	16.37	3.65	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	372	372	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 10am to 7pm. 186/192 households had 2 complete audio recordings. 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 per minute is the average Conversational Turn Count per minute between adults and target child. "% meaningful speech" is the share of the audio categorized as vocalizations from the target child or speech/vocalizations from adults or other children near the target child. For further details on the LENA outcomes, please refer to Appendix, Section A.1. Robust standard errors in parenthesis, p-values in curly brackets.

Table 5: Cost-Effectiveness Calculations

	UnitCost	StdDevsEffect	CostperStdDev
At-scale costs; Mother-reported measures	.46	0.10	4.56
At-scale costs; Observed measures	.46	0.06	7.34
RCT costs ; Mother-reported measures	3.01	0.10	29.85
RCT costs ; Observed measures	3.01	0.06	48.01

Note: At-scale costs would only include the cost of printing each calendar and delivering them to health clinics ($0.46). 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 3. We use the LENA index as the observed measure of infant language development. Following Verguet et al. (2022), we take the average of the language and cognitive effects to get the average standard deviation effect of the intervention.

Appendix A

A.1 Appendix Figures

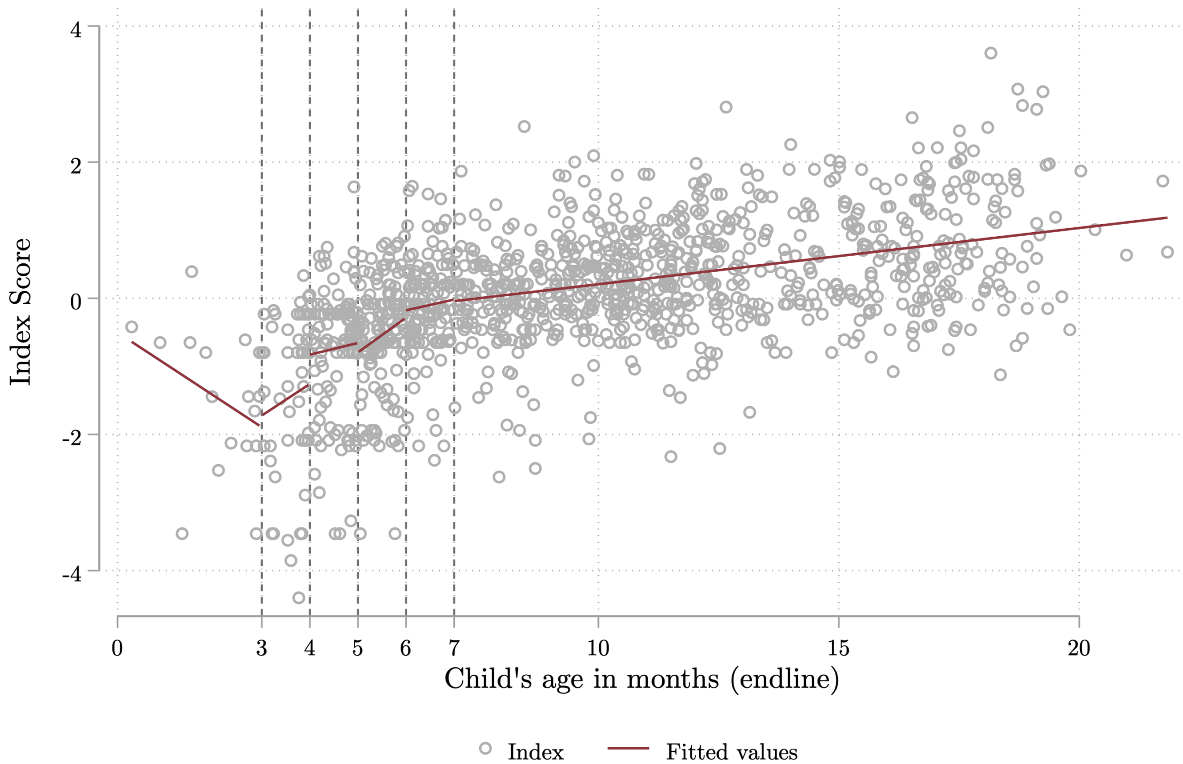
Figure A.1: Calendar for Treated Respondents

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 infant each day that week.

Figure A.2: Calendar for Control Respondents

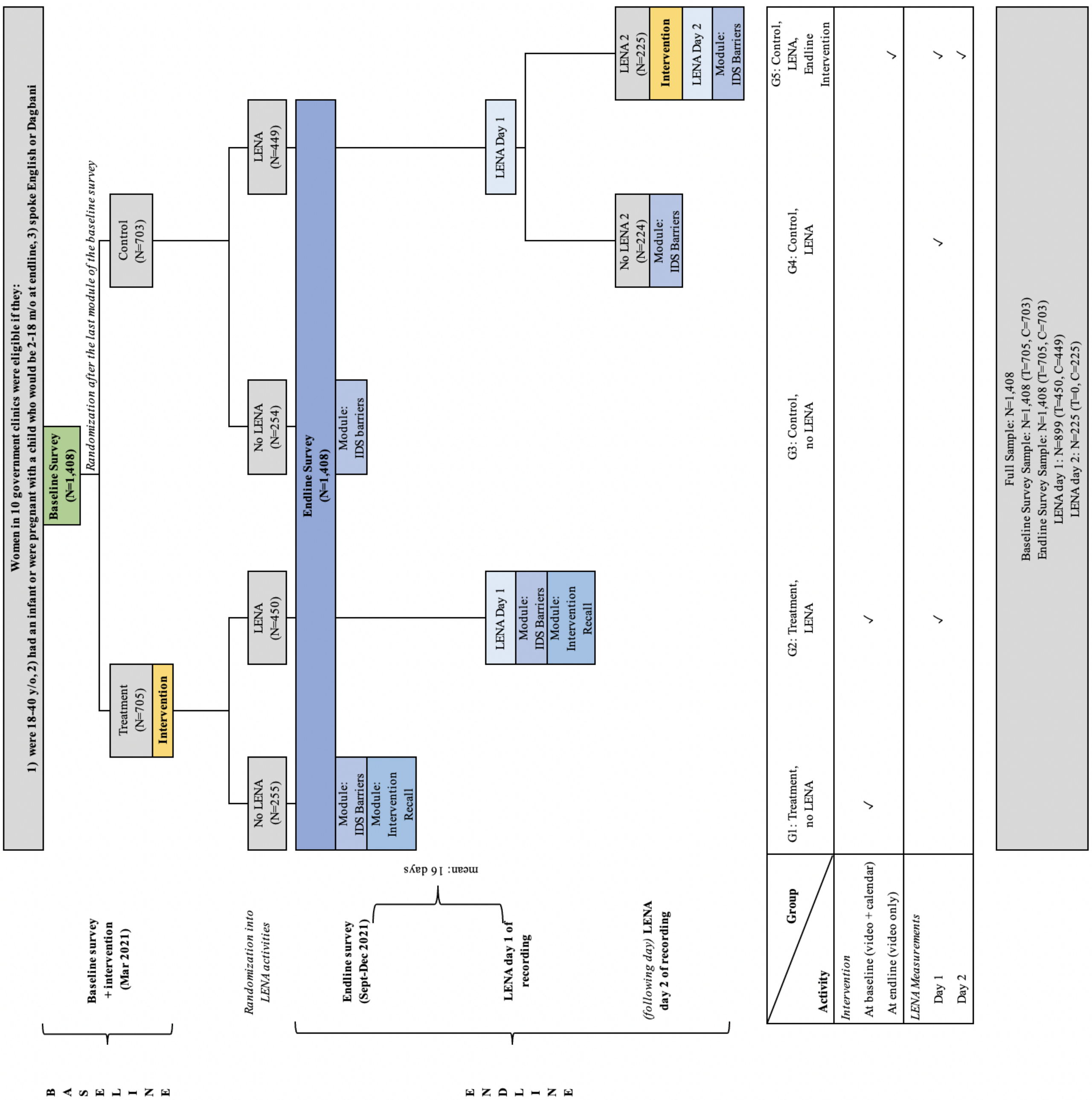
Note: 50% of the sample (N=703) received a regular calendar at the end of the baseline survey as a token of gratitude for participating in the survey. Control respondents did not see the IDS-themed video.

Figure A.3: Why We Focus on Children Aged 3 Months or More for the Objective Child Development Measure: Observed Child Development Index by Child Age



Note: Endline data. The Observed Child Development Index is an Anderson indexed, normalized over the control group. It is based on a selection of items from the Oxford Neurodevelopment Assessment (Ox-NDA). The assessment was administered by the surveyor to the child during the survey. For further details on the items included in the Observed Child Development Index, please refer to Table A.9. In the Figure above, we pool children aged 0 to 2 months in one group as there are few observations in that group (N=12 children with an Index Score). The Oxford Neurodevelopment Assessment is recommended for children aged 10 to 14 months. Since the Index Score is only positively correlated with children's age starting at age 3 months, we restrict our analysis of the Observed Child Development Index to the sample to children aged 3 months or more.

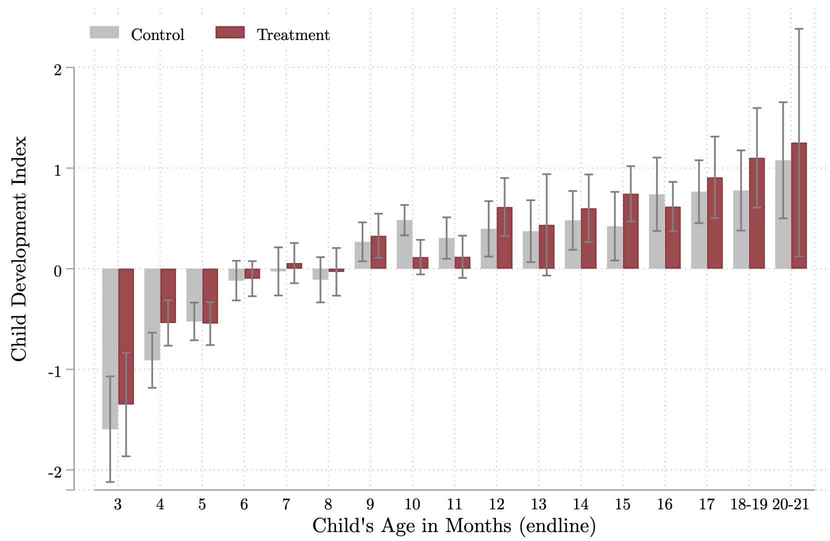
Figure A.4: Experimental Design and Timeline



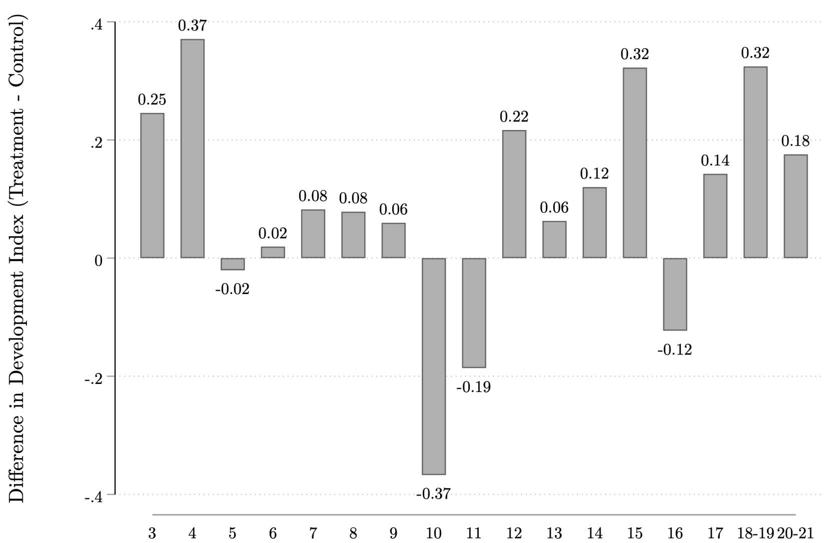
Note: See main text Section 2 for further details on the study design and timeline. On average, 6.4 months elapsed between the baseline and endline surveys.

Figure A.5: Observed Child Development Index by Child's Age (in Months)

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

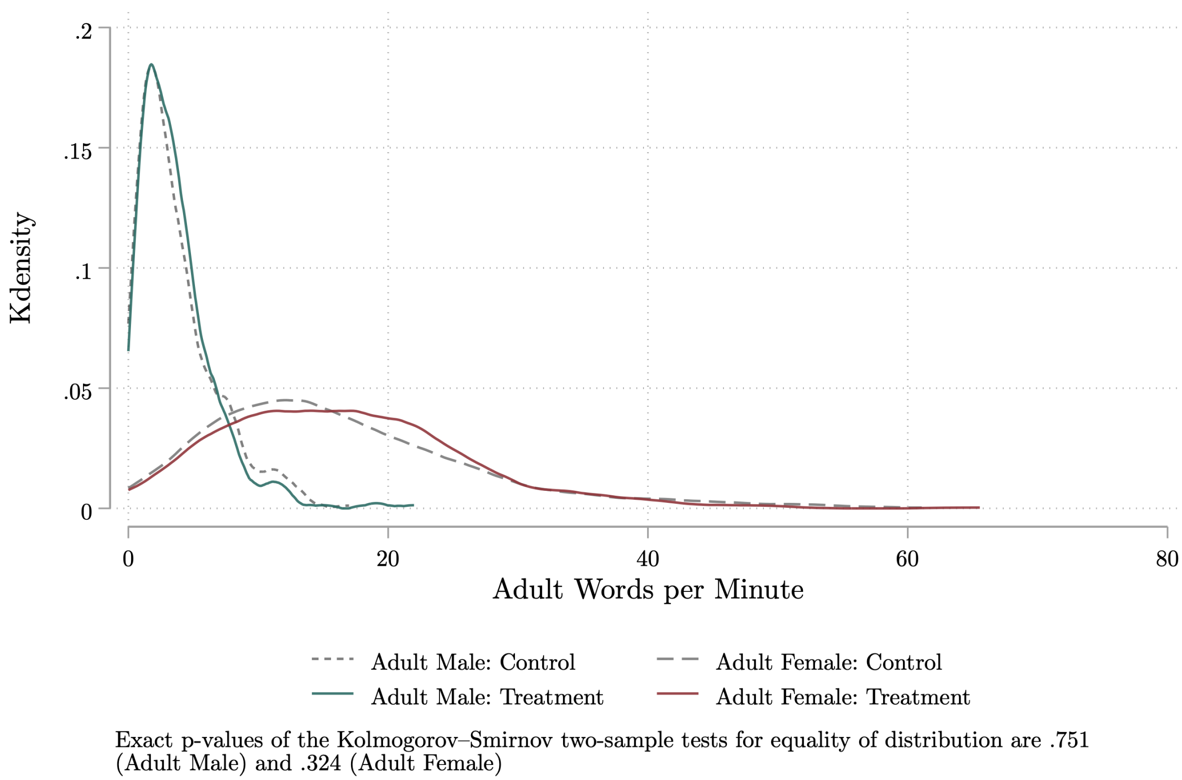


(b) Panel B: Difference in Means (Treatment - Control)



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 group. The sample is restricted to children aged 3 months and older, when the index score starts being positively correlated with children's age (see Figure A.3). We pool children aged 18-19 months and 20-21 months as there are few observations in those groups (see Figure A.8 for children's age distribution at endline). The Observed Child Development Index is an Anderson indexed, normalized over the control group. It is based on a selection of items from the Oxford Neurodevelopment Assessment (Ox-NDA). For further details on the items included in the Observed Child Development Index, please refer to Table A.10.

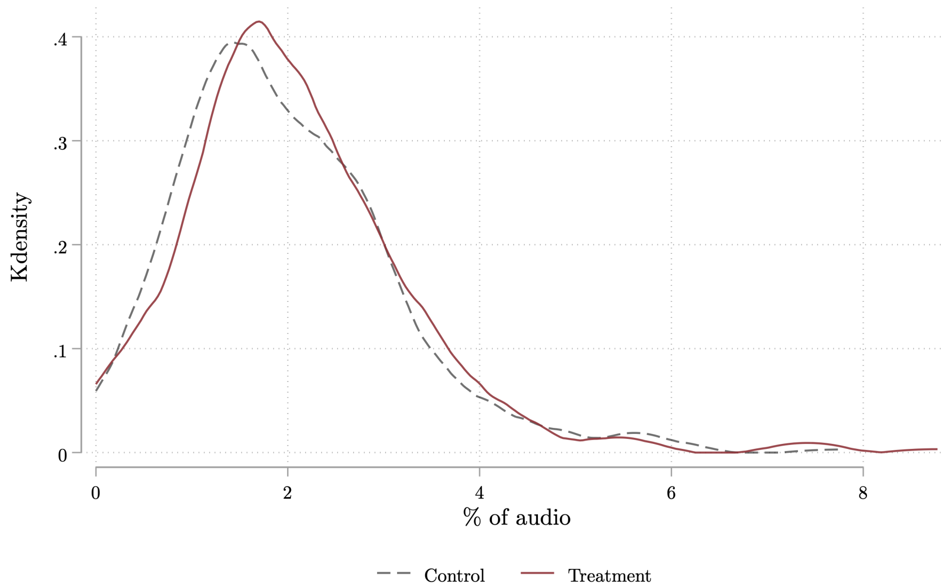
Figure A.6: LENA Measurements: Adult Word per Minute by Speaker Gender and Treatment



Note: LENA data. N=775 recordings. Please refer to Figure B.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 environment. Adult word count per minute is the estimate 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.1.

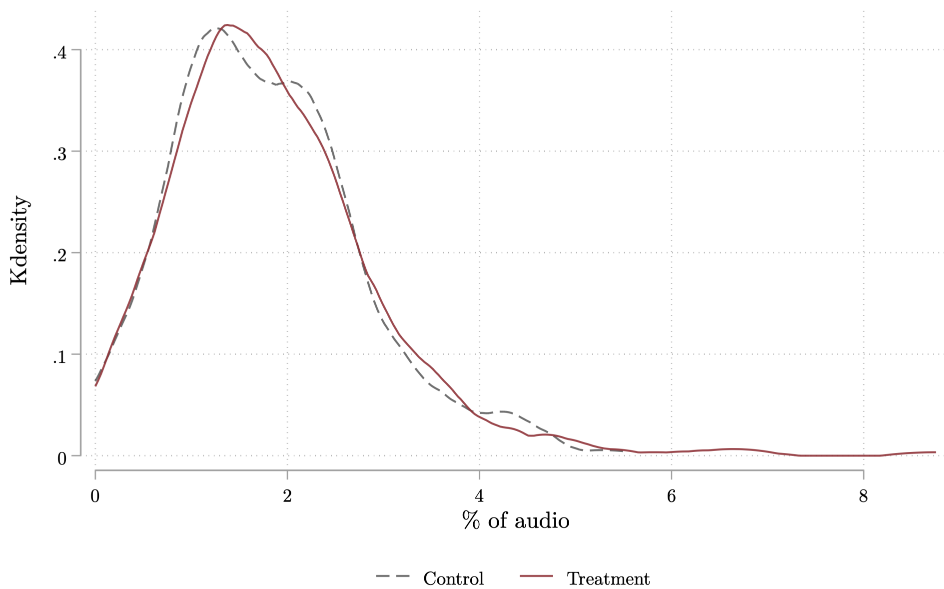
Figure A.7: LENA Measurements: Target Child Sounds by Treatment

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



Exact p-value of the Kolmogorov–Smirnov two-sample test for equality of distribution is .181

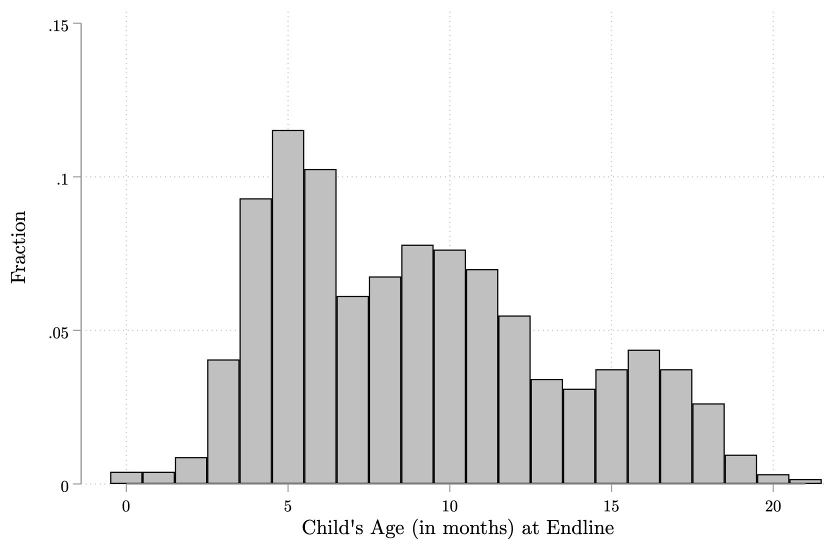
(b) Panel B: Non-Speech Sounds (% of total audio time)



Exact p-value of the Kolmogorov–Smirnov two-sample test for equality of distribution is .915

Note: LENA data. N=775 recordings. Please refer to Figure B.2 for details on the sample. The LENA technology allows us to analyze the target child segment sounds to distinguish and quantify target child **vocalizations** (including words, babbles, and pre-speech communicative sounds or "protophones" such as squeals, growls, or raspberries) from **non-speech** (including fixed signals and vegetative sounds). For further details on the LENA outcomes, please refer to Section A.1.

Figure A.8: Distribution of Infant Ages at Endline



Note: Endline data. N=1,258/1,408 households with children participated in the endline survey. Average infant age is 9.6 months in the treatment group and 9.5 months in the control group.

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: 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	0.89	0.31	1,381	0.89	0.31	0.90	0.30	0.646
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
Age at first child (years)	22.23	3.50	1,242	22.14	3.37	22.33	3.62	0.327
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
# of children	2.21	1.55	1,408	2.28	1.54	2.15	1.57	0.105
# of children at home	2.07	1.50	1,408	2.14	1.49	2.00	1.52	0.071
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
Target child is first born	0.28	0.45	1,408	0.26	0.44	0.30	0.46	0.129
F-test p-value								0.927
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.

Table A.3: 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
Child aged $\geq$ 3 months (if available)	0.99	0.12	1,186	1,203	0.99	0.09	606	0.98	0.14	580	0.078
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 respondent were shown the intervention video (see Section 2 and Figure A.4 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 10am and 7pm, and, hence, is kept in the analysis.

Table A.4: Reported Barriers to IDS

	(1) Pure Control	(2) Baseline Intervention (6-8 mo ago)	(3) Endline Intervention (day before)	(4) P-value Baseline 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 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.4). 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 2nd day of recording. Column 1 presents the means for the pure control group (never received the intervention), column 2 for respondents who received the intervention at baseline (their views incorporate their experience with IDS over the past 6 to 8 months between the endline survey and the intervention), 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 colums 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.5: 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 development index	Child development index	Child babbles	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Without Controls							
Treatment	0.125	0.126	0.118	0.108	0.077	0.061	0.054
	(0.060)	(0.056)	(0.055)	(0.055)	(0.056)	(0.033)	(0.071)
	{0.036}	{0.026}	{0.031}	{0.048}	{0.167}	{0.066}	{0.447}
Control mean	0.00	0.00	0.00	0.00	0.00	0.52	0.00
Controls	No	No	No	No	No	No	No
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,258	1,258	1,258	1,258	1,167	867	775
Panel B: With Clinic-Day Fixed Effects							
Treatment	0.137	0.133	0.110	0.106	0.052	0.065	0.061
	(0.058)	(0.055)	(0.036)	(0.040)	(0.049)	(0.030)	(0.074)
	{0.018}	{0.016}	{0.002}	{0.008}	{0.295}	{0.029}	{0.406}
Control mean	0.00	0.00	0.00	0.00	0.00	0.52	0.00
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	No	No	No	No	No	No	No
Clinic-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,258	1,258	1,258	1,258	1,167	867	775
Panel C: With Surveyor Fixed Effects							
Treatment	0.180	0.119	0.090	0.069	0.031	0.038	0.049
	(0.054)	(0.051)	(0.031)	(0.038)	(0.047)	(0.027)	(0.072)
	{0.001}	{0.021}	{0.004}	{0.067}	{0.514}	{0.170}	{0.494}
Control mean	0.00	0.00	0.00	0.00	0.00	0.52	0.00
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Surveyor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,258	1,258	1,258	1,258	1,167	867	775

Note: Endline and LENA day 1 recording data. This table presents the results from running the main specifications (presented in main text Table 3) without any control (Panel A); replacing clinic FE by clinic-day fixed effects (Panel B); adding surveyor fixed effects (Panel C). Robust standard errors in parenthesis, p-values in curly brackets. Since 3 surveyors at endline did less than 30 surveys each, we grouped them as one surveyor when including Surveyor Fixed Effects (in Panel C).

Table A.6: Treatment Effects on Parental Beliefs

	Treatment Effect			Control Group		
	Coefficient (1)	SE (2)	P-value (3)	Mean (4)	SD (5)	N (6)
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 the following statements:						
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 (variable indicated on the left) is regressed on a dummy equal to 1 if the household received the intervention at baseline (treatment). 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 "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 3, all regressions include clinic fixed effects and control for child's age (in days), survey date, and surveyor gender.

Table A.7: Treatment Effects on Self-Reported Parenting Behavior

	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 weeks, how often did you...						
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 (variable indicated on the left) is regressed on a dummy equal to 1 if the household received the intervention at baseline (treatment). 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 3, all regressions include clinic fixed effects and control for child's age (in days), survey date, and surveyor gender.

Table A.8: Treatment Effects on Self-Reported Child Behavior

	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...						
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
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
Attempt to say words (yes/no)	0.044	0.021	0.038	0.42	0.49	1,257

Note: Each line reports the result of a different regression for which the (variable indicated on the left) is regressed on a dummy equal to 1 if the household received the intervention at baseline (treatment). 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. Respondents were asked to choose from a 3-point Likert scale (not yet (1), sometimes, and often (3)). As for the main text Table 3, all regressions include clinic fixed effects and control for child's age (in days), survey date, and surveyor gender.

Table A.9: Treatment Effect on Observed Child Development ($\geq$ 3 Months Old)

	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	0.046	0.031	0.141	2.78	0.59	1,155
Watches toy placed in front	0.013	0.022	0.570	2.91	0.41	1,163
Identifies spoon correctly when asked	-0.032	0.059	0.592	2.26	1.04	1,125
Child Assessment: 1=worst to 4=best						
Imitates or tries to imitate bi-syllabic words	-0.009	0.035	0.806	1.28	0.62	1,116
Reacts to name when playing	0.119	0.072	0.100	2.75	1.27	1,157
Stops reaching for toy when told no	-0.015	0.062	0.806	1.76	1.09	1,100
Uses or mimics words in play context	0.007	0.030	0.812	1.19	0.49	1,123
Babbles or attempts to when prompted	0.054	0.046	0.241	1.40	0.81	1,125
Combines word and gesture (correctly or not)	-0.043	0.047	0.353	1.46	0.86	1,124

Note: Each line reports the result of a different regression for which the (variable indicated on the left) is regressed on a dummy equal to 1 if the household received the intervention at baseline (treatment). 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 sample is restricted to children aged 3 months or more since the assessment score is not positively correlated with age prior to 3 months (see Figure A.3 for details). The assessment was administered to the child by the surveyor during the survey. Tasks were adapted from the Ox-NDA tool. Observations are missing when the surveyor was unable to assess the child (because child became agitated, started crying, was uncooperative etc.). Each task was evaluated using a scale form 1 to 5 with 1 being the highest score and 4 the lowest one. A score of 5 indicates the surveyor was unable to asses child (because the infant was out of sight, sleeping, crying, became too agitated, etc.) and are recoded as missing. As for the main text Table 3, all regressions include clinic fixed effects and control for child's age (in days), survey date, and surveyor gender.

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)
Outcomes in the index						
Female adult word per min	0.172	0.698	0.805	16.57	10.41	775
Male adult word per min	-0.048	0.215	0.824	3.75	3.00	775
Conversational turns per min	0.004	0.017	0.815	0.41	0.25	775
Child vocalizations per min	0.052	0.058	0.377	1.53	0.82	775
Other child speech (% audio)	0.144	0.153	0.347	3.45	2.14	775
Key child vocalizations (% audio)	0.095	0.082	0.247	1.99	1.14	775
Other outcomes (not in the index)						
Adult word count per min	0.070	0.786	0.929	20.37	11.75	775
Female adult speech (% audio)	0.113	0.302	0.708	7.30	4.50	775
Male adult speech (% audio)	0.026	0.102	0.798	1.87	1.42	775
Key child sounds (% audio)	0.174	0.148	0.242	4.13	2.03	775
Key child non-vocalizations (% audio)	0.066	0.074	0.377	1.79	0.98	775
Meaningful speech (% audio)	0.298	0.499	0.550	16.80	7.08	775
Speech far from child (% audio)	0.750	1.108	0.499	39.07	15.67	775
Silence (% audio)	-0.336	1.120	0.764	29.79	15.31	775
Non-speech noise (% audio)	-1.361	0.483	0.005	7.15	7.23	775
TV/Electronics (% audio)	0.027	0.618	0.965	7.18	9.70	775

Note: LENA data. Each line reports the result of a different regression for which the (variable indicated on the left) is regressed on a dummy equal to 1 if the household received the intervention at baseline (treatment). 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 10am to 7pm (this excludes 10/785 LENA day 1 recordings which have less than 9 hours (rounded up) of recording). As in Columns 7 and 8 of main text Table 3, regressions include clinic fixed effects and controls for the day of the week, survey date, child's age (in days), household size, time (min) child was held on someone's back while wearing the device, time (min) the shirt/device was removed from the child. For further details on the LENA outcomes, please refer to Section A.1.

Table A.11: Treatment Effect Split by Susceptibility to Social Desirability Bias

	Mother's Interview				Observed		LENA
	Mother's belief index	Mother's behavior index	Child language score	Child development index	Child development index	Child babbles	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.181	0.143	0.115	0.098	0.069	0.059	0.066
	(0.060)	(0.059)	(0.038)	(0.042)	(0.052)	(0.031)	(0.077)
	{0.003}	{0.015}	{0.003}	{0.020}	{0.185}	{0.056}	{0.392}
Treatment x Did not associate intervention w/ surveyor	-0.260	-0.033	-0.030	0.011	-0.055	0.003	0.011
	(0.121)	(0.103)	(0.053)	(0.066)	(0.084)	(0.050)	(0.104)
	{0.032}	{0.749}	{0.572}	{0.866}	{0.510}	{0.957}	{0.917}
Control mean	-0.00	-0.00	-0.00	-0.00	0.00	0.52	-0.00
P-val Treatment and Did not associate	0.509	0.278	0.097	0.087	0.870	0.210	0.432
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,248	1,248	1,248	1,248	1,160	860	774
Observations: Did not associate	131	131	131	131	124	87	72

Note: Endline and LENA day 1 recording data. See main text Table 3 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 baseline intervention and were not asked those questions. N=1,248 instead of N=1,258 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. See main text Table 2 for further details on questions and sample. Robust standard errors in parentheses, p-values in curly brackets.

Table A.12: Robustness of Treatment Effects on Observed Child Development

	Observed Child Development Index	
	Replaced by	
	Missing	Lowest Score
	(1)	(2)
Treatment	0.057	0.045
	(0.049)	(0.050)
	{0.238}	{0.362}
Control mean	0.00	0.00
Controls	Yes	Yes
Clinic FE	Yes	Yes
Observations	1,167	1,167

Note: Endline data. This table presents the results from running the main specifications (presented in main text Table 3). In main text Table 3 and column (1) of this table, test components reported as "Unable to assess" by the surveyor (because the infant became agitated, refused to participate etc.) are left as missing when computing the Anderson index. In column (2), the index is computed replacing missing components by the lowest possible score. Robust standard errors in parenthesis, p-values in curly brackets.

Table A.13: 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 development index	Child development index	Child babbles	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.187	0.106	0.086	0.062	0.052	0.044	0.045
	(0.051)	(0.051)	(0.031)	(0.037)	(0.048)	(0.027)	(0.071)
	{0.000}	{0.040}	{0.006}	{0.090}	{0.283}	{0.103}	{0.531}
Control mean	0.00	0.00	0.00	0.00	0.00	0.52	0.00
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	No	No	No	No	No	No	No
Observations	1,258	1,258	1,258	1,258	1,167	867	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 equation. Selected controls in each regression include: (1) education (years), dummy for completing tertiary education, dummy for reading English, partner's education (years), dummy for partner completing secondary education, baseline belief about age at which one should start telling stories to infant, dummy for never married and not cohabiting, dummy for separated, dummy for widowed, 1 baseline health facility dummy, 5 endline surveyor dummies, 4 endline location dummies, and dummy for missing partner's education; (2) dummy for Dagbani as main language at home, interview language (endline), dummy for never married and not cohabiting, dummy for separated, dummy for widowed, dummy for other African tribe, 5 endline surveyor dummies, 4 endline location dummies, and dummy for missing partner's education; (3) dummy for youngest child eligible (baseline) baseline, child's age in days (endline), dummy for child born since baseline eligible (endline), pregnancy status (baseline), dummy for never married and not cohabiting, dummy for separated, dummy for widowed, 1 baseline health facility dummy, 5 endline surveyor dummies, 5 endline location dummies, and dummy for missing partner's education; (4) youngest child's age in years (baseline), breastfeeding intentions (baseline), currently breastfeeds youngest child (baseline), child's age in days (endline), interview language (endline), dummy for feeding youngest child with a combination of solid food and breast milk (baseline), dummy for separated, dummy for widowed, 4 endline surveyor dummies, 2 endline location dummies, and dummy for missing partner's education.(5) youngest child eligible (baseline), child's age in days (endline), dummy for child born since baseline eligible (endline), dummy for separated, dummy for widowed, 1 endline surveyor dummy, and dummy for missing partner's education; (6) youngest child eligible (baseline), child's age in days (endline), dummy for child born since baseline eligible (endline), pregnancy status, dummy for separated, dummy for widowed, 2 endline surveyor dummies, 2 endline location dummies,and dummy for missing partner's education;(7) dummy for partner completing A-levels, dummy for partner retired, dummy for partner in school, dummy for roof concrete, 1 endline location dummy, dummy for missing number of adults aged at least 50 years. 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)
Baseline 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=Baseline 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 day of recording (see Figure A.4 for further details on the experimental design and timing of the "IDS Barriers" Module). Respondents were asked about barriers that may prevent families from talking to their babies. 3 specific barriers respondents were asked about were: "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 child's age (in days), endline survey date, LENA survey date, 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 development index	Child development index	Child babbles	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.143	0.090	0.071	0.055	0.063	0.036	0.064
	(0.070)	(0.064)	(0.041)	(0.045)	(0.057)	(0.034)	(0.084)
	{0.041}	{0.156}	{0.087}	{0.228}	{0.271}	{0.292}	{0.447}
Treatment x 1st-time mother	-0.051	0.132	0.122	0.159	-0.027	0.087	-0.021
	(0.128)	(0.127)	(0.078)	(0.091)	(0.110)	(0.063)	(0.148)
	{0.691}	{0.300}	{0.120}	{0.082}	{0.807}	{0.169}	{0.887}
1st-time mother	0.116	-0.000	-0.085	-0.022	-0.041	-0.022	-0.174
	(0.084)	(0.090)	(0.055)	(0.062)	(0.075)	(0.044)	(0.105)
	{0.166}	{0.997}	{0.125}	{0.723}	{0.587}	{0.625}	{0.098}
Control mean not 1st-time mother	-0.02	-0.01	-0.06	-0.07	-0.03	0.50	0.06
P-val Treatment and 1st-time mother	0.385	0.045	0.004	0.007	0.696	0.021	0.725
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,258	1,258	1,258	1,258	1,167	867	775
Observations: 1st-time mother	347	347	347	347	333	238	214

Note: Endline and LENA day 1 recording data. See main text Table 3 for details on specifications and outcomes. "1st-time mother" is a dummy equal to 1 if the target child is the first born of the respondent. Robust standard errors in parenthesis, p-values in curly brackets.

Table A.16: 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 development index	Child development index	Child babbles	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.102	0.075	0.092	0.062	0.082	0.036	0.074
	(0.066)	(0.068)	(0.043)	(0.046)	(0.055)	(0.033)	(0.083)
	{0.123}	{0.267}	{0.030}	{0.180}	{0.137}	{0.278}	{0.375}
Treatment x Colored stars	0.135	0.170	0.041	0.104	0.013	0.061	0.029
	(0.100)	(0.086)	(0.058)	(0.069)	(0.072)	(0.043)	(0.116)
	{0.178}	{0.049}	{0.479}	{0.130}	{0.854}	{0.155}	{0.803}
Control mean	-0.01	-0.00	0.02	0.02	0.03	0.52	0.01
P-val Treatment and Colored stars	0.012	0.002	0.013	0.011	0.144	0.014	0.343
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,136	1,136	1,136	1,136	1,061	860	724
Observations: Colored stars	188	188	188	188	173	167	117

Note: Endline and LENA day 1 recording data. See main text Table 3 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 baseline 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 2). See Figure A.1 for further details on the calendar. Robust standard errors in parenthesis, p-values in curly brackets.

Table A.17: 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	0.88	0.33	386	0.89	0.31	0.86	0.34	0.437
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
Age at first child (years)	23.03	3.39	259	22.80	3.43	23.27	3.35	0.266
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
# of children	0.74	0.71	393	0.81	0.82	0.67	0.60	0.056
# of children at home	0.72	0.66	393	0.78	0.72	0.67	0.60	0.095
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
Target child is first born	1.00	0.00	393	1.00	0.00	1.00	0.00	.
F-test p-value								0.085
Observations	393			184		209		

Note: Baseline data. Sample restricted to mothers whose child enrolled in the study was their first born. 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.

Table A.18: 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								0.376
Observations	393			184		209		

Note: Baseline data. Sample restricted to mothers whose child enrolled in the study was a first born. 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.

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 captured by the device is not directly accessible. Instead, it is processed by a cloud-based LENA software which provides detailed information on the audio environment of the child. Available outcomes include minutes of the audio coming from different sources (child, other children, adults, TV/electronics, silence, non-speech noise), categorization of those sounds by type (meaningful speech, far speech, words, vocalizations, etc.), number of words by adults or other children, number of child vocalizations and number of conversational turns. Those are described in further details in the next section.

Figure B.1: LENA device and shirt

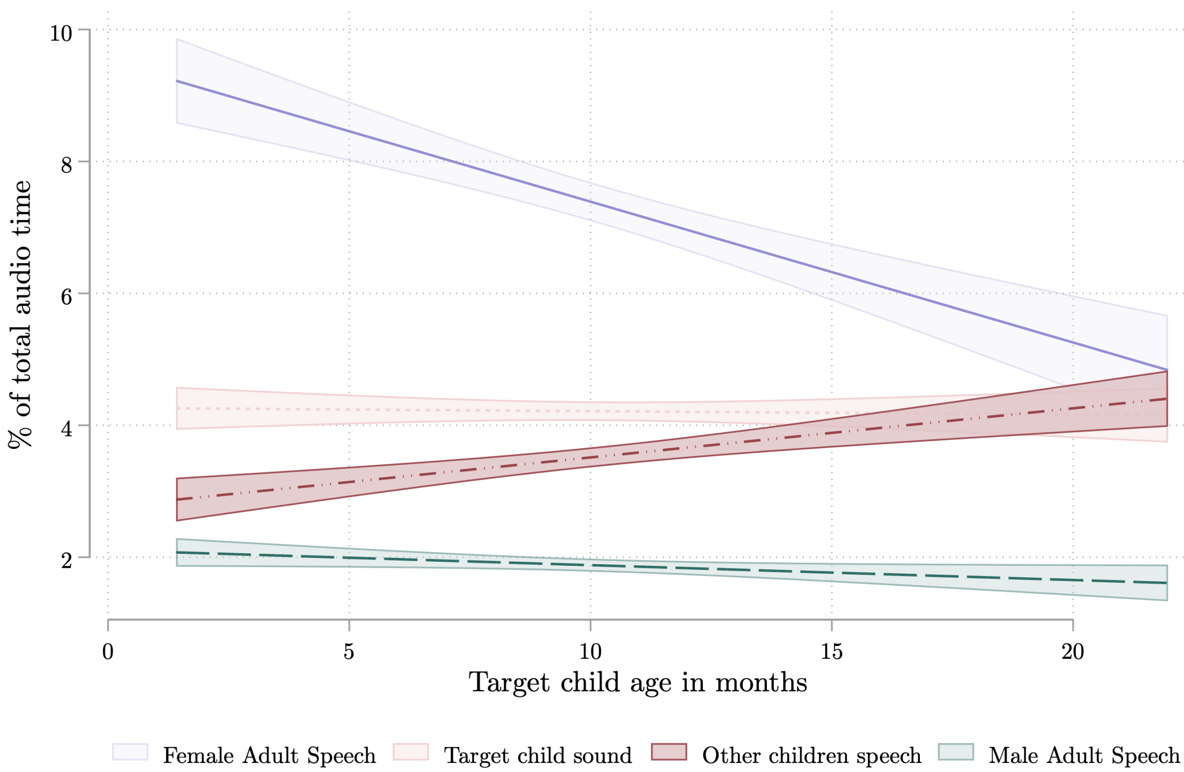
Source: Picture of a child participating in the experiment and wearing the LENA shirt with the device inserted in the front pocket. The LENA device was purchased from the LENA Foundation and the shirt was designed by our research team.

Description of LENA outcomes

In this paper, we focused on the LENA outcomes summarized below. For further details and more complete descriptions of those outcomes, please refer to the LENA technical guide available at https://www.lena.org/wp-content/uploads/2020/07/LTR-12_How_LENA_Works.pdf (Gilkerson and Richards, 2020).

- **Adult Word Count:** estimated number of words spoken by post-pubescent individuals.
- **Other Child Speech:** vocalizations by pre-pubescent children within 6 to 10 feet of the child wearing the device.
- **Target Child Total Sound:** any sound originating from the mouth of the child wearing the device, including speech-related vocalizations (such as babbles, words, and sentences) and non-speech sounds (such as vegetative sounds (sounds related to respiration or digestion such as burps, breaths) and fixed signals (instinctive reactions to the environment such as cries, screams, laughs)).
- **Target Child Vocalization Count**: estimated number of times the child produced communicative vocalizations (including words, babbles, and pre-speech communicative sounds or "protophones" such as squeals, growls, or raspberries). This excludes vegetative sounds or fixed signals.
- **Conversational Turns Count:** estimated number of alternations between the child wearing the device and an adult in the vicinity. More precisely, it estimates the number of alternations between a segment of sound including a vocalization from the target child wearing and an adult sound segment (including a word).

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 10am to 7pm (this excludes 10/785 LENA day 1 recordings which have less than 9 hours (rounded up) of recording). Outcomes are % of total audio time. "Sound by target child (as % of audio)" is the share of the recording tagged as a sound emitted by child (vocalizations or 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 10am to 7pm (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 LENAr System," *LENA Foundation Technical Report*, 2020