

## Informing Mothers about the Benefits of Conversing with Infants: Experimental Evidence from Ghana<sup>†</sup>

By PASCALINE DUPAS, CAMILLE FALEZAN,  
SEEMA JAYACHANDRAN, AND MARK WALSH\*

*We evaluate a low-cost intervention designed to boost parents' verbal engagement with infants, which tends to be limited in developing countries. In our randomized experiment, recent or expectant mothers watched a three-minute informational video and received a themed calendar. Six months later, treated mothers reported stronger belief in the benefits of verbal engagement, more frequent parent-infant conversation, and more advanced infant language skills. Treatment effects on objective measures of parent-child conversation frequency and infant skills were positive but insignificant. We find larger immediate treatment effects on objective parent-child conversation, suggesting potentially larger long-term effects had the behavior change stuck more. (JEL D83, D91, I26, J13, J16, O12)*

While parents universally use “baby talk” to soothe an infant or get their attention, engaging in a richer form of infant-directed speech (IDS)—which 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 (Hart and Risley 1995; Hoff 2003) and across (Farran et al. 2016) societies. Given the benefits of parent-infant conversations for cognitive development (Monnot 1999; Weisleder and Fernald 2013), these SES gaps are likely to compound the disadvantages that children in poorer families face. The problem may be especially acute in lower-income countries, where 43 percent of children under 5 years old (over 250 million children) are at risk of not reaching their developmental potential (Black et al. 2017).

One explanation for parental underinvestment in conversing with infants is inaccurately low expectations about the benefits. A large body of literature in the United

\*Dupas: Princeton University, NBER, and CEPR (email: [pdupas@princeton.edu](mailto:pdupas@princeton.edu)); Falezan: MIT (email: [cfalezan@mit.edu](mailto:cfalezan@mit.edu)); Jayachandran: Princeton University (email: [jayachandran@princeton.edu](mailto:jayachandran@princeton.edu)); Walsh: Stanford University (email: [mwalsh24@stanford.edu](mailto:mwalsh24@stanford.edu)). Heather Royer was coeditor for this article. This research was supported by grants from the World Bank’s Strategic Impact Evaluation Fund and Stanford University. It builds on pilot work made possible by funding from the Weiss Fund (grant number STN-008). 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. We are grateful to three anonymous referees for their constructive feedback. The research protocol was approved by the Ghana Health Service Ethical Review Committee (GHS-ERC 002/12/20) and by the Stanford University IRB (51503). We registered the RCT in the AEA RCT Registry (ID AEARCTR-0007161) and at ClinicalTrials.gov (ID NCT04807907) (Dupas, Jayachandran, and Walsh 2021a,b). Data and code are available on openICPSR (Dupas et al. 2025).

<sup>†</sup>Go to <https://doi.org/10.1257/pol.20230283> to visit the article page for additional materials and author disclosure statement(s).

States has shown that lower-SES parents are less aware of the returns to verbal engagement with infants (List, Pernaudet, and Suskind 2021). In the Northern Region of Ghana, our study setting, only 11 percent of mothers reported that parents should start talking to their baby at birth,<sup>1</sup> and 61 percent reported that they should begin talking to the child once he or she is older than 6 months (Duflo et al. 2024). While one might have hoped that rising educational attainment in low- and middle-income countries (LMICs) would narrow this gap, Duflo et al. (2024) find that a subsidy program that increased secondary school completion in Ghana did not change maternal beliefs about the importance of conversing with infants. The persistence of misperceptions is perhaps unsurprising, as it is not intuitive that speaking to a one-week- or even three-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 be a practice that arises not organically, but only by parents explicitly being taught its value. If this explanation is correct, a cheap information intervention might be enough to correct parental beliefs, cause behavior change, and cost-effectively enhance infant development outcomes.

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

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

<sup>1</sup> This is lower than in other contexts. For example, in urban areas of Turkey, 50 percent 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 LMICs such as Morocco (Zellman, Karam, and Perlman 2014) and Nepal (Shrestha et al. 2019).

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

Given the "light-touch" nature of the intervention, we planned for the possibility that mothers might initially change their behavior but revert to their preintervention behavior by the time of the follow-up survey (six to eight months later). To distinguish between participants never embracing the recommended practices and participants adopting them initially but not persisting, we also administered our informational video treatment to a subset of control-group respondents the morning after they completed their daylong endline LENA recording. We refer to this as the "endline intervention" to distinguish it from the main intervention described above that was delivered at baseline. For these respondents, we recorded the child's auditory environment again over the next day. To estimate the effect of the endline intervention, we compare the LENA measures for these respondents from *the day before* receiving the intervention to the same measures *the day after*. The newly informed mothers speak 1.4 more words per minute ( $\approx 8.4$  percent of the mean,  $p = 0.058$ ) postintervention relative to their preintervention levels. This effect is almost 11 times larger than the main intervention effect estimated after six to eight months. The large immediate impact of the endline intervention on parental behavior shows that mothers are willing and able to verbally engage with their children when (i) they are told that they should and (ii) 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 that they should do it, mothers know *how* to do it. But the fact that treatment effects of the main intervention are much lower after six to eight months suggests that sustaining the behavior over time is difficult.

When we ask mothers who received either the main intervention or endline intervention about likely barriers to parent-infant conversations, the most common answers are fear of social sanctions (scorn) and difficulty in habit formation. The relative importance of these two perceived barriers varies by time since the intervention delivery. Mothers who received the main intervention at baseline (i.e., six to eight months prior to being asked about barriers) are 41 percent less likely to report social scorn or mockery as the main barrier ( $p = 0.002$ ), compared to

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

mothers who received the intervention at endline (i.e., less than 24 hours prior to being asked about barriers). In contrast, habit formation is equally likely to be cited as the main barrier across the two groups. One interpretation of this pattern is that mothers quickly get over the social awkwardness of verbally engaging their infants, while transforming a new behavior into a sticky habit is fundamentally difficult (Rothman et al. 2015; Webb and Sheeran 2006; Lally and Gardner 2013).

Recent meta-analyses have already shown that there is strong evidence that interventions encouraging “responsive caregiving” (which includes parent-infant conversations) promote maternal knowledge and mother-infant interactions, but our intervention is cheaper and lighter-touch than any of the studies included (Jeong, Pitchik, and Yousafzai 2018; Jeong et al. 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 the experiments of Suskind et al. (2018) and List, Pernaudet, and Suskind (2021) in the Chicago metropolitan area in the United States. Suskind et al. (2018) find significant effects on parental beliefs from mothers watching a ten-minute video but does not measure parental behavior or child development. List, Pernaudet, and Suskind (2021) evaluate the effect of mothers watching ten-minute videos when their child is one, two, four, and six months old and measures persistent effects on beliefs and short-run effects on parental verbal inputs, but noisy null effects on mother-reported vocabulary. Our study tests an intervention that is significantly shorter and finds positive effects on infant language and cognitive development, which were unmeasured outcomes in Suskind et al. (2018) and may have been undetected in List, Pernaudet, and Suskind (2021) due to a lack of statistical power.<sup>3</sup>

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

We estimate that the cost per child beneficiary in our research trial was \$3.01. At scale, one could use existing health center staff rather than surveyors to hand out the calendar and show the video, lowering the cost to \$0.45 per child. The low cost, combined with the treatment effects we estimate, implies that the intervention could be among the most cost-effective known ways for LMICs to increase infant language development. We calculate that at scale it would deliver a 1 SD improvement in a child’s cognitive or language development for \$4.52 to \$10.01, depending on whether we use our reported or observed measures of child development (or \$30 to \$67 under our research trial conditions). In Verguet et al.’s (2022) meta-analysis of 12 early childhood interventions, the median intervention costs \$328 per SD

<sup>3</sup>List, Pernaudet, and Suskind’s (2021) sample size is 475, compared to our sample size of 1,408.

improvement in child cognitive or language development.<sup>4</sup> Even if the true effect of our intervention, or the effect achievable at scale, were a quarter of the magnitude we estimated (so the cost becomes \$18.09 to \$40.02 per SD improvement), the intervention would still be more cost-effective than 11 of the 12 interventions assessed by Verguet et al. (2022).

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

## II. Study Design

### A. Sampling and Intervention

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

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

Half of the respondents were randomly allocated to the treatment group and selected to watch a three-minute intervention video (see <https://www.facebook.com/ghanababytalk>) and receive the intervention calendar at the end of the baseline survey (see Supplemental Appendix 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

<sup>4</sup>We convert Verguet et al.’s (2022) estimates to 2021 US dollars to facilitate comparisons. Among the 12 interventions they assess, 10 cost over \$50 per child, and the cheapest costs \$22 per child.

<sup>5</sup>The average monthly household income in the Northern Region is ~\$38, while the national average is ~\$156 (Antwi and Lyford 2021).

<sup>6</sup>Of the 1,765 women approached, 1,462 were eligible. 283 were ineligible because of their child’s age or due date, 17 because of their age, and 3 because they did not speak English or Dagbani. Of the 1,462 eligible women, 1,408 completed the survey and were administered the intervention. Thirty-eight chose not to participate partway through the baseline survey, 15 refused to participate, and 1 did not pass the COVID-19 symptom screening.

that “back-and-forth moments” are particularly important for child development. The video then provides a few ideas for how to converse with your infant, such as describing what you see “when you are walking across the village or town with her,” telling your baby what you are preparing “when you are cooking,” “telling stories,” “singing songs,” or describing pictures in/“reading books.” This narration is paired with images of family members talking to an infant while doing the suggested activities. In short, the video informs mothers about the benefits of verbal engagement with infants and about how to verbally engage an infant. The intervention calendar highlights a few key points from the video, displays images from the intervention video, and has hollow stars next to each week that respondents were encouraged to color in if they talked to their infant at least once a day during that week.<sup>7</sup>

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

We implemented the intervention at the public health clinics after the patients’ prenatal or postnatal visits, which mirrors how we expect that this intervention would be implemented at scale.<sup>8</sup> 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 if the video is shown to a group of patients, perhaps on a television monitor in the waiting room.<sup>9</sup>

### *B. Sample Characteristics and Baseline Behavior*

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

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

As expected, and consistent with the qualitative background research that led us to conceive this study, baseline knowledge about the role of verbal engagement in early childhood development is limited. Table 2 presents baseline IDS beliefs and behavior for our sample. On average, respondents report that parents should start talking

<sup>7</sup>The calendar also included a link to the web page with the video. There were 26 longer-than-three-second viewers of the Dagbani version of the video during the study period and 10 longer-than-60-second viewers. The low usage of the web page could be due to the internet data charges to stream a video being expensive for this population.

<sup>8</sup>We partnered with officials in the Ghana Health Service who agreed that this was a reasonable expectation.

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

<sup>10</sup>We also cannot reject the null of the joint test for the baseline variables on IDS-related beliefs and behaviors presented in Table 2, providing additional evidence that the randomization was implemented correctly.



TABLE 1—BASELINE CHARACTERISTICS AND BALANCE

	Full sample			Treatment		Control		Treatment = control
	Mean	SD	Obs.	Mean	SD	Mean	SD	<i>p</i> -value
Age (years)	27.75	5.17	1,403	27.93	5.15	27.57	5.19	0.194
Dagomba ethnicity	0.82	0.38	1,407	0.83	0.38	0.82	0.38	0.825
Main language spoken: Dagbani	0.88	0.33	1,408	0.88	0.33	0.88	0.33	0.951
Highest level of education:								
None	0.37	0.48	1,406	0.38	0.49	0.36	0.48	0.500
Primary school	0.28	0.45	1,406	0.28	0.45	0.28	0.45	0.879
Secondary school	0.22	0.42	1,406	0.21	0.41	0.23	0.42	0.463
Can read (English/Dagbani)	0.61	0.49	1,408	0.59	0.49	0.62	0.48	0.205
Housewife/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
No. household members: Under 5	1.90	1.60	1,407	1.92	1.66	1.87	1.54	0.537
No. household members: 5–15 y/o	1.88	2.08	1,405	1.96	2.08	1.80	2.08	0.142
No. household members: Above 16	4.85	3.22	1,400	4.83	3.23	4.86	3.21	0.877
Has children	0.89	0.31	1,408	0.90	0.29	0.88	0.32	0.138
Has child 6 y/o or younger	0.75	0.43	1,408	0.77	0.42	0.74	0.44	0.229
Has child older than 1 month	0.69	0.46	1,408	0.70	0.46	0.68	0.47	0.284
Has child older than 3 months	0.64	0.48	1,408	0.64	0.48	0.63	0.48	0.516
Age at first child (years)	22.23	3.50	1,242	22.14	3.37	22.33	3.62	0.327
No. of children	2.21	1.55	1,408	2.28	1.54	2.15	1.57	0.105
Age of youngest child (months)	15.31	20.74	1,182	15.12	20.12	15.50	21.36	0.754
Youngest child eligible	0.61	0.49	1,408	0.62	0.49	0.60	0.49	0.389
Pregnant with an eligible child	0.39	0.49	1,408	0.38	0.49	0.40	0.49	0.360
Focal child is firstborn	0.28	0.45	1,408	0.26	0.44	0.30	0.46	0.129
<i>F</i> -test <i>p</i> -value								0.662
Observations		1,408			705		703	

*Notes:* 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 and hence is missing for some observations. The *F*-test *p*-value reported at the bottom of the table is for the joint significance of the differences between the treatment and control groups for all of the variables reported in the table. For the *F*-test, missing values (due to refusal/don't know or a logic skip (e.g., "age of youngest child" when there are no children)) are replaced by the variable average value and flagged by a dummy.

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 that children start saying meaningful words themselves (20 months). These reports demonstrate that the beliefs of many women in our sample diverge from evidence-supported practices for enhancing child development, such as extensively conversing with newborn infants.

### C. Endline Measurements

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

TABLE 2—BASELINE IDS BELIEFS AND BEHAVIOR

	Full sample			Treatment		Control		Treatment = control
	Mean	SD	N	Mean	SD	Mean	SD	p-value
<i>Beliefs on IDS and child development</i>								
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 months) 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 (months) 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
<i>Self-reported IDS behavior</i>								
Tells stories to youngest child	0.51	0.50	1,059	0.50	0.50	0.52	0.50	0.366
Asks youngest child to repeat words	0.61	0.49	1,059	0.60	0.49	0.61	0.49	0.764
When child was 1 m/o: Described objects when cleaning/organizing	0.40	0.49	972	0.36	0.48	0.43	0.50	0.036
When child was 3 m/o: Described things to child when walking	0.64	0.48	895	0.65	0.48	0.63	0.48	0.594
<i>Inequality aversion</i>								
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 first child if she provides better care to second child	0.69	0.46	1,408	0.68	0.47	0.69	0.46	0.758
F-test p-value								0.765
Observations		1,408			705		703	

*Notes:* Baseline data. Treatment is a dummy equal to 1 if the respondent received the main intervention (at baseline). Child's age outcomes are in months. In the panel "Beliefs on IDS and child development," the 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. "Tells stories to youngest child" and "Asks youngest child to repeat words" were only asked to respondents with a child aged six years or less, and the two subsequent questions to those with a child aged between one month and six years and between three months and six years. The *F*-test *p*-value reported at the bottom of the table is for the joint significance of the differences between the treatment and control groups for all the variables reported in the table. For the *F*-test, missing values (due to refusal/"don't know" or a logic skip (e.g., "age of youngest child" when there are no children)) are replaced by the variable average value and flagged by a dummy.

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

We describe the measurement tools used and how we combine them for analysis in Section IIF. All survey outcomes were collected in the main endline visit except for those on treatment recall and perceived barriers to IDS, which were measured at



the very end of all endline-related activities (i.e., post-LENA measurements when applicable) for a given respondent.

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

The LENA surveyors visited respondents before 10 AM on the day of the recording activity to give mothers the shirt with the LENA device, they answered any questions about the device and/or instructions, and they stayed while mothers dressed the child with the shirt. Surveyors asked mothers to have the child wear the shirt until the next day. Ninety-seven percent of mothers who completed the endline survey consented after the surveyor described the LENA process. We restrict the LENA analysis to audio data collected from 10 AM to 7 PM with no interruption (nine hours of recording).<sup>12</sup> The same surveyor came back the next day to pick up the LENA device and conduct a short survey on the respondent's experience with the LENA, barriers to conversing with babies, and (for treatment respondents) recall of the main intervention. Given the cost of the LENA devices and the LENA pickup surveyors, we could only afford to use the LENA measurements with 900 respondents. We randomly chose 900 respondents from our full sample, stratified by treatment status.<sup>13</sup> We obtained 785 LENA recordings (see Supplemental Appendix Table A.2 for details on survey and LENA activity participation rates).

#### D. Endline Intervention

Had we found null effects at endline, it would have been important to understand whether participants never adopted the recommended practices or adopted them initially but then stopped. It is also possible that the treatment effects could have grown over time as participants gained experience and comfort with IDS. Thus, to assess

<sup>11</sup> This device was validated in Ghana by the Harvard Laboratory for Developmental Studies. See Supplemental Appendix Section B for more information.

<sup>12</sup> This nine-hour time window ensures that we have comparable data for all observations, as households received the LENA device between 6 and 10 AM depending on when the LENA surveyor arrived. The LENA device could record 16 hours of audio, but after 7 PM, a few LENA devices turned off (because either they ran out of battery or households turned them off (purposely or not) to bathe the child). Ninety-nine percent of recordings had 9 uninterrupted hours of audio and were kept in the analysis.

<sup>13</sup> We originally sampled 900 respondents but discovered at endline that 1 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 (observations = 450 from the treatment group, observations = 449 from the control group).

how the immediate effect of the intervention differs from the six-to-eight-month effect, we also compare the child's auditory environment *the day before* their mother watches the video to the *the day after* through our "endline intervention."

We use the LENA to measure the effect of the endline intervention. Relative to our other options (self-reports or direct observation by the surveyor), the LENA should engender less of an experimenter demand effect. Moreover, a positive treatment effect on IDS that is driven by experimenter demand still enables us to rule out the existence of a "technological" barrier or contextual factor that prevents mothers from talking to their infant even when they would like to do so.

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

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

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 point in the calendar year. This prevents a confound such as the endline measurement occurring

<sup>14</sup>We stratified by whether the focal child would be under or over one month old at baseline and, for households with a focal child over one month old at baseline, whether they scored above or below the median on the baseline self-reported mother behavior index. Of the 225 selected households, 24 were ineligible or not available at endline. Six 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 Supplemental Appendix Table A.2 for further details).

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

<sup>16</sup>Of the 225 endline intervention participants, 112 were also sampled to watch, just prior to watching the intervention video, a one-minute video of a Ghanaian mother verbally engaging with their infant. 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.

<sup>17</sup>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.

during the lean season and treatment effects being smaller in the lean season, which could cause treatment effects to only be observed in the short run. With the standard approach, environment-contingent treatment effects like this could be misinterpreted as fade-out. Second, the approach reduces study costs if outcome measurement has a fixed-cost component (e.g., to train a team of surveyors on how to deliver the LENA device to study participants). Since the short- and long-run measurements occur simultaneously, fixed costs are incurred only once. There are also drawbacks, such as the short-run effects being estimated from children who are older than those who identify the long-run effects.<sup>18</sup>

### *E. Treatment Recall, Social Norms, and Perceived Barriers*

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

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

In our final set of endline questions for control and treatment respondents, we asked their opinions on the barriers to parent-infant conversations for families in

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

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

TABLE 3—TREATMENT RECALL AND SELF-REPORTED BEHAVIOR CHANGE

	Mean	SD	Count	N
<i>Main intervention sample (treatment)</i>				
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			
<i>Endline intervention sample</i>				
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			

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

their community (see Supplemental Appendix Table A.3). Among respondents who did not watch the video, 43 percent 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 percent), “it’s mocked/frowned upon in the community” (32 percent), and “it’s not clear that it makes any difference for the child” (28 percent).

### F. Outcome Measures

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

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

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

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

Specifically, for language, mothers were asked about specific words and phrases their child understands and/or attempts to pronounce (see Supplemental Appendix

<sup>20</sup> We do not impute missing index components when calculating the components’ weights and the indices. Except for the observed child development index, missing index components are “Refuse to answer” and “Don’t know.” For mother-reported indices, between 95 and 100 percent of respondents answered all components. For the observed child development index, missing components are due to the surveyor selecting the answer “Unable to assess” (because the infant became agitated, refused to participate, etc.). We drop observations missing more than 50 percent of components (observations = 19).

Table A.6). Mothers were also asked three questions about whether the child has started to talk. We do not use an Anderson index to compute the child language score; instead, we compute a score using item response theory (IRT), which involves estimating a one-parameter logistic model on the mother's responses to these questions, where the model assigns a difficulty level to each question and, then, a latent trait to each individual based on their answers to the questions adjusting for the question's difficulty level.<sup>21</sup>

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

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

*Surveyor-Measured Child Development.*—We collect a direct observation (by surveyors) of child development by adapting questions from the ASQ and the Ox-NDA, shown in Supplemental Appendix Table A.8.<sup>22</sup> Specifically, we use two ASQ-like items adapted to be administered by surveyors (as opposed to collecting mothers' reports): whether the child's eyes follow the mother when she moves and whether the child's eyes follow a toy placed in front of her.

Most items in the original Ox-NDA ask the surveyor to take an action (e.g., placing a spoon, cup, plate, ball, and pen in front of the child and asking the child "which one is the spoon?") and to record observations about the child's response (e.g., whether they pointed to/picked up the spoon, pointed to/picked up a different object, or did not respond). While the Ox-NDA is designed for children 10–14 months old, 57 percent of our focal children were 9 months or younger at endline. In July 2022, we piloted the items after adapting them to the local context, and identified the most promising seven items (in terms of expected variation) to include in our measure.

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

When prompted to do some of the Ox-NDA items, some children disengaged from the test or refused to perform tasks they had performed at other instances in the survey, such as babbling. Two weeks into the data collection, we added a question on whether the surveyor observed the child babbling (at least one syllable) *at*

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

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



*some point* during the home visit. As a result, we have two observed measures of babbling: one as part of the Ox-NDA test and one for whether the child babbled at any point during the survey. As the latter measure was added to the survey two weeks after starting data collection, it is only available for 71 percent (888/1,258) of respondents who completed the endline survey.

*LENA Outcomes.*—We focus on two LENA measures: female adult words per minute (a measure of parental verbal input) and child vocalization count per minute (a measure of child verbal output).

To measure the impact of the endline intervention, we use the second daylong LENA recording of children's auditory environments (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).

### G. Preregistration

We registered the RCT first in the AEA RCT Registry (ID AEARCTR-0007161) and subsequently with more details in ClinicalTrials.gov (ID NCT04807907) (Dupas, Jayachandran, and Walsh 2021a, b).

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

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

<sup>23</sup> Assuming a LENA sample size of 900 with 9 percent attrition and a 100 percent take-up rate.

<sup>24</sup> Conversational turn count is the number of alternations between the child and adults in the vicinity, so it combines parental input and child output.

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

### H. Descriptive Analyses

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

## III. Empirical Framework

### A. Treatment Effects of Main Intervention

We identify the impact of the video-plus-calendar intervention on our outcomes of interest at endline (six to eight months after the intervention) by estimating the following equation via ordinary least squares (OLS):

$$(1) \quad Y_i = \beta_0 + \beta_1 T_i + \beta_2 \mathbf{X}_i + \eta_i + \epsilon_i,$$

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.  $\mathbf{X}_i$  is a vector of controls including the child's age in days, the

<sup>25</sup> Child vocalizations are also weakly positively correlated with surveyor-observed and mother-reported child language development.

<sup>26</sup> Parental beliefs about the importance of IDS or self-reported behavior are also not positively correlated with maternal education or SES.

date of the survey, and an indicator for the surveyor being female. For outcomes derived from the LENA recording, we control for the child's age in days, household size, the day of the week the audio was recorded, and interruptions to the LENA's recordings.<sup>27</sup>  $\eta_i$  represents clinic fixed effects.  $\epsilon_i$  is the error term, and the estimated standard errors are robust to heteroskedasticity. We adjust for multiple hypothesis testing among our primary outcomes using Romano and Wolf (2016).<sup>28</sup>

The experimental variation in  $T_i$  generated by the RCT enables us to estimate the causal effect of the intervention on our outcomes of interest as long as the stable unit treatment value assumption (SUTVA) holds. Our estimate of this effect will be captured by  $\beta_1$ , so we will be interested in testing whether  $\beta_1$  is significantly different from zero.

One threat to our identifying assumption, SUTVA, is that control respondents may learn about the intervention from treatment respondents. These spillovers would likely downward bias our estimates by improving the outcomes of the control respondents. To explore the magnitude of spillovers, we gathered rough measures of the extent of social diffusion of our intervention. At baseline, we asked treated respondents after they saw the video "Do you know anyone who has already seen this video?" (Enrollment in the study was on a rolling basis over three weeks.) Seven percent reported knowing someone who had seen the intervention video. At endline, 8 percent of 195 control respondents who received the endline intervention and were asked the same question reported knowing someone who had seen the video, and 16 percent of 615 treatment respondents (who had received the "main intervention" at baseline) had discussed the video with friends (see Table 3); a subset of these friends could be in the control group. We did not probe control group respondents further on what they had heard about the video prior to their baseline survey, nor on what they discussed with others postbaseline. Hence, we are unable to provide further details on how much that control respondents learned about the video or its key message through social spillovers. Such spillovers would lead us to underestimate the treatment effect, but the magnitude of this underestimation is likely very small.

### B. Before/After Immediate Effect of the Endline Intervention

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

$$(2) \quad Y_{it} = \gamma_0 + \gamma_1 \text{EndlineIntervention}_t + \gamma_2 \mathbf{Z}_t + \omega_i + \mu_{it},$$

where  $t$  denotes whether this was the first or second daylong LENA recording for a given focal child, and  $\text{EndlineIntervention}_t = 1$  if  $t = 2$  (i.e., for the observation collected after the respondent received the endline intervention) and 0 otherwise.  $\mathbf{Z}_t$

<sup>27</sup> Interruptions include the device being removed or the child being on someone's back where the sound might be muffled.

<sup>28</sup> We assess the robustness of our results to alternative specifications in Supplemental Appendix Table A.9.

represents a vector of LENA-recording-specific characteristics such as the day of the week that the audio was recorded and interruptions to the LENA's recordings.  $\omega_i$  represents household fixed effects. The estimated standard errors are robust to heteroskedasticity.

The use of household fixed effects means that our coefficient of interest,  $\gamma_1$ , identifies the effect from within-household (i.e., within-focal-child) changes over time in treatment status. More precisely,  $\gamma_1$  estimates the change in the child's verbal inputs or outcomes between the *day before* the endline intervention and the *day after* it (i.e., the day after their mother watches the video). We interpret  $\gamma_1$  as the immediate effect of their mother watching the video on the child's verbal environment.<sup>29</sup>

## IV. Results and Discussion

### A. Effects of the Intervention after Six to Eight Months

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

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

The magnitude of these effects are lower than heavier-touch interventions but encouraging given the light-touch nature of the intervention. Relative to the pooled effect sizes computed by Jeong et al. (2021) in their meta-analysis of around two dozen randomized interventions with children under 12 months at baseline, our treatment effects on parental beliefs are 26.3 percent of the average effect of heavier-touch interventions; for parental behaviors and infant language development, our effects are 42.8 percent and 46.4 percent of the average effects, respectively.

Using indices has the disadvantage that the outcomes are less concrete, so we also present the effects for a few concrete outcomes.<sup>30</sup> Treated mothers are 10.4 percentage

<sup>29</sup>The mother knows that she is being recorded, and we cannot speak to whether the results generalize to when she is not being recorded. Another potential confound is that the posttreatment measure is always the second day of recording, while the pretreatment measure is always the first day of recording. Past studies using the LENA did not report that LENA outcomes differed substantively between the first and second usages.

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

TABLE 4—TREATMENT EFFECTS, 6 TO 8 MONTHS AFTER INTERVENTION

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

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

points more likely to report that parents should start talking to their infant at birth, a key message of the video and calendar ( $p < 0.001$ ) (see Supplemental Appendix Table A.4). The effect size is 31.5 percent of the control group mean. Treated mothers are also 6.5 percentage points ( $p = 0.019$ , 14.8 percent of the control mean) more likely to report that an adult read to or looked at a book with the child and 4.5 percentage points ( $p = 0.089$ , 14.5 percent of the control mean) more likely to report that an adult told stories to the child in the last 4 weeks (see Supplemental Appendix Table A.5). Based on their reports, their children understand 0.48 more words ( $p = 0.008$ , 6.8 percent of the control mean) and can say 0.31 more words ( $p = 0.014$ , 29.3 percent of the control mean) from the 16 words listed by the surveyor (see Supplemental Appendix Table A.6 for the list of words and treatment effects on easier-to-interpret/more concrete outcomes).

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

There is a significant positive effect on whether the child babbled at any point during the survey. As mentioned earlier, we added this question after two weeks of data collection. Among the 71 percent of respondents for whom the measure was recorded, the intervention led to a 5.4 percentage point increase ( $p = 0.054$ ) in the infant babbling, which is 10.6 percent of the control mean (Table 4, column 6).

*Heterogeneity by Child Age.*—We also estimate the treatment effects on the mother-reported and surveyor-observed outcomes separately by age group, reported in Table 5.<sup>31</sup> Heterogeneity by age is interesting per se, but we show this breakdown mainly because the various instruments we use are more appropriate for certain age groups. In particular, the child development measures are not designed for very young infants. For mother-reported outcomes, the positive pooled effects reported above are driven by children aged six months and older. This is reassuring since these are the ages when, a priori, the measures should be more reliable. The treatment effect on surveyor-observed child development, which is statistically insignificant in the full sample, shows no clear age gradient (see Supplemental Appendix Figure A.7 for effects by child's age in months). The treatment effect on the probability of babbling is concentrated among young infants. Among young infants, the intervention doubles the proportion who babbled, from 10 percent to 20 percent ( $p = 0.049$ ; Table 5, column 4).

*LENA-Recorded Outcomes.*—The treatment effects on our primary LENA-recorded outcomes are positive but insignificant. We estimate an insignificant 1 percent increase in female adult words per minute and a 3 percent increase in child vocalizations per minute (0.131 words with  $p = 0.851$  and 0.052 with  $p = 0.377$ , respectively; Table 4, columns 7–8).<sup>32</sup>

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

<sup>31</sup> See Supplemental Appendix Figure A.6 for full distribution of infant age at endline.

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



TABLE 5—TREATMENT EFFECTS BY AGE GROUPS, 6 TO 8 MONTHS AFTER INTERVENTION

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

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

The negative correlations between background noise and child vocalizations, conversational turns, female adult word count, male adult word count, and other child speech in the control group support this interpretation.

One limitation of our LENA results is that the estimates are noisy since our sample size is half as large as for the self-reported outcomes due to budget constraints. As anticipated in our power calculations (see Section IIG), the treatment effects on LENA measures have larger standard errors (about 0.07 when using standardized outcomes) than the treatment effects on the other outcomes. This could contribute to the insignificant effects on the LENA outcomes, while we observe significant and

substantial effects on the mother-reported outcomes. Future research could attempt to survey more households or record households for multiple days to see if the reported maternal behavior changes translate into true small, hard-to-detect changes.

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

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

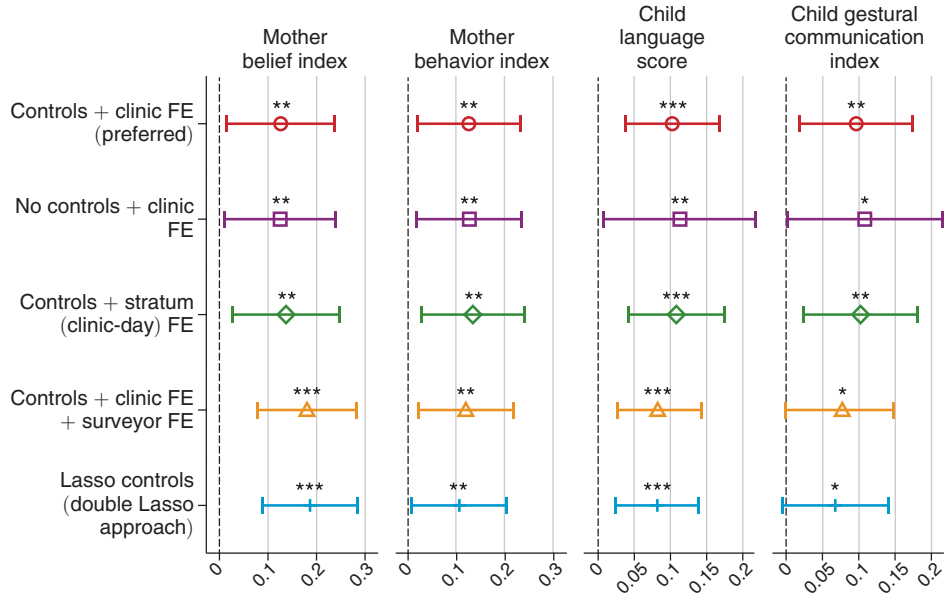
For the belief index, we find results consistent with experimenter demand effects. Those who associate the intervention with the survey organization have a significantly larger treatment effect. This pattern is not mirrored for any of the other mother-reported outcomes. The differential treatment effects among those who associate the survey with the intervention are close to 0 and statistically insignificant. Experimenter demand could also cause LENA inputs to be higher (the mother uses IDS more when her behavior is being recorded), but, reassuringly, we find no evidence for this.

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

*Robustness Checks.*—To test the robustness of our results, we estimate the treatment effects using alternative specifications: excluding control variables other than clinic fixed effects, including clinic-day (stratum) fixed effects instead of just clinic fixed effects, adding surveyor fixed effects, and using double Lasso to select control variables. We describe the tests in more details below, but Figure 1 graphically summarizes them by showing the treatment coefficient and 95 percent confidence intervals from each robustness specification for our eight main outcomes (see Supplemental Appendix Tables A.9 and A.12 for regression tables.). In brief, across the different robustness checks, the results remain significant at the 5 percent level for mother-reported mother's beliefs, parental verbal inputs, and child language

<sup>33</sup> Since we are conditioning on an endogenous variable, differences in baseline characteristics may also be driving differences in endline outcomes.

Panel A. Mother-reported measures



Panel B. Observed and LENA measures

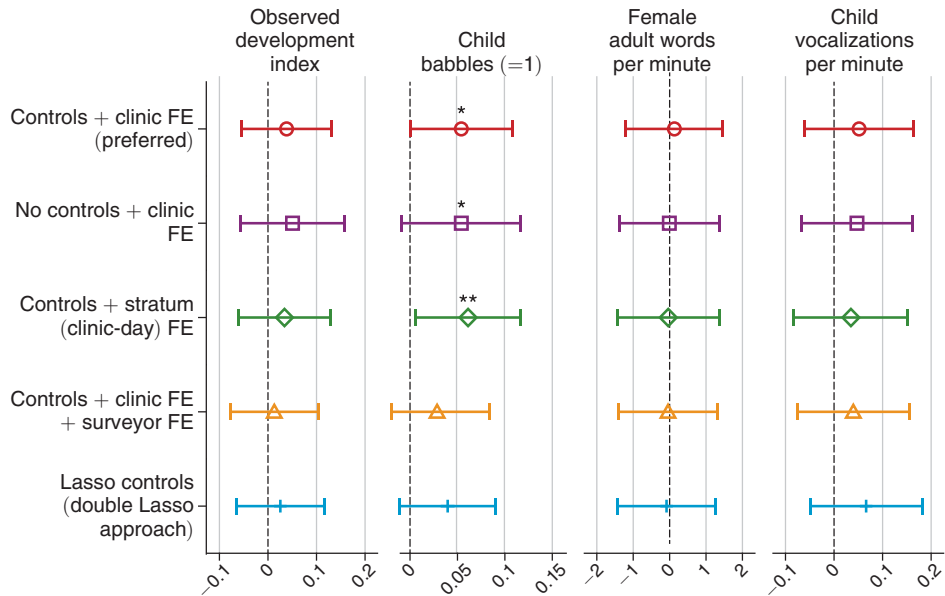


FIGURE 1. ROBUSTNESS ANALYSIS

Notes: Endline and LENA day 1 recording data. The figure plots the treatment coefficients and 95 percent confidence intervals for regressions of the outcome variable listed at the top on a treatment dummy and the fixed effects (FEs) and controls indicated on the y-axis. The first specification is the one used to estimate treatment effects in Table 4. The second to fourth rows change the main specifications by removing control variables other than clinic FEs, replacing clinic FEs by clinic-day FEs, or adding surveyor FEs (also reported in Supplemental Appendix Table A.9). The estimate in the fifth row uses the double Lasso approach of Belloni, Chernozhukov, and Hansen (2013) as implemented by Ahrens, Hansen, and Schaffer (2019) to choose control variables (also reported in Supplemental Appendix Table A.12). All specifications use robust standard errors. For further details on outcomes, please see Table 4.

scores and at the 10 percent level for the child gestural communication score. They remain positive but insignificant for child development and LENA outcomes.<sup>34</sup> The significance of the effect on observed child babbling outcome is not robust to the inclusion of surveyor fixed effects or Lasso-selected controls.

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

Surveyors might be idiosyncratic in how they interpret and record mothers' responses or assess the child's behavior, so surveyor fixed effects could be correlated with outcome measures, in which case including them could improve precision. However, every surveyor surveyed both treatment and control respondents, and, by design, surveyor fixed effects should not be systematically correlated with treatment. This is indeed the case in the data: Surveyor fixed effects are significantly correlated with all of the outcome measures except the LENA outcomes (reassuringly), but they are not correlated with treatment status (see Supplemental Appendix Table A.13). When we include surveyor fixed effects, there are nonetheless some changes in the treatment coefficients. The effect on observed child babbling remains positive but is no longer statistically significant (3.2 percentage points,  $p = 0.232$ ; see Supplemental Appendix Table A.9, column 6). The other results that are significant in the main specification remain significant with surveyor fixed effects.

We use the double Lasso approach of Belloni, Chernozhukov, and Hansen (2013) as implemented by Ahrens, Hansen, and Schaffer (2019) to flexibly choose control variables, separately for each outcome. Overall, we see little change in the coefficients and small reductions in standard errors (see Supplemental Appendix Table A.12). Significance levels do not change except for the effect on infant babbling, where the  $p$ -value increases from 0.054 to 0.135, and on child gestural communication index, where the  $p$ -value increases from 0.018 to 0.083.

We also compute Romano-Wolf stepdown adjusted  $p$ -values to adjust for multiple hypothesis testing and report them in Table 4 (Romano and Wolf 2016). Focusing on the results with a  $p$ -value below 0.1 using the conventional  $t$ -test, the effect on reported child language score remains significant (stepdown  $p = 0.023$ ), while the stepdown  $p$ -values are above 0.1 for reported maternal beliefs ( $p = 0.160$ ), parental behavior ( $p = 0.158$ ), child gestural communication ( $p = 0.127$ ), and observed child babbling ( $p = 0.207$ ).

### B. Effects on Newly Informed Mothers

We next present the effects of the endline intervention, or the immediate effects of the mother watching the video on the child's verbal inputs and outputs. The treatment effects (estimated using equation (2)) are reported in Table 6. On average, the child hears 1.83 more adult words ( $p = 0.036$ ) the day after the endline intervention compared to the day before. In our context, women speak far more words to

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

TABLE 6—TREATMENT EFFECTS ON NEWLY INFORMED MOTHERS:  
EVIDENCE FROM THE “ENDLINE” INTERVENTION

	Adult words per minute (1)	Female adult words per minute (2)	Male adult words per minute (3)	Conversational turn count per minute (4)	Percent meaningful speech (5)
Second day (postintervention)	1.827 (0.864) {0.036}	1.389 (0.728) {0.058}	0.509 (0.285) {0.076}	0.015 (0.016) {0.352}	0.927 (0.487) {0.058}
Mean preintervention (day 1)	20.12	16.46	3.67	0.42	16.65
Mean postintervention (day 2)	21.81	17.67	4.14	0.44	17.62
Controls	Yes	Yes	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	372	371	371	372	372

Notes: 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 second day of recording, households were shown the intervention video. Of the 225 households sampled for a second day of recording, 192 consented to the recording and saw the intervention video. The analysis is restricted to recording times between 10 AM to 7 PM. Of 192 households, 186 had 2 complete audio recordings. Adult words per minute by gender is not available for one recording. Regressions include controls for the day of the week that the audio was recorded (dummies), the total time (in minutes) that the shirt/LENA device was removed from the child, and the total time (in minutes) that the child was held on someone’s back while wearing the device. Household fixed effects are included. Conversational turn count is the number of alternations between the focal child and adults in the vicinity. Percent meaningful speech is the share of the audio categorized as sounds from the focal child or speech from adults or other children near the focal child. For further details on the LENA outcomes, please refer to Supplemental Appendix Section A2. Robust standard errors in parenthesis; *p*-values in curly brackets.

infants than men do, so the effect is primarily driven by a rise in female adult words (accounts for 76 percent of the effect). However, there is also an impact on male adult words (0.51 words, *p* = 0.076), suggesting spillovers of the intervention to other members of the household who did not view the video. As expected, we do not see increases in measures of child verbal output such as child vocalizations per minute; one would expect these gains to only materialize in the longer run as child language skills accrue from increased IDS. We see a modest but insignificant increase in “conversational turns” per minute (0.015 turns, *p* = 0.352), which requires engagement between adults’ verbal inputs and the focal child’s verbal output.

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

C. Self-Reported Barriers to IDS

When we asked respondents about potential barriers preventing families from talking to their babies, the most common answers for those who never watched the video were “it’s hard to remember/make a habit” (35 percent) or “it’s mocked/frowned upon in the community” (32 percent) (see Supplemental Appendix Table A.3). Comparing

this group to those who received the endline intervention (i.e., the newly informed mothers), we find that the endline intervention increased reporting of mocking/social scorn as the main barrier by 19.7 percentage points ( $p < 0.001$ ; see Supplemental Appendix Table A.14, column 3). However, such a reaction appears short lived. Mothers who received the treatment six to eight months earlier (the treatment group who received the main intervention) are much less likely to report social scorn as the main barrier ( $p = 0.020$ ). In contrast, habit formation is equally likely to be cited as a main barrier among the main treatment group and the endline intervention group ( $p = 0.967$ ).

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

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

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

## V. Cost-Effectiveness

As a measure of cost-effectiveness, we estimate the cost per child’s SD improvement in cognitive and language outcomes. We assume that the at-scale costs of the intervention would only include printing out the calendars and delivering them to health clinics. As our measure of the benefits of the intervention, we average the effects on child language and gestural communication (from Table 4).<sup>37</sup> Using these

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

<sup>36</sup> We added questions on calendar use midsurvey, so we do not have this data for 90 treatment respondents.

<sup>37</sup> Cost-effectiveness is similar when we use Cohen’s  $d$  as the measure of the intervention effect, following Verguet et al. (2022), instead of our regression coefficients.



TABLE 7—COST-EFFECTIVENESS CALCULATIONS

		At-scale costs		RCT costs	
	SD effects	Unit cost (\$)	Cost/SD (\$)	Unit cost (\$)	Cost/SD (\$)
<i>Panel A. Full intervention effects size</i>					
Mother-reported measures	0.10	0.45	4.52	3.01	30.26
Observed measures	0.04	0.45	10.01	3.01	66.93
<i>Panel B. One-half intervention effects size</i>					
Mother-reported measures	0.05	0.45	9.05	3.01	60.52
Observed measures	0.02	0.45	20.01	3.01	133.85
<i>Panel C. One-fourth intervention effects size</i>					
Mother-reported measures	0.02	0.45	18.09	3.01	121.03
Observed measures	0.01	0.45	40.02	3.01	267.71

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

estimates, we calculate that the intervention delivers a 1 SD improvement in child development for \$4.52 (see Table 7). This would be lower than any of the interventions included in the meta-analysis of “responsive caregiving” intervention by Verguet et al. (2022) (the majority of which are home-visiting interventions).<sup>38</sup>

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

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

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

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

<sup>39</sup> After adjusting for inflation, the most cost-effective intervention in Verguet et al. (2022) delivers a 1 SD improvement for \$35.96.

truly have positive effects. This possibility of null effects is counterbalanced by high cost-effectiveness if the intervention does have even moderate positive effects. Because of the upside potential, there is a case for scaling up the intervention based on what we have learned. Future research using more precise observed measures or larger sample sizes could remove the uncertainty around cost-effectiveness, enabling even risk-averse policymakers to scale up the policy. Furthermore, combining the intervention with support for mothers in forming a parent-infant conversation habit could be even more cost-effective than the intervention alone, given our finding that habit formation is the main barrier to mothers sustaining their short-run levels of verbal input.

## VI. Conclusion

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

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

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

## REFERENCES

- Ahrens, Achim, Christian B. Hansen, and Mark E. Schaffer. 2019. “PDSLASSO: Stata Module for Post-selection and Post-regularization OLS or IV Estimation and Inference.” Statistical Software Components S458459.
- Anderson, Michael L. 2008. “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* 103 (484): 1481–95.
- Antwi, K. D., and C. P. Lyford. 2021. “Socioeconomic Determinants of Rural Households’ Food Security Status in Northern Ghana.” *Journal of Agriculture and Food Sciences* 19 (2): 86–94.
- Attanasio, Orazio P. 2015. “The Determinants of Human Capital Formation during the Early Years of Life: Theory, Measurement, and Policies.” *Journal of the European Economic Association* 13 (6): 949–97.

- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen.** 2013. "Inference on Treatment Effects after Selection among High-Dimensional Controls." *Review of Economic Studies* 81 (2): 608–50.
- Black, Maureen M., Susan P. Walker, Lia C. H. Fernald, Christopher T. Andersen, Ann M. DiGirolamo, Chunling Lu, Dana C. McCoy, et al.** 2017. "Early Childhood Development Coming of Age: Science through the Life Course." *Lancet* 389 (10064): 77–90.
- Duflo, Esther, Pascaline Dupas, Elizabeth Spelke, and Mark Walsh.** 2024. "The Intergenerational Effects of Secondary Education: Experimental Evidence from Ghana." NBER Working Paper 32742.
- Dupas, Pascaline, Seema Jayachandran, and Mark Walsh.** 2021a. *Promoting Infant-Directed Speech in Ghana*. ClinicalTrials.gov, Pre-analysis plan. <https://clinicaltrials.gov/study/NCT04807907>.
- Dupas, Pascaline, Seema Jayachandran, and Mark Walsh.** 2021b. *Promoting Infant-Directed Speech in Northern Ghana*. AEA RCT Registry. <https://doi.org/10.1257/rct.7161-1.0>.
- Dupas, Pascaline, Camille Falezan, Seema Jayachandran, and Mark Walsh.** 2025. *Data and Code for: "Informing Mothers about the Benefits of Conversing with Infants: Experimental Evidence from Ghana."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. <https://doi.org/10.3886/E206742V1>.
- Ertem, I. O., G. Atay, D. G. Dogan, A. Bayhan, B. E. Bingoler, C. G. Gok, S. Ozbaz, D. Haznedaroglu, and S. Isikli.** 2007. "Mothers' Knowledge of Young Child Development in a Developing Country." *Child: Care, Health and Development* 33 (6): 728–37.
- Farran, Lama K., Chia-Cheng Lee, Hyunjoo Yoo, and D. Kimbrough Oller.** 2016. "Cross-Cultural Register Differences in Infant-Directed Speech: An Initial Study." *PLoS ONE* 11 (3): e0151518.
- Fernandes, Michelle.** 2021. "The Oxford Neurodevelopment Assessment (OX-NDA) for Infants Aged 10 to 14 Months." *THE OX-NDA MANUAL – V2.1*.
- Gilkerson, Jill, and Jeffrey A. Richards.** 2008. *The LENA Natural Language Study*. LENA Foundation.
- Gilkerson, Jill, and Jeffrey A. Richards.** 2020. *A Guide to Understanding the Design and Purpose of the LENAr System*. LENA Foundation.
- Hart, Betty, and Todd R. Risley.** 1995. *Meaningful Differences in the Everyday Experience of Young American Children*. Paul H. Brookes Publishing.
- Hoff, Erika.** 2003. "The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development via Maternal Speech." *Child Development* 74 (5): 1368–78.
- Jeong, Joshua, Helen O. Pitchik, and Aisha K. Yousafzai.** 2018. "Stimulation Interventions and Parenting in Low- and Middle-Income Countries: A Meta-Analysis." *Pediatrics* 141 (4): e20173510.
- Jeong, Joshua, Emily E. Franchett, Clariana V. Ramos de Oliveira, Karima Rehmani, and Aisha K. Yousafzai.** 2021. "Parenting Interventions to Promote Early Child Development in the First Three Years of Life: A Global Systematic Review and Meta-analysis." *PLoS Medicine* 18 (5): e1003602.
- Lally, Phillippa, and Benjamin Gardner.** 2013. "Promoting Habit Formation." *Health Psychology Review* 7: S137–S158.
- List, John A., Julie Pernaudet, and Dana Suskind.** 2021. "It All Starts with Beliefs: Addressing the Roots of Educational Inequities by Shifting Parental Beliefs." NBER Working Paper 29394.
- Marchman, Virginia A., and Philip S. Dale.** 2023. "The MacArthur-Bates Communicative Development Inventories: Updates from the CDI Advisory Board." *Frontiers in Psychology* 14: 1170303.
- Monnot, Marilee.** 1999. "Function of Infant-Directed Speech." *Human Nature* 10 (4): 415–43.
- Romano, Joseph P., and Michael Wolf.** 2016. "Efficient Computation of Adjusted  $p$ -values for Resampling-Based Stepdown Multiple Testing." *Statistics & Probability Letters* 113: 38–40.
- Rothman, Alexander J., Peter M. Gollwitzer, Adam M. Grant, David T. Neal, Paschal Sheeran, and Wendy Wood.** 2015. "Hale and Hearty Policies: How Psychological Science Can Create and Maintain Healthy Habits." *Perspectives on Psychological Science* 10 (6): 701–5.
- Shrestha, Merina, Manjeswori Ulak, Tor A. Strand, Ingrid Kvestad, and Mari Hysing.** 2019. "How Much Do Nepalese Mothers Know about Child Development?" *Early Child Development and Care* 189 (1): 135–42.
- Suskind, Dana L., Christy Y. Y. Leung, Robert J. Webber, Alison C. Hundertmark, Kristin R. Leffel, Iara E. Fuenmayor Rivas, and William A. Grobman.** 2018. "Educating Parents about Infant Language Development: A Randomized Controlled Trial." *Clinical Pediatrics* 57 (8): 945–53.
- Suskind, Dana L., 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.** 2016. "A Parent-Directed Language Intervention for Children of Low Socioeconomic Status: A Randomized Controlled Pilot Study." *Journal of Child Language* 43 (2): 366–406.

- Verguet, Stéphane, Sarah Bolongaita, Anthony Morgan, Nandita Perumal, Christopher R. Sudfeld, Aisha K. Yousafzai, and Günther Fink.** 2022. "Priority Setting in Early Childhood Development: An Analytical Framework for Economic Evaluation of Interventions." *BMJ Global Health* 7 (6): e008926.
- Webb, Thomas L., and Paschal Sheeran.** 2006. "Does Changing Behavioral Intentions Engender Behavior Change? A Meta-analysis of the Experimental Evidence." *Psychological Bulletin* 132 (2): 249–68.
- Weisleder, Adriana, and Anne Fernald.** 2013. "Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary." *Psychological Science* 24 (11): 2143–52.
- Xu, Dongxin, Umit Yapanel, and Sharmi Gray.** 2009. *Reliability of the LENA Language Environment Analysis System in Young Children's Natural Home Environment*. LENA Foundation.
- Zellman, Gail L., Rita Karam, and Michal Perlman.** 2014. "Predicting Child Development Knowledge and Engagement of Moroccan Parents." *Near and Middle Eastern Journal of Research in Education* 2014 (1): 1–17.